Informatica
from Salesforce

eBook

Scale AI with Data You Can Trust

A fast-track guide for data leaders to drive competitive edge

Where data & AI come to LIFE™

Contents

Part 1

Trusted Data Is the Foundation of AI Success

The AI revolution is here, and it's transforming how organizations make decisions, automate processes and innovate at scale. Agentic systems, powered by generative AI and autonomous agents, are reshaping competitive landscapes.

But as IT leaders and chief data officers race to operationalize AI, one truth is becoming increasingly clear: AI is only as trustworthy as the data it's built on.

While over 80% of CXOs recognize the rise of AI as the most significant opportunity, over half of them say that ineffective data flows between functions and time-consuming compliance and reporting are key constraints to true data-led enablement.[1]

Informatica's 2025 CDO Insights survey shows that 82% of data leaders find poor data quality is the top barrier to realizing ROI from AI investments.[2] Fewer than half feel fully prepared to scale AI responsibly, no matter how good the model, citing lack of confidence in the underlying data as the primary concern.

Legacy approaches to data quality, such as static rules, one-off attempts, siloed tools or bolt-on fixes weren't designed to handle the volume, velocity and variety of data powering modern AI systems. They can't keep up with the demands of real-time analytics, generative models, or autonomous agents. Instead, they add to tool sprawl and open quality and security gaps — slowing progress and increasing risk.

The question is no longer why data quality matters. It's how to get it right.

In this eBook, you'll find:

- What's changed and how data and technology leaders can rethink data quality for the AI era
- A clear, actionable guide, including a roadmap and best practices, to build a data quality foundation that enables trustworthy, scalable and responsible AI
- A checklist of the 10 must-have capabilities of an AI-ready data quality platform — helping you confidently begin your future-ready data quality journey

AI-era data quality isn't just about fixing errors. It's about enabling trust, speed and scale to elevate business outcomes.

[1] https://www.thomsonreuters.com/en-us/posts/corporates/c-suite-survey-2025/
[2] **https://www.informatica.com/lp/cdo-insights-2025_5039.html.**

What Got You Here Won't Get You There

Let's talk about why legacy data quality approaches fall short in the AI era. Advanced analytics, generative AI and agentic systems are redefining both the meaning of data quality and the methods to achieve it.

For data leaders, the stakes have never been higher. AI is no longer a future ambition; it's a present imperative. But the path to trusted, scalable and responsible AI starts with one foundational shift: rethinking data quality.

The Definition of Data Quality Has Changed

In the AI era, data quality isn't just about correctness. It's about:

- **Consistency** across sources and systems
- **Contextual relevance** to business goals and AI models
- **Connectivity** across the data ecosystem
- **Continuity** through real-time updates and adaptive rules

This expanded definition demands a new approach; one that goes far beyond traditional profiling and cleansing.

The Path to Data Quality Has Changed

Legacy practices such as manual rule creation, reactive cleansing and siloed governance can't keep pace with AI's real-time data ingestion and processing demands. They also struggle to support model drift and feedback loops or cross-functional AI pipelines.

AI-ready data quality requires new capabilities spanning automation, observability, scalability and embedded intelligence.

The Tools Enabling Data Quality Have Changed

Point solutions and bolt-on fixes only add to tool sprawl and vendor complexity. They leave critical gaps in governance, security and trust. Siloed approaches fragment your data strategy and your AI outcomes. Handcoding data quality logic is complex resource-intensive and not enterprise-ready. It often results in brittle, non-scalable solutions that cannot keep place with modern AI demands.

What's needed now is a unified, cloud-native solution that bakes data quality into every stage of your data lifecycle; not as an afterthought, but as a core capability deeply enmeshed with your data integration and larger data management strategy.

Figure 1 illustrates how Informatica's AI-ready data quality (DQ) foundation delivers measurable competitive advantages across key business dimensions.

Parameter of Comparison	Legacy DQ Frameworks	AI-Era DQ Demands
Approach	Reactive, rule-based	Proactive, intelligent, adaptive
Tooling	Point solutions, bolt-on	Embedded in data pipelines and platforms
Governance Integration	Minimal, disconnected	Governance and security are deeply integrated
Feedback Loops	Absent or manual	Continuous, ML-driven refinement
Trust Outcomes	Basic validation	Contextual, explainable, bias-aware
Enterprise Scale & Impact	Siloed, project-specific; fragmented investments, slow ROI	Enterprise-wide, unified platform; faster ROI through AI automation

Figure 1. Legacy vs. AI-Era Data Quality Paradigms: A Comparison

Rethinking Data Quality in the AI Era: Unified, Cloud-Native, AI-Powered

To build a future-ready foundation for AI, organizations must move beyond point solutions and integrate data quality into the entire data management lifecycle — from ingestion and transformation through consumption. This means choosing a solution that's unified, cloud-native, vendor-agnostic and powered by AI itself.

Informatica Cloud Data Quality (CDQ) as a service of **Intelligent Data Management Cloud™ (IDMC)** delivers exactly that: an end-to-end platform that embeds data quality into every stage of your data ecosystem.

- End-to-end coverage from ingestion and transformation to consumption
- AI-powered profiling and rule recommendations
- Near real-time monitoring and troubleshooting
- Vendor-agnostic, cloud-native architecture

With Informatica, data quality becomes a strategic enabler — not a reactive fix. It empowers confident decisions, reduces risk and accelerates innovation across analytics, GenAI and agentic systems.

Figure 2 outlines the competitive advantages of an AI-ready data quality foundation.

Dimension	Without a Modern Data Quality Foundation	With Informatica AI-Ready Cloud Data Quality Foundation
Revenue Growth	• Missed revenue opportunities from poor predictions and misinformed decisions. • Inconsistent data reduces campaign ROI and customer conversion. • Slower analytics cycles delay go-to-market.	• Trusted data drives more accurate insights, enabling data-driven decisions and better targeting. • Higher AI model accuracy better recommendations, personalization, cross-sell/upsell. • Faster time-to-market for data-driven products.
AI Readiness and Innovation	• 60-80% of data scientist time is wasted fixing or labeling bad data. • AI models produce biased or inaccurate outputs or cannot scale. • Innovation stalls due to lack of trusted data foundation.	• Reliable, ready-to-use data accelerates AI and analytics initiatives. • Reusable data quality assets across domains speed innovation at scale. • AI/ML teams spend more time building models, less time cleaning data
Operational Efficiency	• Fragmented, manual data prep in silos. • Repeated rework across projects. • Escalating data management costs.	• Automated data profiling, cleansing and rule management reduces manual effort. • Continuous monitoring prevents costly downstream fixes. • Lower total cost of ownership through reuse and automation.
Regulatory Compliance & Risk	• Inaccurate reporting risks non-compliance and fines. • Hard to prove lineage, accuracy and data source trustworthiness. • Increased audit fatigue and remediation costs.	• Centralized, governed and auditable DQ ensures compliance with GDPR, HIPAA, CCPA, etc. • Reduces financial and reputational exposure from reporting errors. • End-to-end lineage, auditability and policy enforcement to support explainable AI.

Figure 2 outlines the competitive advantages of an AI-ready data quality foundation. (Continued)

Dimension	Without a Modern Data Quality Foundation	With Informatica AI-Ready Cloud Data Quality Foundation
Customer Experience	• Duplicate or inaccurate records lead to poor personalization and service issues. • Customer frustration and brand attrition.	• Clean, unified and enriched customer data powers personalization, retention and satisfaction. • Consistent view of customers across systems and touchpoints.
Reputation & Trust	• Data errors damage credibility with executives, regulators and customers. • Perceived as lagging in data maturity.	• Reliable, transparent data builds stakeholder confidence in analytics and AI decisions. • Positions the organization as a data-driven leader.
Decision Agility	• Decision delays due to data distrust and manual validation. • Reduced ability to respond to market changes.	• Real-time, trusted insights accelerate decisions and scenario modeling. • Business users are empowered with governed self-service.
Data Governance Synergy	• Fragmented governance efforts. • Limited visibility into data lineage and impact.	• Integrated metadata, lineage and governance with catalog and policy management. • Holistic view of data health across the enterprise.
ROI on AI	• AI models cannot scale or degrade quickly as data drifts or changes. • Ongoing model retraining costs and failures increase.	• Continuous DQ observability ensures models remain accurate over time. • Foundation for long-term, enterprise-scale AI success.

Figure 2. The Competitive Advantages of Investing in a Modern, AI-Ready Data Quality Foundation

The Real-World Consequences of Poor-Quality Data

Poor data quality can lead to high-profile failures across industries.

- **Retail:** Fragmented customer data can degrade personalization efforts, leading to churn and lost revenue.
- **Finance:** Flawed data can distort credit scoring models, resulting in mispriced loans and compliance violations.
- **Healthcare:** Inconsistent patient records can delay treatment and trigger regulatory scrutiny.

Part 2

How to Build an AI-Ready Data Quality Foundation

Defining AI-Ready Data: 3 Pillars of Trust

Pillar 1: Accuracy and Integrity

AI-ready data is correct, consistent, connected and continuous data across all environments

AI-ready data must reflect reality with precision. That means eliminating duplication, resolving inconsistencies and ensuring seamless connectivity across cloud, streaming and hybrid environments — often in real time.

For example, in healthcare, inaccurate or fragmented patient data across systems can lead to misdiagnoses, treatment delays and compliance failures. Conversely, high-fidelity data — unified and validated in real time — enables AI to support clinical decisions, optimize resource allocation and improve patient outcomes.

With Informatica Cloud Data Quality (CDQ), healthcare providers can ensure that every AI-driven insight is grounded in accurate, trusted data — even across complex hybrid cloud environments.

Informatica Capabilities

- **Cloud-native profiling and cleansing** across structured and unstructured data sources
- **Near real-time monitoring and anomaly detection** to catch issues before they impact models
- **End-to-end lineage and metadata intelligence** for full traceability and auditability

Pillar 2: Completeness and Context

AI-ready data needs unified, enriched views for trusted decision-making

AI systems thrive on holistic understanding. They perform best when they can see the full picture, not fragmented glimpses. Without complete, contextual data, even the most advanced models can make flawed assumptions, leading to poor decisions and missed opportunities.

For example, in retail, incomplete customer data, such as missing transactions, disconnected profiles, or outdated preferences, can derail personalization efforts and skew churn predictions. But when data is enriched, unified and contextualized, AI can deliver tailored experiences, optimize inventory and drive loyalty.

Retail leaders use Informatica to build complete, contextual customer views, enabling AI to personalize at scale and predict behavior with confidence.

Informatica Capabilities

- **360-degree data mastering** for customers, products and suppliers
- **Context-aware enrichment** using external and internal sources
- **Semantic matching and deduplication** to unify fragmented records

Pillar 3: Alignment and Fitness for Use

Strategically aligned data that meets AI model requirements

Not all data is created equal — and not all of it is fit for AI. To be useful, data must match the format, granularity and semantics required by specific models and business objectives. For global enterprises deploying agentic AI across regions, inconsistent formats and definitions can lead to model drift, misinterpretation and compliance risk. Standardization and strategic alignment are key.

In hospitality, for example, hotel chains may use agentic AI systems to automate guest personalization across regions, from room preferences to loyalty offers. But inconsistent date formats, varying definitions of "premium guest," and fragmented booking records across geographies can lead to misaligned recommendations and customer dissatisfaction.

With Informatica Cloud Data Quality, international hotel chains can harmonize global data schemas, enforce semantic consistency and align data with model requirements. This lets the AI agents deliver context-aware, culturally relevant experiences, boosting upsell conversions and improving guest satisfaction scores across markets.

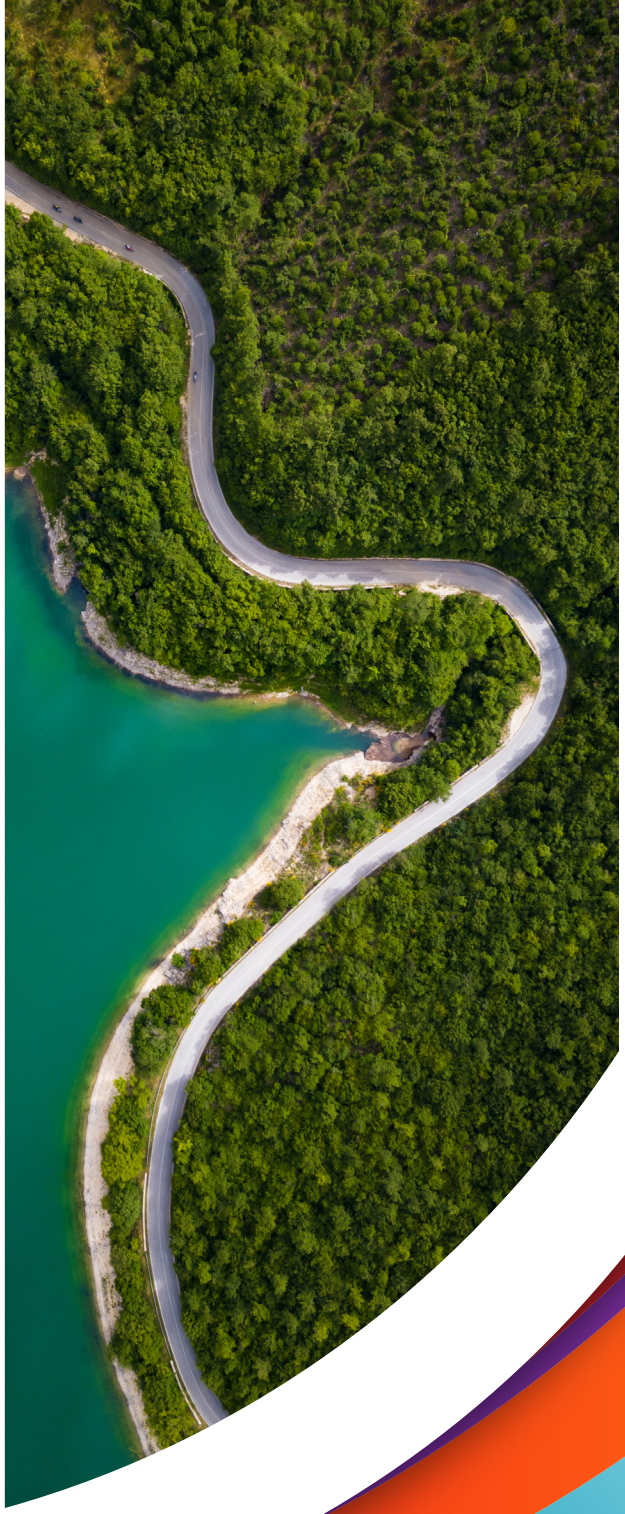
Informatica Capabilities

- **Automated schema harmonization** across geographies and systems
- **Business glossary and policy enforcement** to align data with strategic goals
- **ML-driven recommendations** for format and transformation optimization

Operationalizing Strategic Data Quality: The CDOs Roadmap

For AI, analytics and transformation to succeed, data must be business-ready, context-rich and continuously reliable. This requires a robust data quality foundation.

- **Align data quality with** business outcomes such as customer experience, operational efficiency, risk mitigation, and innovation
- **Start with high-impact, outcome-linked use cases** where enabling trusted data can quickly demonstrate value, such as improving cross-channel personalization, accelerating regulatory reporting, or increasing supply chain visibility.
- **Scale through a data quality center of excellence (CoE)** that unifies governance, tools, and best practices across business units, embedding data quality into every data product and pipeline.
- **Measure business value:** quantify revenue uplift from better targeting, efficiency gains from automation, and risk reduction from trusted reporting. Treat data quality as a living KPI, not a one-time milestone.
- **Build a culture of trust and accountability,** empowering data stewards, analysts, and engineers with transparent metrics and assigning shared responsibility of quality. Turn data trust into an enterprise-wide discipline that supports AI-driven innovation.



Part 3

Building the Architecture of Trust

3 Best Practices for an AI-Ready Data Quality Foundation

Ensuring data quality for reliable AI and agentic systems isn't just a technical requirement — it's a strategic advantage. Informatica's **Intelligent Data Management Cloud (IDMC)** offers a proven framework to operationalize trust at scale. These three best practices help data leaders build a resilient, future-ready foundation for AI.

Best Practice 1: Intelligent Automation in Data Quality

Let AI power your data quality for AI

Modern data ecosystems are too complex and fast-moving for manual data quality management. Informatica's CLAIRE® AI engine brings intelligent automation to every stage of the data quality lifecycle; from discovery to integration and remediation.

AI-powered automation lets you scale data quality across thousands of pipelines — without scaling complexity. CLAIRE's benefits include:

- **Metadata Management:** CLAIRE automatically catalogs and classifies metadata, enabling semantic consistency across sources.
- **Data Observability**: Near real-time monitoring detects anomalies, schema drift and quality degradation before they impact AI models.
- **AI-Powered Profiling and Cleansing:** Informatica CDQ uses machine learning to recommend rules, detect outliers and automate cleansing at scale.
- **Identity Resolution and Deduplication:** Master data integrity is critical for AI trust. Informatica's advanced matching algorithms unify records across systems, delivering trusted 360-degree views.

Best Practice 2: Unified, Enterprise-Scale Data Quality Framework

Trust must be consistent across every cloud, every domain

AI doesn't operate in silos — and neither should your data quality strategy. Informatica enables a unified approach that spans hybrid and multi-cloud environments, integrating seamlessly with governance, MDM and data engineering workflows.

You gain:

- **Human-in-the-Loop Validation:** Combine automation with expert oversight to ensure critical decisions are grounded in verified data.
- **Cloud-Native Scalability:** IDMC is built for distributed architectures, enabling consistent data quality enforcement across AWS, Azure, GCP and on-premises systems.
- **Governance Integration:** Informatica's data governance and catalog tools ensure that data quality rules align with business policies, regulatory mandates and model requirements.

Best Practice 3: Embed Data Quality Proactively Across the Data Lifecycle

Trust isn't a checkpoint. It's a continuous practice.

AI systems require continuous data health to remain reliable. Informatica enables proactive, embedded data quality across ingestion, transformation and consumption stages, making trust a built-in feature — not a bolt-on fix.

This approach enables:

- **Continuous Data Observability:** Monitor data quality in motion, not just at rest — catching issues before they impact dashboards or models.
- **Lifecycle Integration:** Embed data quality into ETL, ELT and MLOps pipelines to ensure trusted data powers every decision.
- **Preventative Quality Management:** Shift from reactive cleansing to proactive prevention to reduce risk and accelerate AI outcomes.

Part 4

Next Steps to Operationalize Your AI-Ready Data Quality Foundation

When it comes to operationalizing enterprise data quality strategy, CIOs tend to focus on architecture, tooling, automation and scalability; while CDOs address strategy, policy, stakeholder alignment and ROI.

Choosing the right data quality platform is one of the key decisions that impacts both strategic and technical outcomes.

Figure 3 outlines the 10 key capabilities to operationalize your enterprise data quality strategy.

	Capability	Why It Matters for AI	The Informatica Data Quality (DQ) Advantage
1	**AI-Powered Data Profiling**	Manual profiling is slow, error-prone and doesn't scale across diverse, high-volume data sources. You need complete visibility into your data (accuracy, completeness, patterns) before you feed it into AI/ML.	**Data Quality**, a service of IDMC, uses intelligent, automated profiling to detect anomalies, missing values and data patterns across cloud-native platforms, enabling rapid assessment of data readiness for AI
2	**AI Powered Cleansing / Standardization / Enrichment**	Inconsistent, incomplete and unreliable data undermines analytics, AI models and operational decisions, leading to biased models and poor decision outcomes	Data Quality applies rule-based and DQ Agent - assisted cleansing, formatting and enrichment logic to transform raw data into consistent, complete and context-rich inputs for trustworthy AI models
3	**ML-Driven Rule Recommendations**	Manual rule creation won't scale for large, varied, high-velocity data; AI can help surface anomalies and generate suggestions at scale.	CLAIRE DQ Agent, a rule generation feature in IDMC, assists by suggesting intelligent cleansing and validation rules based on data behavior and metadata.
4	**Continuous Data Quality Monitoring**	Data quality is not a one-off: AI needs ongoing data quality monitoring over time, across dynamic data flows	Near real-time monitoring and observability ensures data reliability at speed, catch drift, schema changes and degradation before AI is impacted.
5	**Embedded Governance and Policy Enforcement**	A lack of real-time visibility into data quality issues leads to undetected drift, anomalies or broken pipelines.	Embedded governance and observability ensures data accountability and compliance at scale. Together, they form the backbone of AI-ready data quality operations.
6	**Cloud-Native Scalability**	AI workloads demand massive, dynamic and distributed data processing. AI pilots stall due to bottlenecks in handling large, diverse datasets.	Informatica Intelligent Data Management Cloud (IDMC) operates seamlessly across hybrid and multi-cloud environments with elastic scaling, pushdown optimization and serverless execution

Figure 3 outlines the 10 key capabilities to operationalize your enterprise data quality strategy. (Continued)

	Capability	Why It Matters for AI	The Informatica Data Quality (DQ) Advantage
7	**End-to-End Lineage and Traceability**	AI models often fail or produce biased, unreliable results because data scientists lack visibility into the origin, transformation and trustworthiness of the data feeding those models.	**Informatica Data Catalog** with integrated lineage and cloud data governance provides end-to-end lineage, traceability of quality scores and metadata intelligence which enriches data assets with context, ownership and usage insights, critical for model explainability.
8	**Master Data Management Integration**	Inconsistent, duplicate or siloed master data undermines AI model accuracy and governance. Without trusted, unified master data, AI systems struggle with poor feature engineering, inaccurate predictions and compliance and explainability risks from opaque or incorrect data.	**Informatica Master Data Management (MDM) and Customer 360** ensures identity resolution and deduplication for trusted 360-degree views. Integration with data quality rules ensures that AI models are trained and operated on golden records, delivering more accurate predictions, better personalization and stronger regulatory alignment.
9	**Semantic Consistency and Metadata Intelligence**	Without semantic consistency, AI models may misclassify entities, leading to flawed insights. Data integration may become brittle, slowing down AI deployment and causing business users to lose trust in AI outputs due to inconsistent terminology or logic.	Through metadata intelligence, data profiling, lineage tracing, standardized business glossaries, classification, governance and catalog, Data Quality ensures semantic consistency—enabling unified definitions, accurate data context and trusted inputs for AI and analytics models.
10	**API-First Architecture for MLOps Integration**	Disconnected, manual and non-scalable data quality workflows in AI pipelines mean DQ logic isn't portable or automatable, making it hard to embed into CI/CD pipelines or retraining workflows.	Informatica IDMC's Data Quality's API-first architecture, with REST APIs, SDKs, serverless execution and native integration with platforms like Snowflake, Databricks and AWS enables seamless, scalable embedding of data quality into AI/ML pipelines for proactive, ongoing quality management.

Figure 3. 10 Must-Have Capabilities in an AI-Ready Data Quality Platform

Trusted Data Is the Only Path to AI Confidence and Accelerated Innovation

In the AI era, data quality isn't just a technical necessity — it's a growth driver. It's the difference between AI that informs and AI that misleads — between agentic systems that scale innovation and those that amplify risk.

To build AI you can trust, you need data you can trust. That requires a data quality foundation built to withstand the complexity, velocity and continuous evolution of modern enterprise data.

Informatica Cloud Data Quality (CDQ) delivers that foundation. It's not just a tool. It's an intelligent, cloud-native platform that embeds trust into every stage of your data lifecycle. Powered by the **CLAIRE AI engine**, Informatica CDQ offers:

- Automated profiling, anomaly detection and cleansing at enterprise scale
- Real-time data observability to prevent drift and degradation before it impacts AI
- ML-driven rule recommendations to adapt to changing data and model needs
- Seamless integration with governance, MDM and cloud data platforms
- End-to-end lineage and explainability to meet regulatory and ethical standards

These capabilities directly contribute to confident decisions, reduced risk and accelerated innovation: the three outcomes every data leader needs to deliver in the AI era.

As you develop your AI-ready, future-resilient data management strategy, ensure your solution includes enterprise-grade data quality capabilities that embed trust into every AI initiative.

Begin the journey with us. **Schedule your strategy session.**

About Us

About Informatica

Informatica from Salesforce is a leader in AI-powered enterprise cloud data management. Its Intelligent Data Management Cloud (IDMC) platform enables organizations to connect, manage and unify AI-ready data across the enterprise. With capabilities spanning data cataloging, integration, governance, quality, privacy, metadata management and master data management, Informatica supports a broad partner ecosystem and helps customers unlock the full value of their data and AI initiatives.

About Salesforce

Salesforce is the #1 AI CRM, empowering companies to connect with their customers in a whole new way through the power of artificial intelligence, data, and trust. For more information about Salesforce (NYSE: CRM), visit: **www.salesforce.com**.

Worldwide Headquarters
2100 Seaport Blvd.
Redwood City, CA 94063, USA
Phone: 650.385.5000
Fax: 650.385.5500
Toll-free in the US: 1.800.653.3871

informatica.com
linkedin.com/company/informatica
x.com/Informatica

CONTACT US

IN19-5282-1225