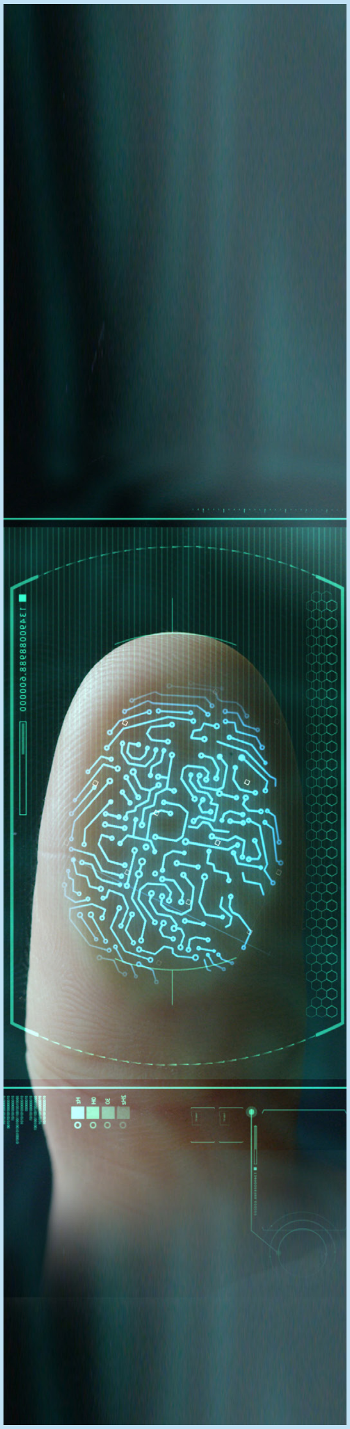
Infosys topaz

VIEW POINT



CONSTRUCTING A DATA CANOPY FOR EFFICIENT AI SYSTEMS

Smart Data Fingerprinting (SDF) can make enterprises data-ready for the future of AI

Infosys®
Navigate your next

INSIGHTS

- Global data creation is estimated to reach 180 zettabytes per year by 2025.
- However, only **20%** of the world's data is currently governed.
- Data compromises in 2023 reached an all-time high, with an increase of **78%** points, compared to 2022.
- To tackle volume, complexity, and security concerns for all types of data, autonomous data management processes like data fingerprinting have become a need of the hour.

Imagine you're exploring the incredible biodiversity of the Amazon rainforest. What's your choice of path?

A blind leap into the unknown without a safety net?

Or

Being prepared for the road ahead with an action plan? A plan that includes (i) familiarizing yourself with a map of the forest's diverse flora and fauna, (ii) staying updated on travel advisories and weather warnings, (iii) prioritizing legal considerations and responsible behavior, and (iv) demonstrating cultural sensitivity and ethical behavior.

The latter, right? Much like encountering the dense and uncharted territories of the Amazon, navigating the expanses of the ever-flourishing 'data rainforest' can be daunting; reiterating the importance of being well-prepared before venturing into unique and complex ecosystems.

Here's an understanding of the journey

Exploring the current state of the data rainforest

With AI shifting the focus to new expanses of data, unstructured data sets in the form of files, emails, websites, software codes, videos, audio, sensor data, IoT data, etc., are scattered in different formats across various locations outside the traditional databases. This makes it difficult to process and protect the data, leading to the downfall of the potential AI impact. Hidden data in uncharted territories can hinder traditional analytics, let alone an AI-powered one.

Furthermore, IDC estimates that only around 20% of the world's data is currently governed. The remaining 80% is a significant volume of data that remains ungoverned globally. Ergo, there is a significant risk for businesses regarding trust, ethics, security, compliance, and privacy, if an AI application consumes such data.

Ungoverned data makes for blind spots, hindering the ability to make the most of AI and generative AI projects to drive innovation. Failing to address this, creates a perfect storm of lost business value and jeopardizes brand reputation.

Though AI/generative AI can unlock groundbreaking solutions and advancements; the volume, velocity, and complexity of all data, requires well-thought-out approaches that deliver continuous data fingerprinting and autonomous processing while taking care of confidentiality and compliance considerations. Ultimately, autonomous data management processes act as a compass to guide enterprises in handling massive datasets effectively.

Traditional maps are not the best tools to find your way forward

The legacy data management solutions in the market are ill-equipped to navigate unchartered data territories. Like maps of yesteryears, they are limited in competency: built for analytical and historical workloads, focused on already governed data, and dependent on manual processes. They also struggle to keep pace with the evolving data landscape, fast-changing AI models and technologies, market regulations, compounding AI, and sovereignty needs.

Current data management solutions often lack interoperability with the organization's existing systems (emails, ERP, audio/video on customer service systems, etc.). This creates data silos and impedes the ability to gain a holistic view of information for effective data governance.

Recognizing these limitations, the global community is actively seeking solutions like data fingerprinting that can help navigate through the lush yet entangled data landscape. To better understand this need, let's take the example of an AI-powered mortgage system that receives a loan application from a consumer. The system must:

- Gather and verify customer data from various sources, such as previous email communications, a call recording expressing pain points, etc.
- Then, analyze the data to assess the consumer's eligibility and creditworthiness for the loan, recommending actionable steps to the loan officer.

In this scenario, how does data fingerprinting secure the outcome? Let's explore.

Smart Data Fingerprinting (patent pending): A guiding ray of light

Infused with AI/generative AI capabilities, autonomous data management processes, like data fingerprinting, provide a granular, in-depth, and all-encompassing vantage point of the data jungle. It fingerprints relevant data across all data types in the organization – forming the roots to build and operate AI systems confidently. Imagine doing this for millions of data objects such as audio, video, files, websites, emails, etc.!

The fog of overwhelming confusion suddenly clears up, and data becomes insights that make sense.

Similarly, for the bank to trust the outcome of the AI-powered mortgage system:

- It must ensure that the intake data (such as emails, call recordings, etc.) is authentic and accurate while ensuring compliance with privacy laws and other regulations.
- Here, an automated data fingerprinting process would responsibly granularize and fingerprint relevant emails, audio recordings, etc., including ingrained access controls of business users, based on the sensitivity and confidentiality of the content.

Businesses need autonomous data fingerprinting to keep up with the volume, velocity, and complexity of such data landscapes.

The data canopy for becoming an AI-first enterprise

Data, from user-generated content to real-time to historical and machine data, can make for a noisy jungle, adding to the challenges of data exploration and unlocking value. Implementing continuous autonomous data fingerprinting proves vital to securely navigate the complexities of such pervasive data in a unified way.

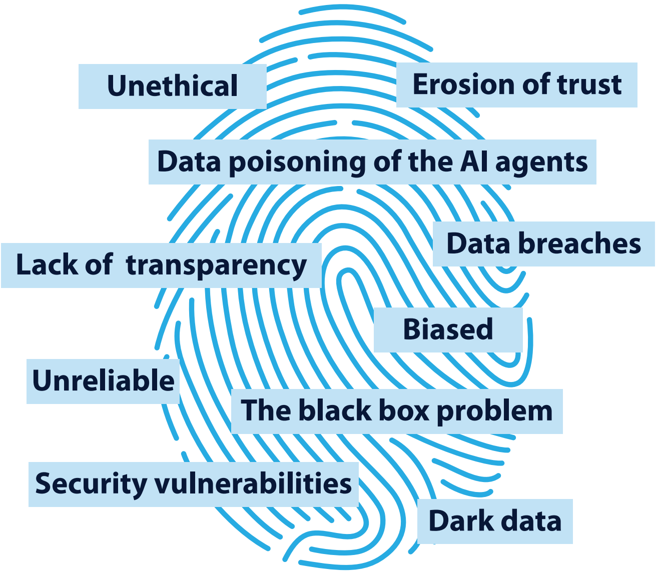


Figure 1: Implications of missing data fingerprinting in AI

Along with covering traces of any data that is generated while navigating the digital landscape, fingerprinting the data forms an immutable and verifiable data transaction chain that ensures traceability, non-repudiation, auditability, and accountability – the foundational elements of trusted data.

In addition to smart data classification, smart cataloging, and smart search, this autonomous process helps identify risks and trigger mitigation, before they are amplified by AI or generative AI systems.

Often, our greatest strength lies not in the initial plunge but in the deeper understanding and preparation that paves the way for a successful business journey.

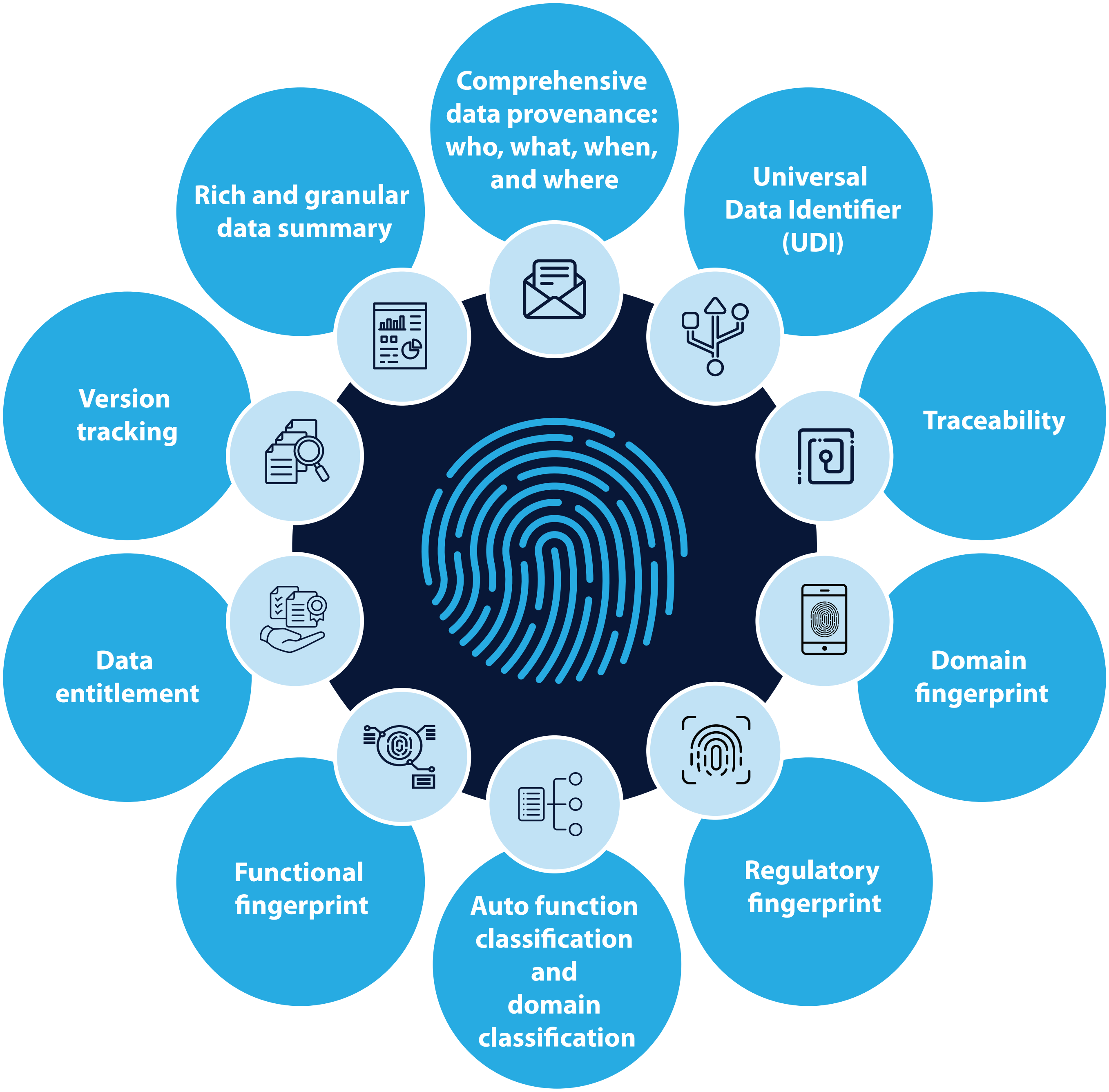
Implementing robust data fingerprinting instills confidence in an enterprise's ability to process data for AI and establishes them as a trusted authority and a credible player in the AI business world. It also helps enterprises build a flourishing ecosystem of customers and partners, leading to a significant shift in the outlook on AI:

- **From risk aversion to innovation at speed and scale:** Data fingerprinting uncovers more data with transparency and traceability; thus, providing a foundation for experimentation for any AI/generative AI project. Easy access to diverse datasets opens new trails to 'responsible data sandboxes' for creative exploration and experimentation to drive organizational growth.
- **From data compliance to trust:** It moves beyond treating data management as a mere compliance exercise. Instead of just box-ticking processes, data fingerprinting emphasizes its ability to generate insights, drive innovation, and unlock significant business value.
- **From silos to collaboration:** It breaks down departmental data silos and fosters a culture of diversity and inclusivity. Further, it promotes seamless data exchanges between different systems, including AI/ML applications. This simplifies cross-functional collaboration and maximizes the impact of AI/data product launches.
- **From fear to empowerment:** Replace fear of data misuse with data literacy and empowerment. It ensures that data is protected from unauthorized access and uses data in accordance with relevant regulations and ethical principles. It also equips data consumers to navigate the data landscape effectively and responsibly.

Enterprises need a partner who can illuminate the path along the data rainforest with a future-proof approach to AI ecosystems, built with autonomous capabilities.

Venture with a strong sense of direction and necessary skills

The **Infosys Smart Data Fingerprinting (SDF)** for AI solution does the major foundational work before diving headfirst into any AI or generative AI project. It is a sustaining solution that highlights the impressions of relevant and useful territories in a phased approach, playing a vital role throughout the data and AI lifecycle.



Comprehensive data provenance: who, what, when, and where
Universal Data Identifier (UDI)
Traceability
Domain fingerprint
Regulatory fingerprint
Auto function classification and domain classification
Functional fingerprint
Data entitlement
Version tracking
Rich and granular data summary

Figure 2: The tenets of data fingerprinting

With SDF, implementing the intelligent data fingerprinting process gives **a unified and holistic view of all forms of data** present across the enterprise. Our solution reimagines traditional and **manual-ridden processes** and propels businesses ahead in the AI race, resulting in a significant accomplishment for an enterprise in their AI/generative AI journey.

Infosys' innovative approach of fingerprinting all data types offers distinct advantages in gaining granular insights over existing solutions:

Flexible configuration	Agility and scalability	Interoperability	Knowledge harvest
Effortlessly tailor the solution to specific enterprise needs with flexible configuration capabilities.	Through an intelligent core that is central to SDF, it continuously learns and adapts to future data needs. In addition to self-refinement, it adjusts parameters and refines its performance over time.	Seamlessly connect with existing systems and tools, integrate effortlessly with the current technology landscape, and break down data silos to achieve unified operations.	Harvest knowledge beyond simply gathering information. Curate fine, ingrained information from disparate, interconnected, and various intertangled data sources to get a comprehensive understanding.

Leap ahead in the race of data and AI with SDF

When implemented cohesively, data fingerprinting strategies become enablers of various value drivers in the enterprise. This strategic asset, not only improves the organization's overall operations, but also opens up additional revenue sources by monetizing data, maximizing business value, improving compliance, and minimizing the downsides.

Becoming an AI-first organization and developing a data culture is a high-magnitude task.

However, the potential benefits for innovation, efficiency, and overall business success, make the data journey worthwhile for companies of all sizes.

Just like the action plan that was meticulously crafted for the Amazon rainforest exploration, Smart Data Fingerprinting can empower enterprises to become data-ready for AI and capitalize on business opportunities to maximize value.

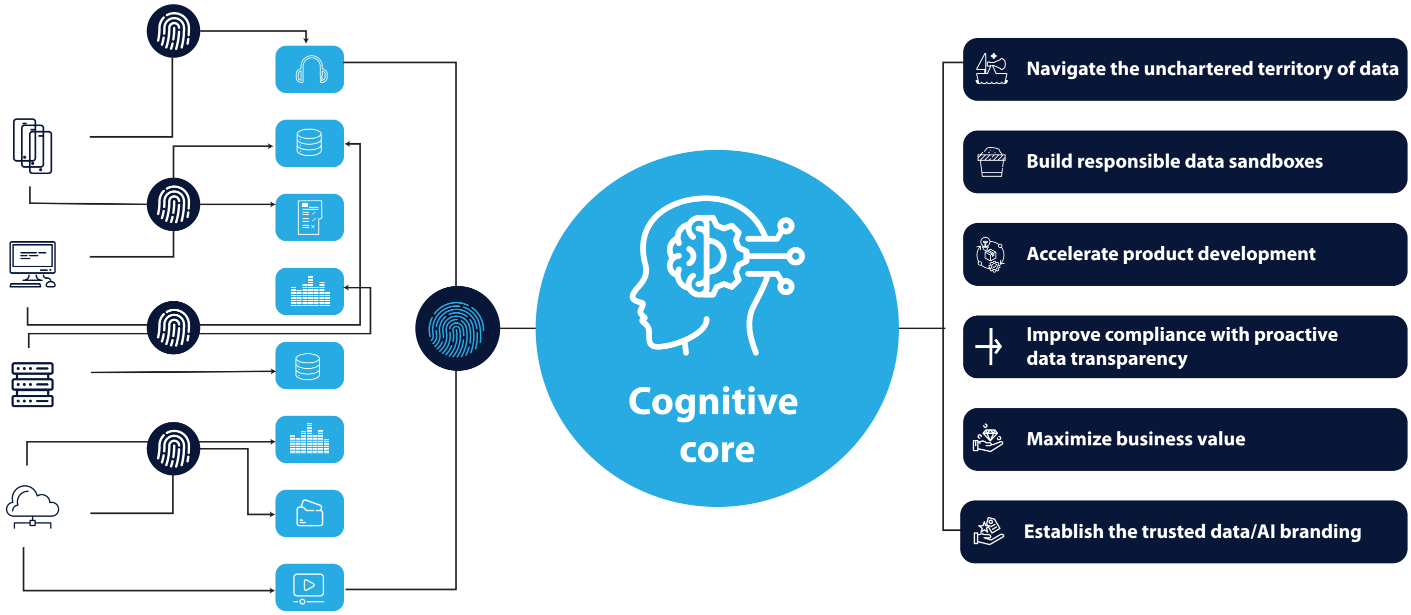

Cognitive core
Navigate the unchartered territory of data
Build responsible data sandboxes
Accelerate product development
Improve compliance with proactive data transparency
Maximize business value
Establish the trusted data/AI branding

Figure 3: Value drivers enabled by data fingerprinting for AI

Infosys Topaz is an AI-first set of services, solutions, and platforms using generative AI technologies. It amplifies the potential of humans, enterprises, and communities to create value. With 12,000+ AI assets, 150+ pre-trained AI models, 10+ AI platforms steered by AI-first specialists and data strategists, and a 'responsible by design' approach, Infosys Topaz helps enterprises accelerate growth, unlock efficiencies at scale, and connected ecosystems.
Connect with us at **infosystopaz@infosys.com.**

For more information, contact askus@infosys.com

Infosys®
Navigate your next

Infosys.com | NYSE: INFY

Stay Connected