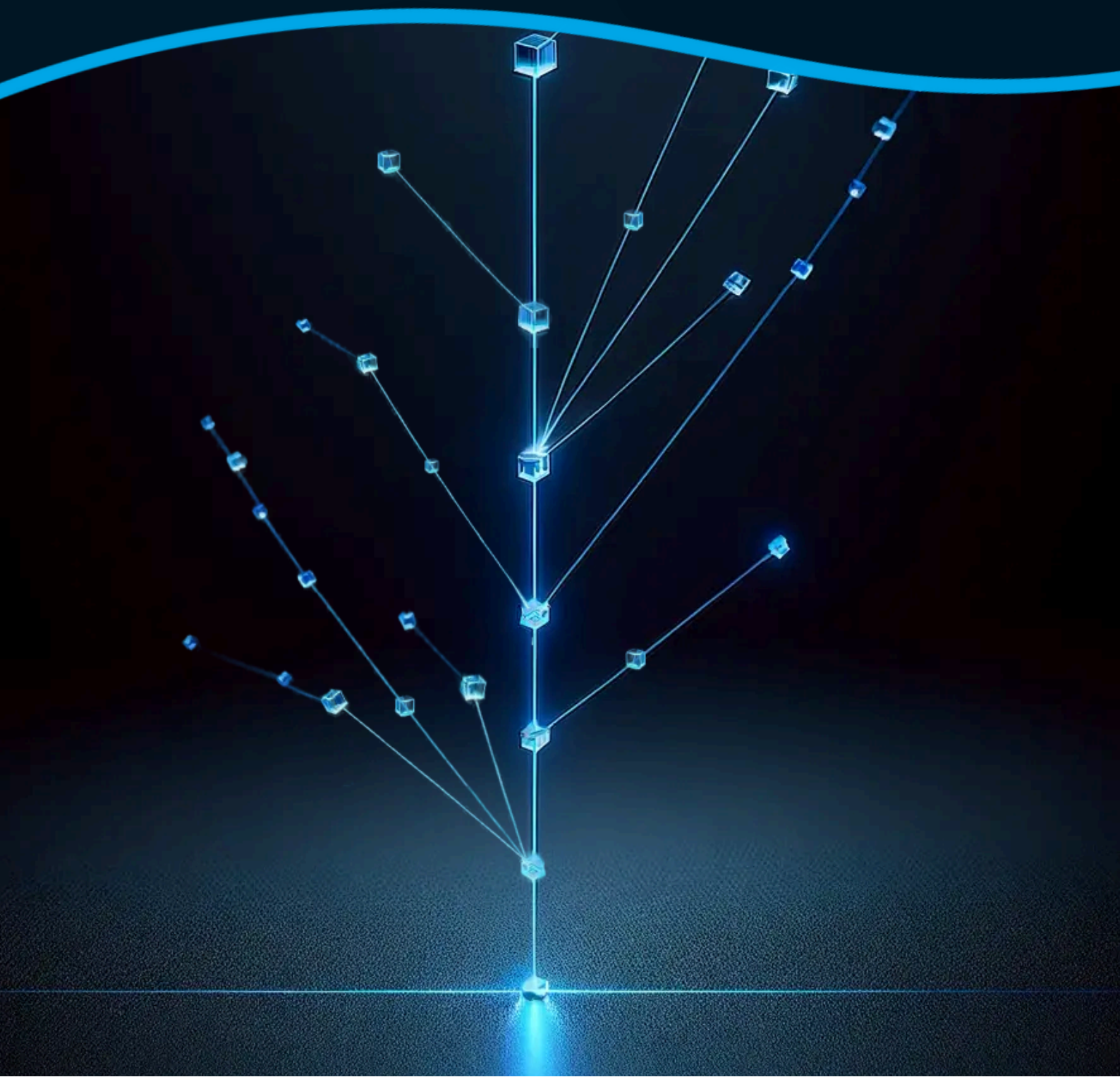
DataTrails

March 2024

RESPONSIBLE AI

Deploying Artificial Intelligence in industry responsibly through transparency, accountability, and explainability.



CONTENTS

01 HUMANITY AT A CROSSROADS

AI adoption and digital transformation are driving forces in modern business, offering unprecedented opportunities for growth, efficiency and improving lives. However, with this promise comes the responsibility to ensure that AI is trustworthy, transparent, and aligned with the goals of the businesses and communities they serve.

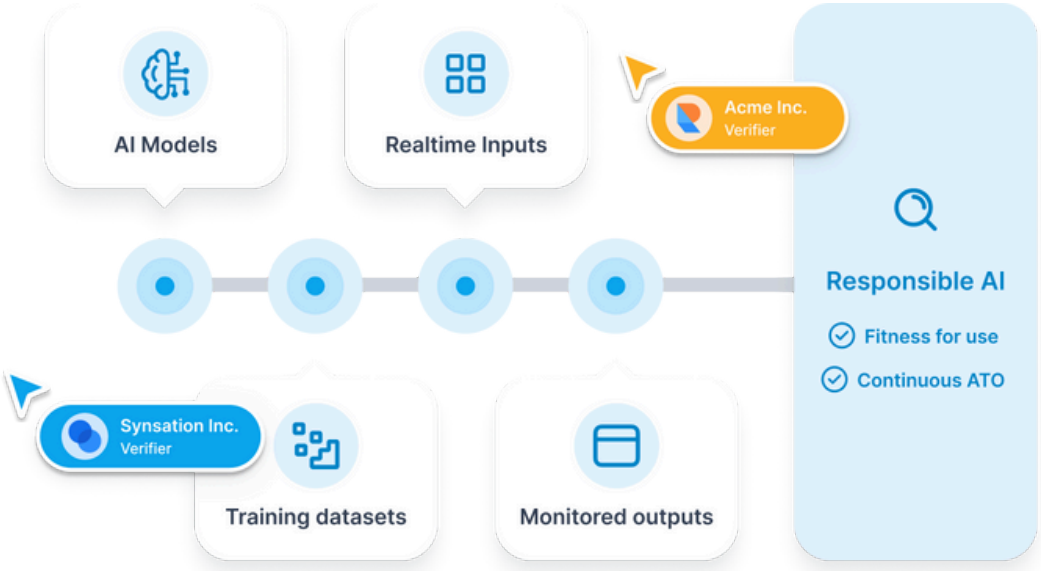
There are significant challenges to keeping this alignment, particularly when it comes to data governance, explainability and trust. For businesses to fully embrace AI and digital transformation, two major hurdles must be overcome:

- knowing that your AI-driven system is fit-for-use
- knowing that its actions and the data it produces can be validated by standard compliance and audit processes.

A key enabler of Responsible AI is the effective management of provenance: the lineage and origin of data and AI models, and the ability to detect potential issues that can arise from any part of the AI software supply chain from training data to reinforcement learning to deployment context.

DataTrails addresses this challenge head-on by providing a platform that facilitates trustworthy multi-party auditability and explainability for the critical decisions made by AI, fostering confidence in workflows and supporting responsible digital operations.

In this paper, we'll explore how DataTrails' provenance management platform plays a central role in achieving Responsible AI, using transparency and accountability to help us harness AI and achieve greater trust in the machines.



02 UNLEASH THE MACHINES...RESPONSIBLY

With great power comes great responsibility. Adopting AI in industry and enterprise workflows hands over decision-making powers, risk, and – little-by-little – corporate strategy to machines. Meaningful tasks are only given to a human employee once that person is suitably qualified and trained, and once in post regular performance monitoring ensures they are doing a good job. While the details may be very different, these fundamental needs are also present with any AI agent given the same responsibilities.

It is unreasonable to expect machines to be 100% perfect all the time, for no better reason than it is impossible to predict what 100% perfect operations and decision-making looks like.

But we must demand that machine decisions be at least as good as human decisions, and as auditable and accountable as well. Before deploying AI into positions of responsibility at scale we need to know that they can fit into our existing social structures and processes for fault-finding and remediation. Without that, we'll never get meaningfully off the starting blocks.

While AI Models and neural networks are often considered to be 'black boxes', making it hard – even impossible – to know why an AI made a particular decision, the proliferation of AI into all aspects of businesses from products to operations has resulted in new requirements for AI to be auditable and explainable.

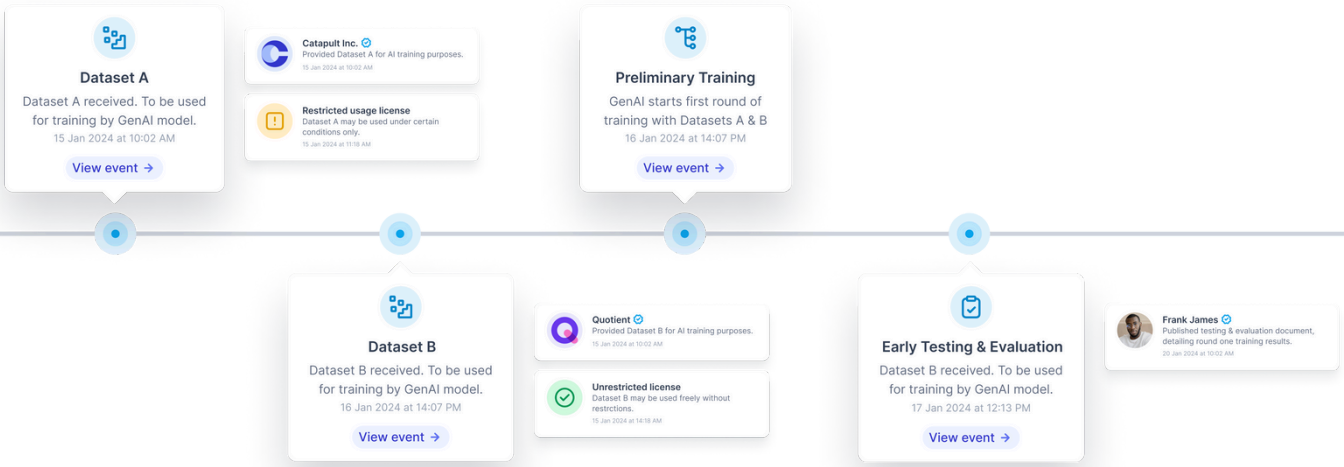
Explainable AI

Explainable AI is a broad field related to technical measures within AI and ML systems that prioritize principles of transparency and explainability in machine-based decisions in order to understand, trust, and improve the decision-making process. While critics of Explainable AI claim that the addition of determinism and traceability reduce the functional benefits of AI models, recent moves in the field to combine model changes with oversight techniques such as supply chain transparency and data authenticity offer much promise.

Responsible AI

Responsible AI is an interdisciplinary field concerned with aligning the use and consequences of AI systems to the moral and social good of society. Combining sociological disciplines such as machine ethics and AI alignment with technical approaches to monitoring and oversight, RAI includes accountability as a cornerstone of responsible operation. What's responsible in one situation may not be in others, so effective Responsible AI relies on the availability of reliable evidence on which to make accountable, responsible decisions.

03 AI TRANSPARENCY IS KEY



The delivery of functioning AI-based solutions from the workbenches of researchers to the offices of customers requires a complex supply chain. Foundation models and training sets need to be sourced, then trained and refined, and every step taken and choice made has major impacts on the output of the final system.

With the rising need for AI systems to be auditable and explainable, those supply chain events need to be recorded in ways that enable developers to make informed choices in creating systems out of constituent parts, and allow users to make informed choices in the use of those systems in their business operations. Just like any supply chain risks creep in and trust breaks down where visibility is low.

By adopting general principles of supply chain transparency and making the major steps auditable, in a single record, no matter where they occur, industry can quickly improve the state of trust and responsible operations in AI.

These auditable transparent records describe the *provenance* of software, models, and data which enables confident choices to be made by both developers and users. Being properly informed of the foundation models used and how they were modified, the quality and origins of training data, and the qualifications and certifications of businesses refining the final system means that operators are not only more sure that they are using AI reponsibly but they can prove it too.

All this information is potentially sensitive, of course, and so while transparency among stakeholders is essential, fears of data leakage outside of stakeholder groups could hold back adoptions and innovation. Fortunately the Internet Engineering Task Force (IETF) and Linux Foundation are both working on open standards for content provenance that combat the erosion of trust in digital technologies whilst affording appropriate business confidentiality.

It's not just matters of initial development and training that impact trustworthiness considerations. As AI agents increasingly communicate and automate multi-party collaboration they will inevitably trade data and learn from each other. On the positive side that means more tailored responses and a larger pool of data to draw on, but along with that baby comes the bathwater of amplified hallucinations, biases, and ethical concerns.

This adds a need for continuous updates to the provenance record to maintain knowledge of who (or which machines) you're learning from, and where (what sources) the data that drives your decisions came from.

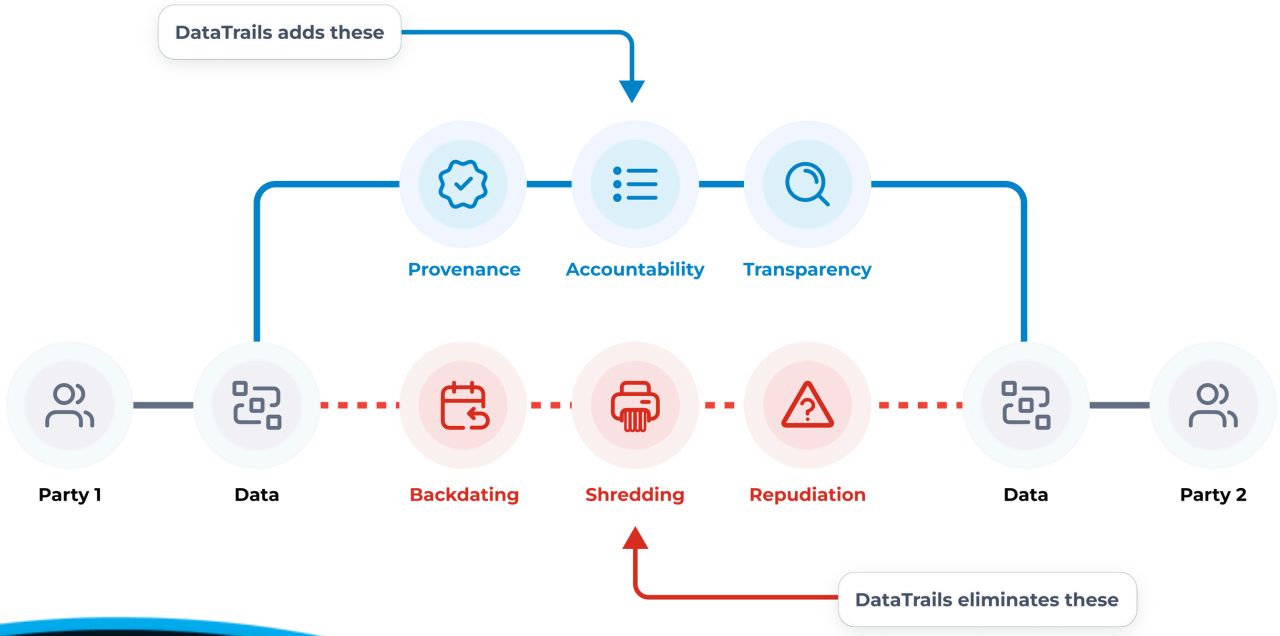
Because every business is different and responsible operations depend on circumstances, access to reliable evidence is a key part of adding a Responsible AI component to your business. This starts with Explainable AI technology that fortifies trust in shared systems and eliminates concerns about back-dating (or forging) evidence, shredding (or withholding) evidence, and repudiation (or disputing) evidence. An immutable transparency log is a very effective antidote to all these issues.

Food for thought

This is not a case of pure quality or binary good vs bad data. It's about the whole process and intention behind their creation.

A surgeon and a lawyer may be equally qualified and subjectively responsible, but which would you trust with heart surgery?

Again, fitness-for-use is the crucial question, and transparency and provenance hold the answer.



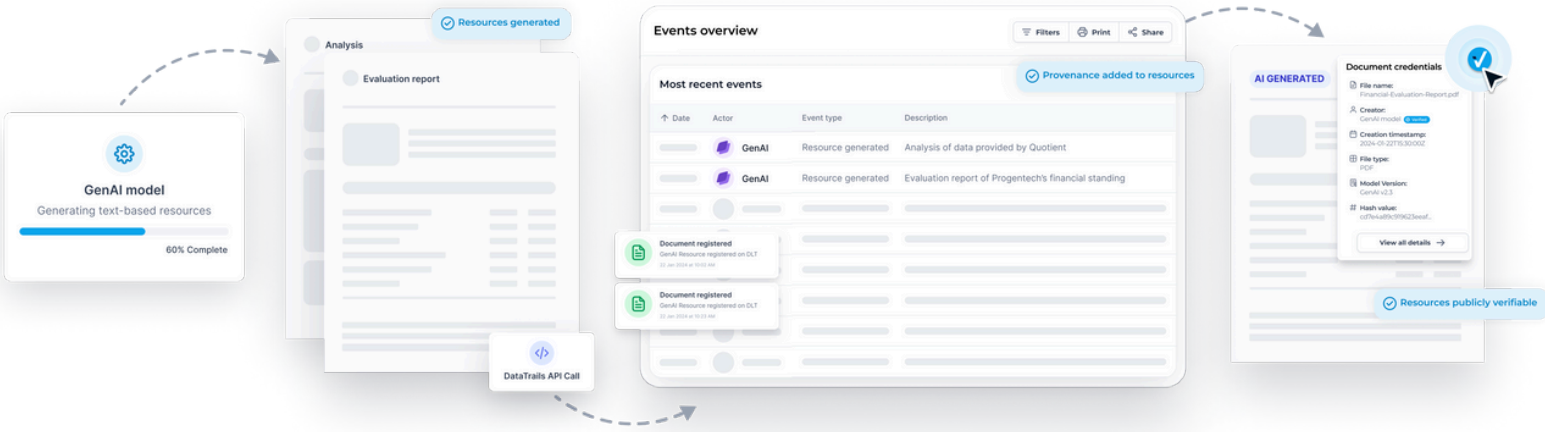
04 OPERATING RESPONSIBLY WITH DATATRAILS

DataTrails revolutionizes data integrity and transparency by enabling control over data flows, validating data sources, and constructing mutually accountable records. This approach not only meets the immediate need for secure and trusted data exchange but also lays the groundwork for AI systems to automate sensitive workflows confidently.

The DataTrails platform can integrate with and journal AI pipelines, enabling transparency of every aspect of the AI lifecycle, from code development and data training to operational deployment. Its flexible provenance metadata format allows a single trail of evidence combining software and structured data evidence with more subjective evidence such as regulatory approvals to create a golden thread of evidence that underpins fitness-for-use claims.

DataTrails' patented immutable audit log system underwrites responsible AI-driven decision-making, ensuring resilience, explainability, and evidence of regulatory compliance. Its fine-grained Access Policy engine enables transparency in the real world, by making sure that all stakeholders in an authorized group have complete and untamperable view of the provenance record whilst ensuring none of that information goes any further.

The robust provenance and metadata management capabilities of DataTrails instill confidence in AI adoption by promoting transparency, integrity, and accountability for continuous assurance of fitness-for-use. With its standards-based approach, DataTrails ensures that AI systems align with ethical guidelines and legal frameworks, building trust with users and society.



DataTrails Explainable Trust for AI features bring trustworthy data automation to AI, providing confidence to workflows and reducing the pain in tracking, tracing, and verifying the lineage of fast-changing AI models, datasets, and connected systems.

PROVENANCE & LINEAGE TRACKING

At its core, DataTrails is a provenance management platform. It excels at capturing and storing detailed lineage information for digital assets, such as data sets, AI models, and processes. This lineage tracking enables users to trace the origin and transformation of data, providing essential context for the inputs and outputs of AI decisions.

DATA INTEGRITY & ACCOUNTABILITY

The immutable audit trail provided by DataTrails ensures data integrity and attribution over the long term. Even as stakeholders and applications change, organizations can track provenance of data and models, and hold stakeholders accountable for decisions made with the confidence that comes from a durable and cryptographically protected evidence base.

TRANSPARENCY & EXPLAINABILITY

By leveraging the lineage information, DataTrails enables explainable AI. Users gain insights into how AI decisions are derived, ensuring transparency in AI outcomes. Understanding the data, algorithms, and logic behind AI-driven choices fosters trust between users and the technology.

COLLABORATIVE GOVERNANCE

Responsible AI requires collaborative governance. DataTrails facilitates collaboration between data scientists, subject matter experts, and regulatory bodies by providing all parties reliable access to all relevant trust data a simple API call to ask and answer complex questions like:

Should I trust this data right now?
The answer comes with a detailed defensible reason, inspectable by stakeholders during or after the event.

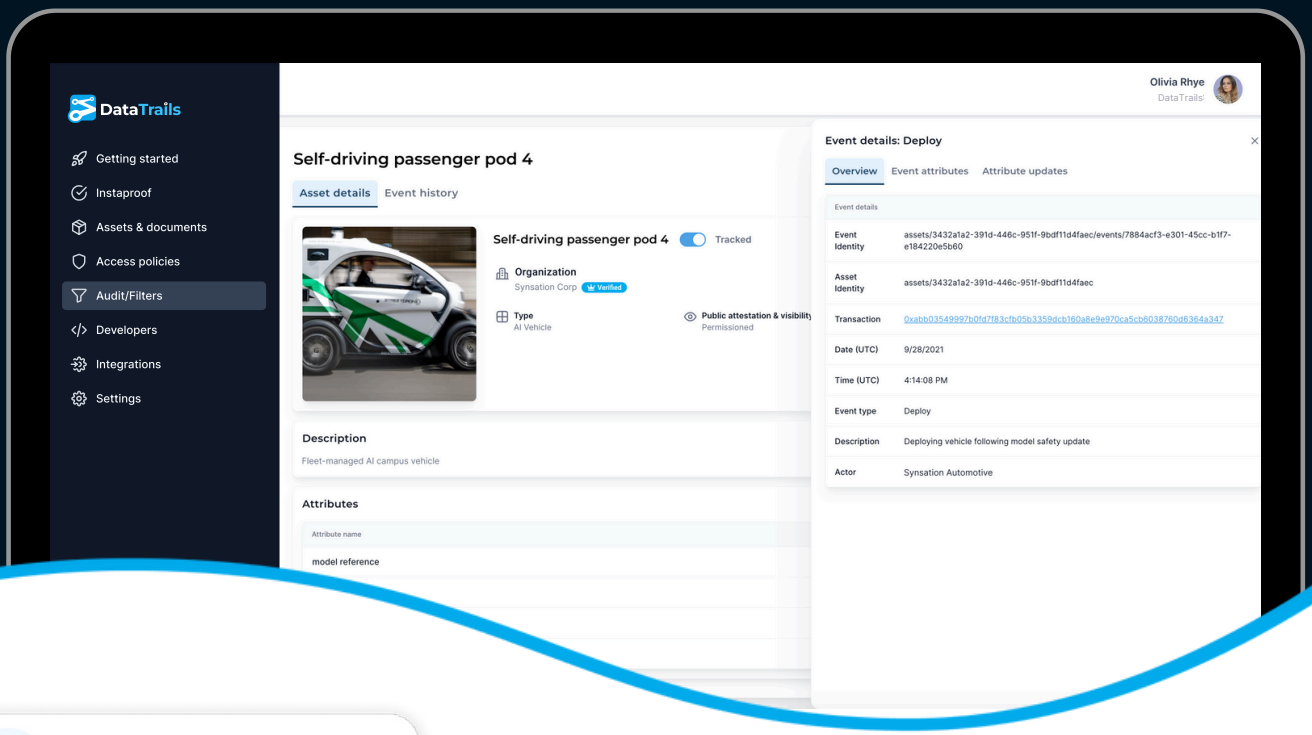
Why did I trust this model back then?
The answer comes with a detailed defensible reason, including unshreddable, unforgeable proof of who knew what when.

Am I getting the kind of updates I expect?
The transparency and collaboration features of DataTrails make it much easier to assess whole-system process anomalies, such as too many or too few events, or data points that greatly differ from the norm.

RETHINKING TRUST:

THE DATATRAILS PLATFORM

05



API-First Platform

Tracked assets
450

Untracked assets
50

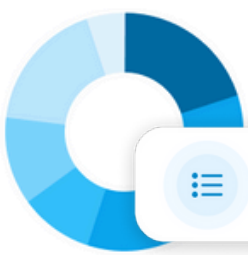
Non-compliant assets
10

Asset views in last 24 hours
All assets
2,000
100% vs last month

Data Provenance

Number of events

Events per asset type



- Asset type 1
- Asset type 1
- Asset type 3
- Asset type 4

Distributed Ledger Technology

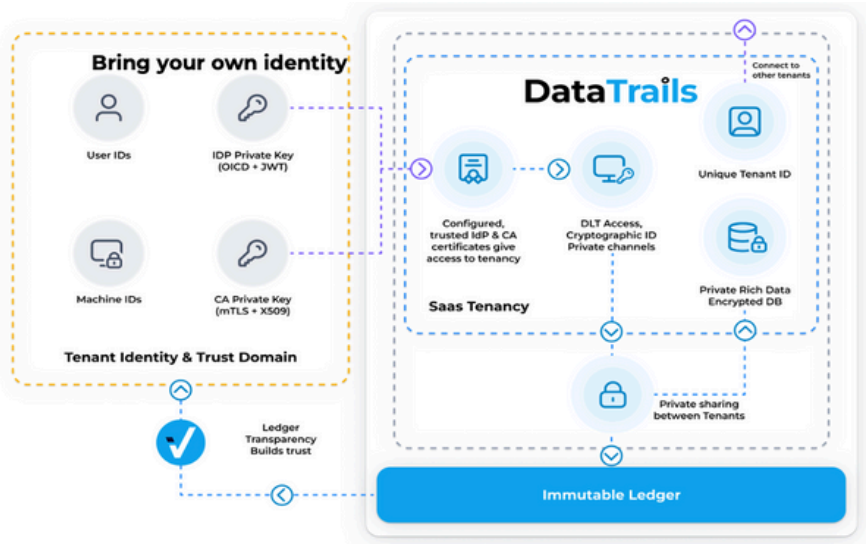
THE DATATRAILS PLATFORM

DataTrails is an API-first platform that offers a robust solution to enhance and protect trust in digital content and evolving supply chains. By separating data from provenance metadata, enforcing rules through cryptography, and leveraging distributed ledger technology, DataTrails empowers all supply chain partners to collaborate effectively, ensure accountability and transparency, and establish a shared base of evidence on which to make trustworthy decisions.

DataTrails enables instant data authentication by capturing and maintaining provenance metadata while enforcing sharing and visibility rules through cryptography and intuitive attribute-based access controls.

These powerful transparency and accountability benefits arise from a simple central concept: **Entities** make **statements** about **artifacts** that become **evidence** of ***Who Did What When***.

- **Entities:** Any authorized party involved in the AI supply chain
- **Artifacts:** Physical or digital items produced as part of supply chain processes that describe or contain the results of supply chain processes such as documents, receipts, logs, bills of materials, certificates, images, or plain data.
- **Statements:** Information about processes and artifacts that help stakeholders to make better informed decisions and build reputation a nd trust in supply chains .
- **Evidence:** Relevant information about the origins and handling of data, models and processes that is made trustworthy and reliable through protection from misattribution, back-dating, or shredding.



No special DLT skills required
Use existing apps & IAM
Can be as simple as a drag-and-drop

Fine-grained permissions
Community-level transparency

Scale options for any business
Don't lose trust data from small players

To ensure trust in these data sources, DataTrails includes several elements of verifiable provenance:

IDENTITY VERIFICATION

First, identity verification is a critical step that requires entities to be identifiable and verifiable, substantiating the claims or evidence they provide. This level of transparency strengthens the trust in the entity and its data.

AUDITABLE TRAILS

Data authenticity is confirmed by enabling auditable trails of data origin and evolution, ensuring that the data genuinely comes from the declared source and is the correct and latest version published by the source.

IMMUTABILITY

To further fortify this trust, DataTrails ensures data integrity by guaranteeing that the data from these sources remains non-repudiable and immutable, preventing equivocation of statements and protecting evidence against unauthorized alterations or tampering during transmission or storage.

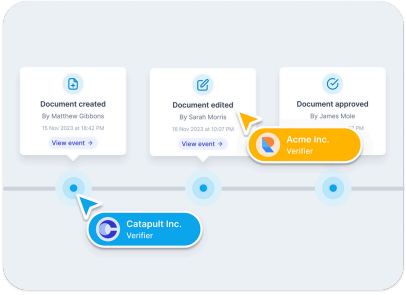
TAMPER-EVIDENCE

To maintain a consistent and reliable data environment, DataTrails employs a transparent and auditable ledger. This ledger can be cross-checked by anyone with standard OSS tooling, making the DataTrails system tamper-evident even against DataTrails itself.

Extensive data governance features are included in the platform that provide fine-grained access controls for sharing provenance both privately and publicly. Original data never needs to enter the DataTrails system so sensitive business secrets can remain in your existing sites and systems on-prem and in the cloud while still benefiting from globally verifiable provenance and integrity. By combining confidentiality, integrity and availability, DataTrails creates a robust basis for enhancing and maintaining confidence and trust in the data content and sources that underpin multi-party digital operations.

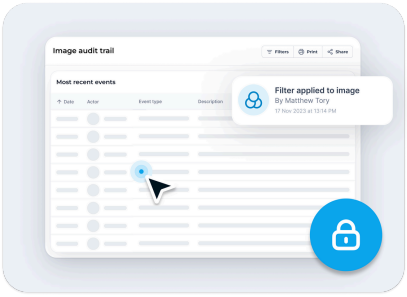
DATATRAILS BENEFITS

As AI technology and digital supply chains continue to evolve, the adoption of DataTrails will contribute to improve efficiency, resiliency, and transparency between businesses, partners, and customers, and safety in AI.



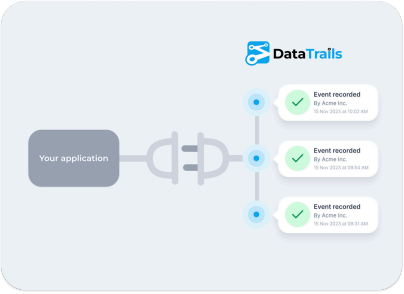
Transparency & accountability

DataTrails keeps track of the origins, history and authenticity of data and models in a tamper-evident ledger which makes it easy for anyone to detect fake data, prove ownership, and bring vital accountability to the digital world.



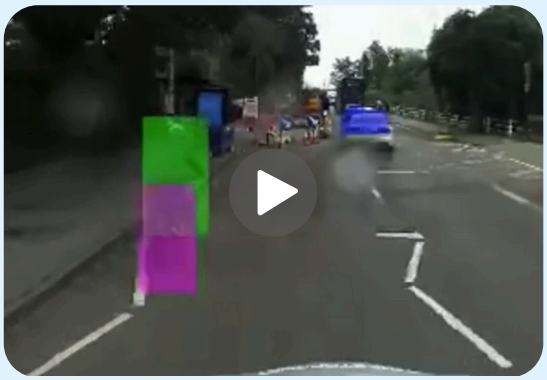
Tamper-evident ledger

Using innovative DLT and cybersecurity technologies DataTrails makes it practically impossible for anyone – even DataTrails – to forge, back-date, or shred data.



No-fuss, simple integration

Implementing provenance usually requires heavy investments and process change, but with DataTrails you can keep your current apps and working processes.



US DOD Explainable AI Demo
Tracing and auditing model development, deployment, and management

Click here to see the recorded demo.

Or use the link below:
https://www.youtube.com/watch?v=_IhWJPJ94_E

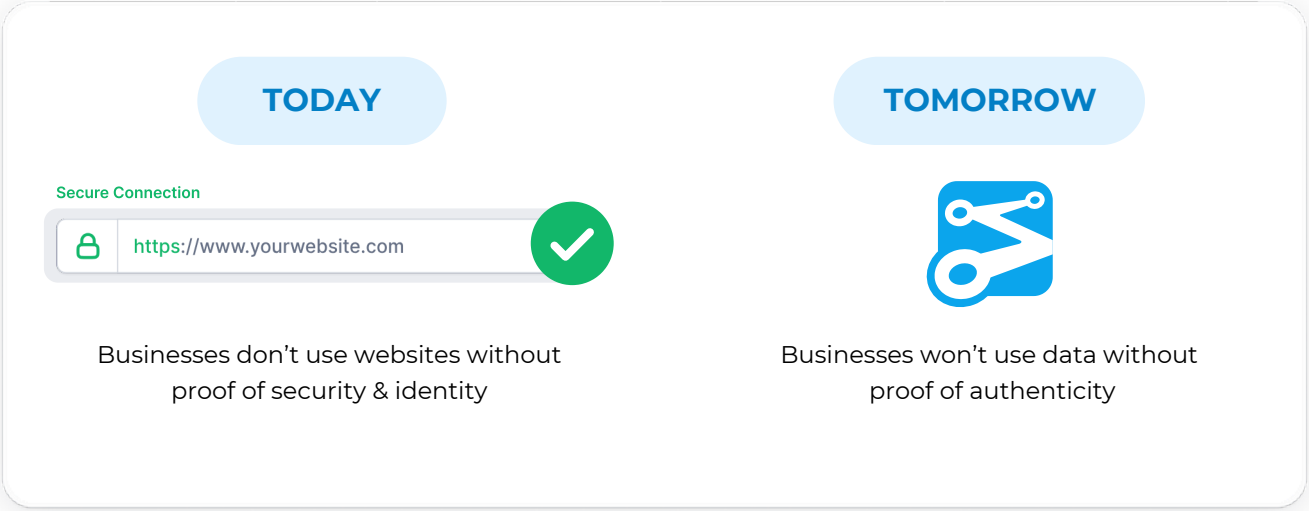
06 CONCLUSION

With provenance records in DataTrails, AI automation becomes more trustworthy and explainable in the real world. With appropriate operational transparency AI has a greater range of freedom as external stakeholders can verify responsible operation, confident that they will be able to investigate and swiftly remediate machine errors. As AI technologies continue to shape our world, the imperative for Trustworthy AI becomes increasingly critical.

With a shared platform for provenance and metadata management, stakeholders can collectively ensure that AI models and solutions can be proven fit-for-purpose not only on day 1, but verified and audited continuously as new knowledge and threats arise. Always know where your decisions came from.

DataTrails is a standards-based provenance management platform empowering organizations with explainable AI technology that tracks lineage, enables transparency, and ensures data integrity with a tamper-proof record of who-did-what-when in the commissioning of multi-party AI systems.

.Through collaborative governance and adherence to open & interoperable principles, DataTrails establishes a solid foundation for the responsible development and deployment of AI, fostering trust between users and the AI-driven technologies of the future.



07 ABOUT DATATRAILS

DataTrails brings long-term confidence to digital business models and data by maintaining a tamper-proof record of who did what when, no matter which stakeholders need to prove or verify responsible operations.

Underpinned by openly verifiable ledger technology, the provenance record created in the DataTrails secure cloud platform provides an immutable audit trail with appropriate levels of transparency to enable all stakeholders to assess the fitness-for-use of any data or technology entering their business processes.

Whether validating documents in real time or looking for simpler, better ways to meet audit requirements, DataTrails delivers the integrity, transparency and accountability required in today's fast-paced digital-first world.

To learn more visit: www.datatrails.ai

08

REFERENCES

- https://en.wikipedia.org/wiki/Explainable_artificial_intelligence
- https://deepai.org/publication/explainable-artificial-intelligence-xai-concepts-taxonomies-opportunities-and-challenges-toward-responsible-ai
- https://www.cutter.com/article/fiducia-ex-machina
- Partnership on AI, Responsible Practices for Synthetic Media, https://syntheticmedia.partnershiponai.org/
- https://www.mdpi.com/2076-3417/11/11/5088

To learn more visit: www.datatrails.ai