

White Paper

Trustworthy AI/ML for Patient Analytics and Research

AI-secure, privacy-first, with continuous monitoring and oversight



Table of contents

Responsible innovation in patient analytics and research	2
Heightened care in AI/ML	2
Adopting a principled, AI-secure approach to AI/ML	2
Shifting baseline for AI-Secure AI/ML	3
Bridging AI and data protection with federated modeling	3
Collection limitation and data minimization	3
Use limitation and purpose specification	4
Security safeguards	4
Accountability and oversight	4
Openness and transparency	4
Federated learning for AI/ML	5
Understanding the data journey for federated learning	6
Source ingestion: pseudonymization and segregation	6
Horizontal federated learning: generating synthetic trends	6
Vertical federated learning: AI-secure AI/ML	7
Safe outputs	7
Beyond de-identification: managing reconstruction risk	8
Leveraging synthetic trends for AI/ML	8
Managing reconstruction risk	9
AI governance and privacy operations (AI PrivOps): an integrated governance function	10
Continuous monitoring of AI PrivOps metrics	10
Oversight without exposure	11
Human-in-the-loop for accountability	11
Ethics Board for patient analytics and research	12
Conclusion	13
Acknowledgment	13



This paper outlines our approach to Artificial Intelligence (AI) and Machine Learning (ML) that withstands today's threat landscape and serves as a blueprint for sustainable innovation. It's how we raise the bar for defensible AI in healthcare applications and beyond, representing a shift from traditional data practices to AI-security as a design feature.

Responsible innovation in patient analytics and research

Life sciences are being transformed by Artificial Intelligence (AI) and Machine Learning (ML). But with that transformation comes a critical question: how do we unlock value from sensitive health data without undermining trust, transparency, or control? Traditional safeguards are no longer enough in an era of AI/ML, where subtle patterns can be used — or misused — in unanticipated ways. The stakes are especially high in healthcare, where data utility must be balanced with rigorous protection.

Robust de-identification methods, which remove identifying elements, can be used but the industry lacks widespread adoption of standardized practices. This absence of fixed standards provides space to explore forward-looking approaches, especially in light of emerging AI/ML threats that will need to be addressed. As AI/ML and other developing technologies reshape the landscape, more sophisticated strategies are needed to balance AI/ML and data protection with responsible use.

Heightened care in AI/ML

This whitepaper introduces a novel privacy-first and AI-secure architecture for defensible AI developed by IQVIA. In response to AI and data protection concerns, the platform combines synthetic data abstractions, federated learning, and integrated AI Governance and Privacy Operations (AI PrivOps) monitoring to enable safe, effective AI/ML without compromising confidentiality.

The solution enforces AI and data protection through architectural features such as input transformation, non-reversibility, and latent space modeling. Aligned with global standards such as ISO/IEC 42001 AI Management System and frameworks by the U.S. National Institute of Standards and Technology (NIST), this system ensures continuous oversight, minimizes AI/ML risks, and promotes defensible AI. This approach meets evolving regulatory and organization expectations, and it sets a new benchmark for ethical healthcare AI innovation.

Adopting a principled, AI-secure approach to AI/ML

Envision a future where AI/ML models for health and wellness applications are proactively engineered with resilience and security at every layer. Sensitive data remains protected, systemic vulnerabilities and risks are managed before they surface, and insights are extracted and utilized without exposure. This is the new frontier of AI and data protection, where the architecture is purpose built for robustness, trust, and availability without compromising analytical power.

This white paper introduces a novel, principled approach that puts AI security at the center of the system architecture. Powered by the ***IQVIA Synthetic Trends Engine***, our approach is grounded on three foundational pillars: synthetic data abstraction, federated learning architecture, and integrated AI governance and privacy operations.



Synthetic data abstraction: Traditional models rely on raw data, increasing the surface area for risk. Instead of relying on raw data, our approach transforms high-dimensional signals into non-reversible trend vectors using AI-secure dimensionality reduction techniques. Synthetic trends capture useful patterns to maintain analytical utility while minimizing downstream reconstruction risk by design, approaching near-zero exposure. This enables inferential bridging for analytics across isolated datasets.



Federated learning architecture: Rather than aggregating data into a central repository, our system employs a federated architecture in which source data are segregated within secure environments. Raw data never leaves its origin, and only synthetic trends, which are themselves AI-secure, are combined for modeling. Decentralized computation ensures that data sovereignty is respected, significantly reducing the risk of exposure and unauthorized access, while still allowing for collaborative analytics.



AI governance and privacy Operations: Every step of the data flow is governed by strict policies and technically enforced oversight mechanisms that are tracked and managed accordingly, including continuous monitoring and auditable logs for end-to-end traceability. Segregated environments, role-based access controls, and robust audit trails ensure that data is used only for its intended purpose, in alignment with enterprise-level governance and risk strategies.

This approach is a response to today's threat landscape — including risks such as AI model inversion, data reconstruction, linkage attacks, and the misuse of data across distributed systems — and serves as a blueprint for sustainable innovation. It's how we raise the bar for what defensible AI looks like in healthcare applications and beyond, representing a shift from traditional data practices to a new paradigm where AI-security is a design feature.

Shifting baseline for AI-Secure AI/ML

AI has transformed the landscape of health and wellness industries, enabling innovative solutions that enhance patient outcomes, streamline clinical research, and improve patient engagement in health-focused interventions. As the appetite for data-driven personalization and combined intelligence continues to grow for health and wellness use cases, the rise of sophisticated analytics, cross-platform identifiers, and probabilistic modeling has raised urgent questions about AI and data protection.

Across the globe, AI-driven applications are expected to navigate an increasingly complex regulatory environment. AI and data protection are converging priorities, with new and evolving laws, regulations, and policy guidance emphasizing risk-based AI governance, fairness, transparency, and accountability. Both global frameworks and national laws are shaping expectations for how health data is collected, analyzed, and used — particularly where AI introduces novel risks such as data leakage, inference, or misuse.

This regulatory evolution underscores the need for AI systems that are secure by design, aligned with organizational objectives, and capable of adapting to shifting oversight requirements. Patient analytics and research, especially the secondary use of health information, is under increasing scrutiny from data protection authorities due to the use of sensitive health-related data sensitive and because the outputs will inform patient care and outcomes. Concerns may include fairness, transparency, and the risk of harm when personal health insights are used irresponsibly or without adequate safeguards.

IQVIA's approach is designed to address those concerns directly. By using synthetic trends, federated modeling, and continuous monitoring, we minimize data exposure while enabling high-quality patient insights. An ethics board can provide an additional layer of accountability, ensuring that we align with regulatory expectations and set a new standard for responsible and trustworthy use of AI in healthcare.

Bridging AI and data protection with federated modeling

IQVIA's federated modeling approach translates abstract principles of AI-security and privacy protection into concrete system behaviors. The platform operationalizes them through embedded architectural controls and data-handling strategies. The core engineering concepts of data protection in our approach — input transformation, non-reversibility, and latent space modeling — are aligned by design with AI-security and privacy protection, ensuring both technical performance and principled data use.

Collection limitation and data minimization

Through input transformation, IQVIA ensures that only essential attributes are retained. Features with high reconstruction risk or low modeling utility are excluded during preprocessing. Data is pseudonymized externally and abstracted early, reducing the need for collection of

detailed raw inputs. The use of synthetic trends derived from group-level statistical abstractions reflects this principle in practice: models are built using only what is necessary — and nothing more.

Use limitation and purpose specification

Non-reversibility reinforces the boundary between raw input and model use. The system is designed so that data collected for AI/ML can only be used for approved purposes. Outputs such as cohort scores or patient segments are labeled with metadata that restricts downstream use to a specific context and time frame. By making outputs unusable for anything but the intended application, non-reversibility enables robust enforcement of purpose limitation.

Security safeguards

Latent space modeling adds a meaningful security layer by removing any semantic traceability to original data. The use of abstracted, non-human-readable embeddings to create synthetic trends prevents even authorized personnel from reconstructing sensitive traits or behaviors. Combined with traditional access controls, encrypted environments, and audit trails, this approach embodies both technical and organizational security safeguards.

Accountability and oversight

Each of the three architectural strategies is embedded within a broader governance model that supports continuous oversight. Input transformation pipelines are versioned and logged. Risk assessments tied to non-reversibility thresholds are stored and periodically audited. Latent space models are reviewed for drift and abuse potential. These controls establish traceability and institutional accountability. An ethics board can also have an oversight and monitoring role to ensure we remain aligned with regulatory expectations and continuously update ethical guardrails.

Openness and transparency

IQVIA is proactively publishing documentation on its modeling pipeline, including how data is transformed, abstracted, and safeguarded. Clients and partners are given access to non-sensitive model lineage reports and can request high-level explanations of model purpose and boundaries. These efforts support informed trust without revealing protected IP or compromising privacy protections.

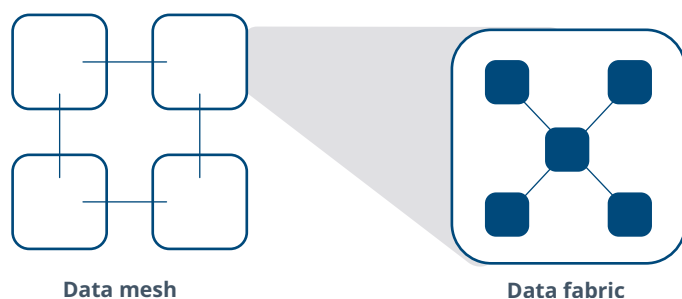
Together, these mappings demonstrate that IQVIA's technical strategy is aligned with privacy norms and operationalizes them. This federated modeling approach is privacy-aware and AI-secure, by design.



Federated learning for AI/ML

In IQVIA's federated modeling approach, source data is decentralized, identifiers are masked, and only synthetic abstractions are used for modeling. To strike a delicate balance between centralized control and decentralized innovation, we use the architecture of a data fabric aligned to a broader data mesh strategy, which we call a **secure health fabric**, as shown in Figure 1.

Figure 1. Data fabric within a data mesh strategy



The secure health fabric configuration is designed around the creation of segregated workspaces that enable individual teams to function autonomously within a cohesive technological framework. This setup supports the development of customized AI applications tailored to specific needs while maintaining the security and governance standards of the overall architecture.

Each workspace produces its own data products, which are developed, tested, and validated independently of others, thereby enhancing security and specialization.

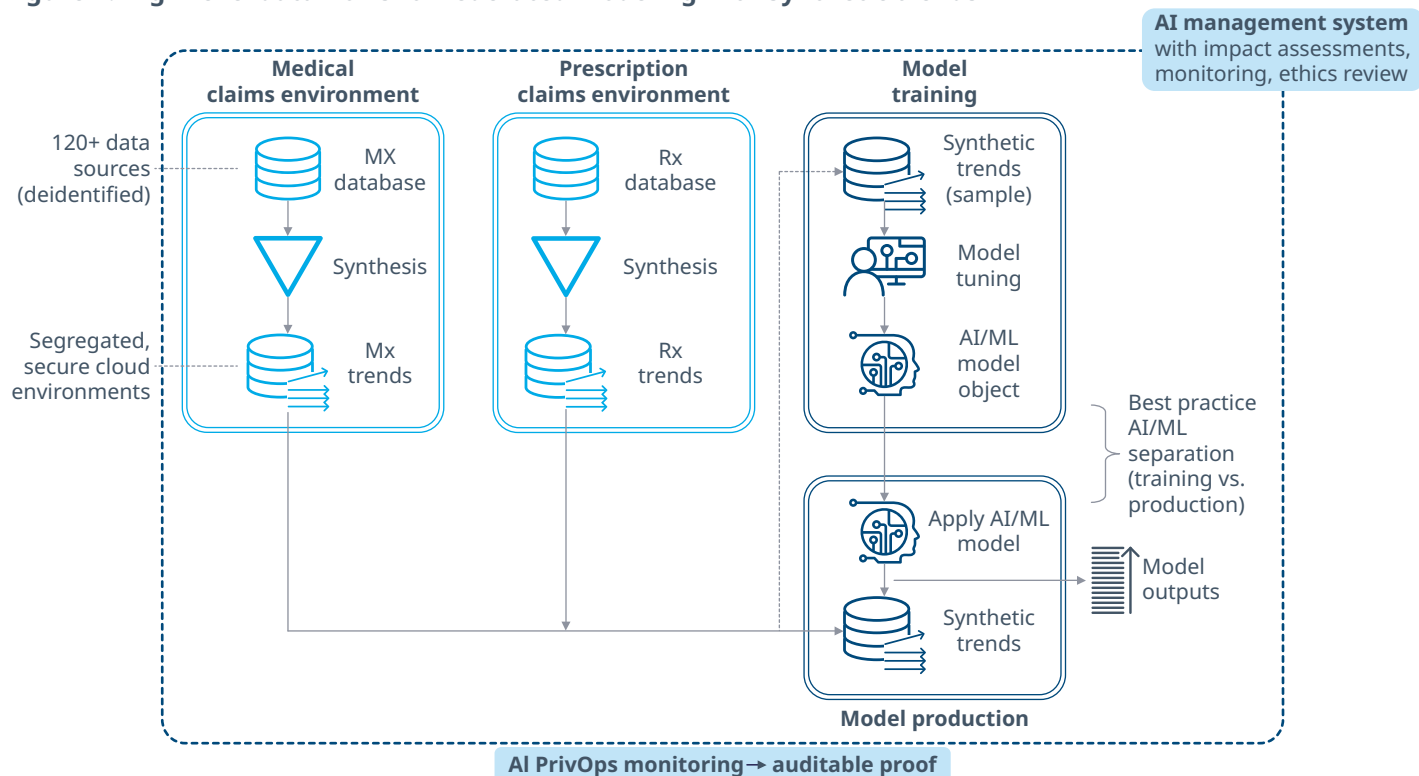
The independent workspaces within the secure health fabric are governed by robust protocols that align to the highest data protection and AI security standards. Automation through AI agents enable real-time management and enforcement of these protocols, while human-in-the-loop checkpoints and operational monitoring systems ensure accountability and trust in how AI agents are used to manage AI-enabled workflows. More than a data management system, the secure health fabric serves as a foundational architecture for securely and efficiently deploying, managing, and scaling AI applications in healthcare.



Understanding the data journey for federated learning

Figure 2 illustrates the overall data journey through the system. There are four key stages within the federated learning process: source ingestion, horizontal federated learning, vertical federated learning, and safe outputs.

Figure 2. High-level data flows for federated modeling with synthetic trends



Source ingestion: pseudonymization and segregation

At the entry points, each data stream is pseudonymized before ingestion. Pseudonymization replaces sensitive identifiers with unique, random pseudonyms that have no intrinsic meaning and are internal to the federated modeling process. To maximize separation of duties and enhance trust, this function can be handled by a neutral Third-Party Provider (TTP), ensuring that no internal team or downstream process ever has access to unique identifiers. Depending on the design, pseudonymization can support deterministic linkage (the same input always maps to the same output, enabling consistent matching across datasets) or probabilistic linkage (which allows matching with a configurable degree of fuzziness to accommodate natural data variability).

Upstream sourcing requirements (contractual, privacy, or governance enforced) associated with each data stream are considered at this stage through additional

de-identifying transformations, to align with regulatory and organizational expectations around data use and sharing. Following these processes, the individual data streams are ingested and maintained in segregated environments with strong data separation and access controls in place.

Horizontal federated learning: generating synthetic trends

In segregated environments, each data stream undergoes processing in isolation to convert raw features into synthetic trends, which are abstracted signals that capture group-level patterns rather than individual-level details. This is accomplished through horizontal federated learning whereby we align data by columns across different datasets. We use techniques such as autoencoding and matrix factorization methods to compute statistics, like means, from each dataset and then pool these to derive global insights.

These methods are configured to retain important signals but discard idiosyncratic traces that could re-construct original data. To further strengthen protections, we can apply differential privacy during this transformation process. Differential privacy introduces controlled, mathematically calibrated noise into data or summary statistics. Differential privacy ensures that no single individual's data significantly impacts the outcome, making it practically impossible to infer sensitive details or identify participants based on embeddings.

Pseudonymized health features are transformed into behavioral trend vectors (representing important patterns and relationships between features). These vectors retain insights from health features without preserving individual records. In segregated environments, different sources of health information (e.g., medical or prescription claims) are independently transformed into synthetic trend vectors that retain meaningful health insights (e.g., synthetic medical trends or synthetic prescription trends).

Vertical federated learning: AI-secure AI/ML

Synthetic trends from segregated bridge environments are routed into the modeling environment; the upfront pseudonymization process enables vertical federated learning to align data by rows, matching records across datasets without ever combining them directly, instead combining trends from each dataset to create comprehensive individual profiles. The combined trends undergo a robust reconstruction risk assessment before being made available for modeling purposes.

To further minimize risk, a sampling-first approach to model training is adopted. Only a subset of the combined synthetic trend data is used to train models initially. This sample is statistically representative but constrained enough to reduce exposure in the event of model overfitting or data leakage. Once models are trained and meet AI PrivOps constraints, they are deployed across the full combined synthetic trend dataset during inference.

This separation between model training and inference is deliberate: the model training is highly restrictive, supporting only exploratory analysis and feature refinement under tight governance and automation. Inference processes are distinct, auditable, and stripped of access to feature generation logic, ensuring that post-training application of models cannot be reverse-engineered into insights about individual or group behavior.

This two-stage model lifecycle (sampled training followed by broad inference) further reinforces the privacy perimeter, limiting who can see the trends data and how deeply the system “learns” from it.

Safe outputs

Before the results of the modeling process are used for decision-making or shared externally, a rigorous output validation process is conducted. This involves ensuring that outputs comply with data protection and governance expectations by applying data aggregation, noise injection, and secure encryption protocols.

Modeled outputs may take various forms, including risk stratification scores, predictive analytics, or cohort classifications. These outputs undergo a validation process to verify that they meet data protection standards, for example aggregation thresholds. Any external sharing is managed through secure, non-reversible pseudonymization methods, ensuring that the receiving systems cannot trace back to individual data points. Metadata controls are applied to restrict downstream use, enforce expiration, and track data lineage.

By separating raw data from model environments, abstracting signals into synthetic trends, and enforcing tight operational boundaries, the overall system transforms AI and data protection from a theoretical concept into a functional constraint. Each stage ensures that data use remains purpose-bound and technically non-reversible, aligning with IQVIA's commitment to responsible innovation in regulated machine learning contexts.

Beyond de-identification: managing reconstruction risk

Federated learning architectures are designed to minimize risk by decentralizing data. However, when using raw or even de-identified features, modern AI systems could draw connections and surface features that humans might miss. When applied to high-dimensional datasets, especially ones that combine various sources of health information, even relatively sparse information can yield sensitive inferences. This raises the possibility that models may inadvertently disclose sensitive information or infer sensitive health attributes, undermining privacy.

Adversarial actors can exploit poorly abstracted model inputs or outputs through inference, reconstruction, or correlation attacks. So, while traditional de-identification techniques such as pseudonymization, generalization, or suppression reduce identifiability, our concern is mitigating the ability of systems to infer attributes under adversarial conditions.

Leveraging synthetic trends for AI/ML

In IQVIA's principled approach, synthetic trends serve as an AI-secure intermediary, abstracting meaningful group-level patterns from raw data through feature aggregation, dimensionality reduction, and controlled noise injection. This method enhances AI and data protection and optimizes analytical utility by maintaining critical signal integrity without exposing individual records, as shown in Table 1.

Table 1: Synthetic trends dataset for AI/ML

ID	OUTCOME	TREND 1	TREND 2	TREND 3	TREND 4	TREND 5
1	0	0.159	1.629	0.138	2.5	0.89
2	1	-1.712	-0.153	-5.8	-0.15	-0.15
3	1	1.090	0.617	11.72	0.83	0.617
4	0	1.771	-1.277	7.63	-0.28	-0.45
5	0	-1.308	-0.817	-4.6	1.14	-0.31

Our approach introduces safeguards rooted in both design and AI-security engineering to ensure that minimizing reconstruction risks and preventing misuse are continuous and enforceable throughout the entire data lifecycle. This is achieved through three core concepts:



Input transformation

Raw signals are converted into standardized and privacy-abstracted representations before any modeling occurs. This includes operations like normalization, signal aggregation, and feature encoding. The outcome is the generation of synthetic trends — insights that preserve patterns without traceability back to original data. This upfront abstraction ensures that no raw data interacts directly with model logic.



Non-reversibility

To prevent any reconstruction of source data, non-reversibility is enforced through dimensionality reduction, noise injection, and mutual information minimization. These techniques create a technical barrier that prevents models from inferring original signals even when auxiliary datasets are available. By mathematically disrupting traceable paths back to individual records, this principle guarantees statistical implausibility of re-construction.



Latent space modeling

Model training and inference are conducted within abstract, non-semantic feature spaces, decoupling learned behavior from recognizable inputs. This approach optimizes learning while introducing an additional layer of privacy, as models operate solely on latent representations that lack direct ties to raw data. Even under adversarial conditions, this abstraction mitigates risks of unintended inference or data leakage.

This structured approach redefines AI and data protection as a proactive design principle, moving beyond traditional methods to embed AI-security directly within the modeling architecture. By prioritizing secure abstractions and non-reversible transformations, we ensure this system is resilient against modern threats while optimizing analytical utility without sacrificing data protection.

Managing reconstruction risk

Because original values in the health data are transformed into an embedding space, a row-level record would need to be reconstructed before any meaningful data about an individual could be misused. We define this possibility, the chance of inferring original values from their transformed representations, as *reconstruction risk*, a precursor to data misuse. Our implementation embeds continuous measurement, quantification, and monitoring across the entire data lifecycle, ensuring that synthetic trends are robust against adversarial attempts and statistical attacks.

Reconstruction risk quantifies the privacy and security exposure of the transformed embedding space. It accounts for the diminishing contribution of higher order embedding dimensions, as well as the effects of distortion and noise injection. While retaining more dimensions may preserve more detail, we deliberately limit this by applying dimensionality reduction. This introduces distortion that acts as a safeguard, making it significantly harder to reconstruct original values and reducing the likelihood of reversal.

In practice, we use this method to reduce large and complex datasets into compact forms that improve computational efficiency while producing secure synthetic trends. In these cases, reconstruction risk is orders of magnitude below practical thresholds, supporting an elevated level of anonymization. This approach offers a practical way to assess AI and data protection through the lens of adversarial reconstruction difficulty.

AI governance and privacy operations (AI PrivOps): an integrated governance function

As federated models grow in complexity and scale, ongoing oversight is needed to ensure that controls operate as intended and remain aligned with policy, law, and public expectations. IQVIA addresses this need through an integrated operational framework for AI Governance and Privacy Operations (AI PrivOps).

AI PrivOps serves as a cross-functional, continuous AI management layer that translates policy into enforcement, monitors live modeling activities, and provides a mechanism for escalation, remediation, and auditability. It enables privacy assurance as a living process woven throughout the model lifecycle.

While data protection begins with architectural choices like federated learning and input transformation, ongoing monitoring and accountability will uphold

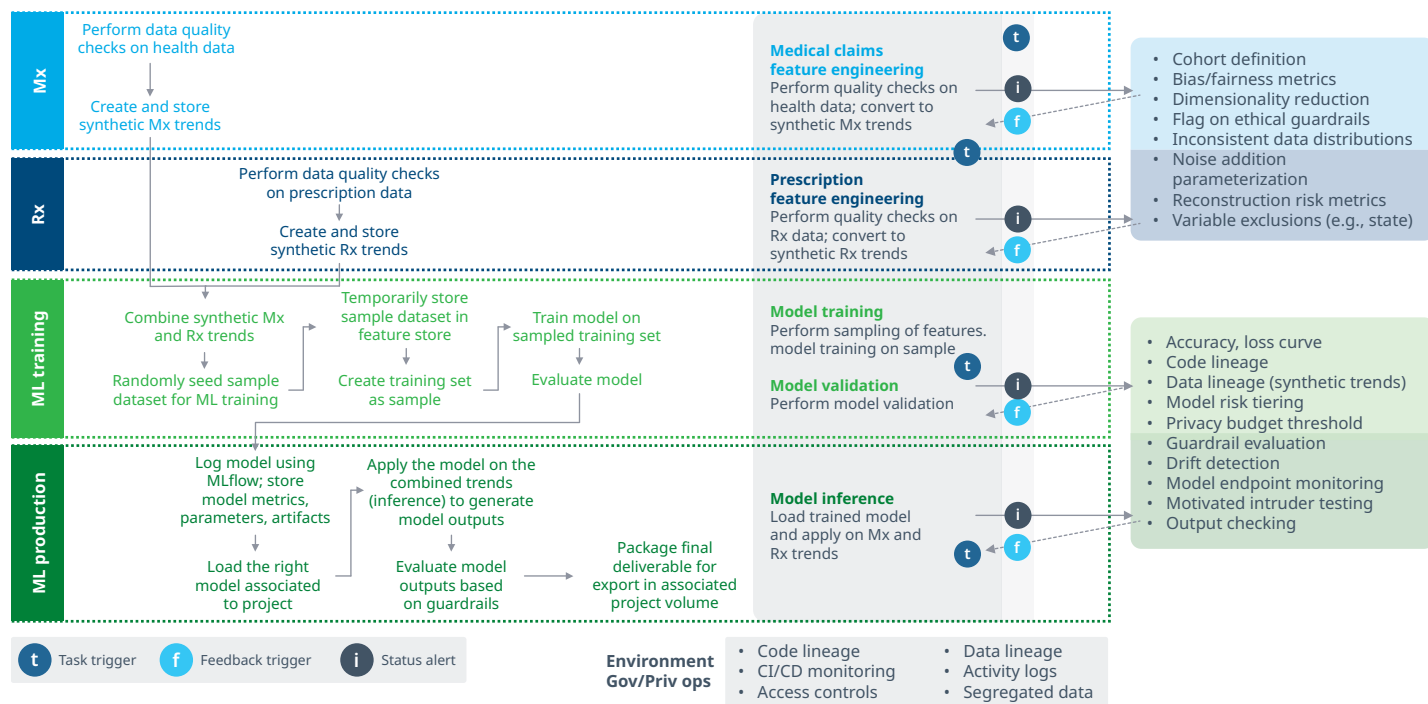
trust and maintain alignment over time. AI PrivOps is the operational backbone that turns privacy-by-design commitments into defensible outcomes.

AI PrivOps is a monitoring and assurance layer that overlays the federated modeling stack. It ensures continuous tracking of privacy risks, enables oversight without exposing sensitive inputs, and provides structured workflows for intervention, review, and improvement, as shown in Figure 3. It supports adaptive governance helping organizations respond to model drift, feature updates, and evolving regulatory expectations.

Continuous monitoring of AI PrivOps metrics

AI PrivOps enables continuous monitoring by embedding metrics at key points across the federated learning workflows, from data ingestion to synthetic trends generation, model training, and deployment. Each stage is instrumented to capture specific classes of metrics that reflect AI and privacy risk, model behavior, and environmental integrity.

Figure 3: AI governance and privacy operations (AI PrivOps)



At the source ingestion and bridging stages, key metrics include outcome definition fidelity, fairness and bias detection, dimensionality compression validation, reconstruction risk measurement, and flagging for variable exclusions (such as geographic or sensitive attribute filters). Noise injection parameters and checks for inconsistent or anomalous data distributions are also logged to detect preprocessing errors or skewed datasets that could compromise AI and data protection.

During model training and inference, AI PrivOps tracks accuracy and loss curves, monitors drift and checks for abnormal fluctuations in model behavior through endpoint metrics. Variability (where applicable) is evaluated post-training, and motivated intruder testing may simulate adversarial scenarios. These techniques identify if outputs could leak sensitive patterns or invite misuse.

Beyond the workflow layer, AI PrivOps also continuously assesses environment-level controls through logs that monitor Continuous Integration and Continuous Delivery (CI/CD) activity, access controls, infrastructure segregation, and code lineage integrity. Together, these layers of real-time and historical measures form a layer of privacy assurance that allows IQVIA to detect issues before they lead to downstream consequences.

Oversight without exposure

While AI PrivOps maintains visibility into all stages of the data lifecycle, it does so without ever needing direct access to raw data or modeling internals. This is made possible by its architectural placement: AI PrivOps operates adjacent to, and independent from, the federated learning pipeline. It runs on segregated infrastructure that is securely connected to the broader modeling environment, allowing it to tap into key status signals and metadata artifacts — such as model lineage, configuration snapshots, and synthetic data summaries — without becoming a bottleneck or point of risk exposure.

This separation ensures that AI PrivOps can conduct effective, real-time oversight without interrupting core model workflows or violating the data minimization principles it aims to uphold. It avoids becoming a perpetual “gating mechanism” by focusing on signal intelligence and policy enforcement through measurement and monitoring rather than direct intervention. Key AI and privacy metrics are rendered onto governance dashboards designed for non-intrusive review, allowing governance teams to assess alignment and effectiveness of controls, investigate anomalies, and verify proper handling without ever crossing into protected data space.

Human-in-the-loop for accountability

AI PrivOps incorporates human-in-the-loop processes that ensure governance is automated and context-aware, policy-aligned. A cross-functional governance board composed of legal, privacy, technical, and ethical stakeholders can oversee critical decisions throughout the model lifecycle. This includes approving high-risk feature sets, adjudicating exceptions, and reviewing edge-case applications.

When anomalies are detected, such as elevated uniqueness in outputs or unauthorized model behavior, AI PrivOps initiates a formal incident response. The workflow includes automatic pausing, investigation, root cause diagnosis, and corrective action such as retraining, feature suppression, or policy escalation. All actions are logged and included in the process audit history.

These same governance structures ensure auditability and traceability. Decisions, transformation, risk score, and mitigation are documented from pseudonymization through to model activation and downstream use. This provides internal accountability and supports external defensibility in audits, partner reviews, or regulatory inquiries.

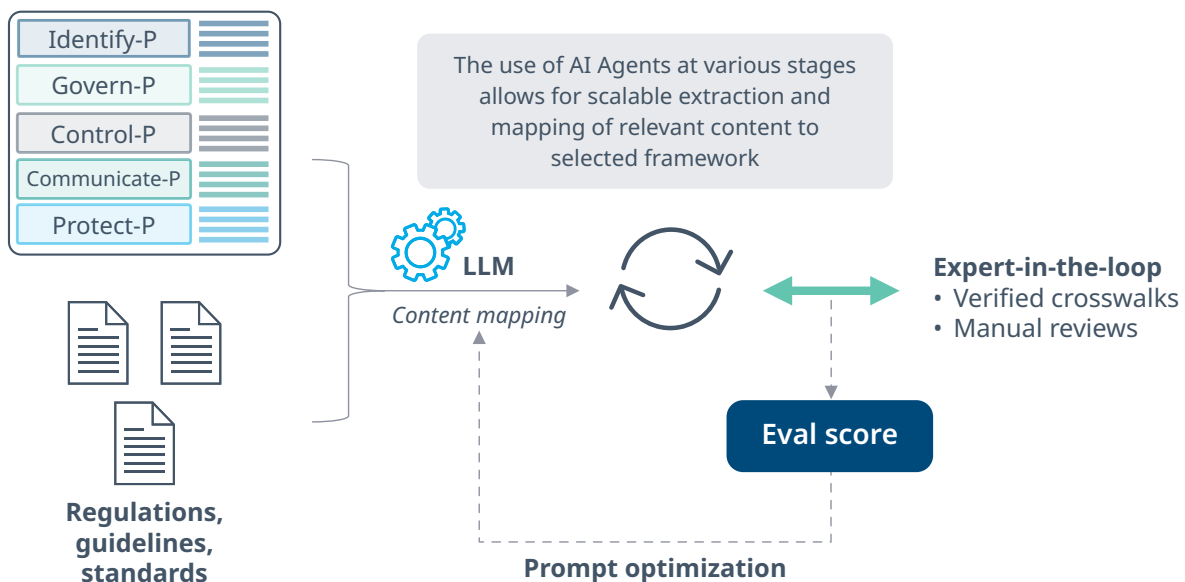
CONTROLS MAPPING FOR SCALABLE AI MANAGEMENT

To build a scalable AI management framework for federated learning, IQVIA employed a structured process known as controls mapping — an approach that aligns technical and procedural safeguards with global regulatory expectations and best practice frameworks for data and AI protection (including ISO/IEC standards). This process begins by identifying applicable legislative, sectoral, and jurisdictional requirements relevant to the data and use case at hand.

Controls from international frameworks provide a neutral, harmonized baseline from which to design practical and auditable privacy protections. Data protection frameworks, for example, can outline hundreds of potential actions across domains such as risk identification, protection, detection, and response. Rather than implementing every possible control, IQVIA applies a risk-based lens — tailoring safeguards based on the sensitivity of data types (e.g., demographic vs. lab) and specific modeling contexts.

This structured mapping process enables prioritization of requirements, ensures continuous alignment with evolving policy landscapes, and facilitates transparent traceability from principle to practice. Large Language Models (LLMs) and AI agents further enhance the efficiency and scale of this effort by rapidly identifying and connecting regulatory requirements to control frameworks — while human-in-the-loop reviews maintain context fidelity and subject matter accuracy. The result is a scalable, repeatable, and defensible mechanism for embedding privacy principles within enterprise AI operations.

Data protection framework



Ethics Board for patient analytics and research

Depending on the level of sensitivity, a cross-functional ethics board can be made responsible for overseeing and monitoring AI/ML efforts, ensuring activities remain aligned with regulatory expectations and continuously

update ethical guardrails. These guardrails would be communicated internally to support employee education and reinforce organizational integrity. The recommended structure of an ethics board would include expertise in data ethics, healthcare law, patient privacy, AI governance, and patient analytics.



Based on procedural guidelines and ethical guardrails, a governance intake form is completed by a business lead, which is designed for use as early as possible when considering the creation of a new analytical use case, and certainly prior to AI/ML. It captures the necessary detail for ethical risk classification while remaining checklist-driven and user-friendly. It acts as a gate in the AI/ML platform and creates auditable proof of responsible AI/ML.

Activities flagged for ethical assessment may require collaboration with the ethics board to clarify impact in case mitigations are recommended. If recommendations are deemed overly restrictive, the business unit should immediately discuss with the ethics board. The completed form is stored in an ethics intake repository and reviewed prior to AI/ML.

Conclusion

IQVIA's Synthetic Trends Engine represents a fundamental shift in how machine learning is applied to sensitive health data. This approach meets and exceeds current expectations for AI and data protection by embedding technical safeguards — like non-reversible transformation, federated processing, and latent space modeling — directly into the system architecture. By minimizing exposure risk while maintaining high modeling utility, the platform delivers measurable value without compromising patient trust.

Crucially, this is a continuous alignment and verification exercise with monitoring and oversight through AI Governance and Privacy Operations (AI PrivOps), real-

time risk monitoring, and governance checkpoints ensure that safeguards remain effective over time. The system is designed to adapt to evolving threats, organizational needs, and regulatory landscapes — turning privacy and AI security into operational features.

This model sets a new ethical benchmark for health-related patient analytics and research. By shifting from reactive privacy controls to proactive AI-secure design, it aligns with global best practices while addressing the unique risks of emerging AI threats. With embedded transparency, human-in-the-loop governance, and architectural discipline, this approach raises the bar for responsible AI — proving that innovation and accountability can go hand in hand.

Acknowledgment

IQVIA Applied AI Science designed and developed the IQVIA Synthetic Trends Engine. It is enabled by our secure health fabric and AI Governance and Privacy Operations (AI PrivOps) monitoring to produce auditable proof of continuous oversight and protection. The Synthetic Trends Engine can be used in a standalone data cleanroom or in a federated learning approach, for a variety of health and wellness applications that require robust implementation of privacy and security measures against emerging AI threats. IQVIA Applied AI Science is a leader in developing advanced AI methods and platforms, powered by Privacy Analytics for third party assessments and privacy operations monitoring.

The background of the slide is decorated with large, abstract, rounded shapes. A dark blue shape is on the left side, and a light blue shape is on the top right side.

CONTACT US
defensibleAI@iqvia.com
iqvia.com/contact