Federated Computing: The Missing Architecture Layer for AI-Driven Drug Discovery

Rhino
Federated Computing

CHALLENGE:

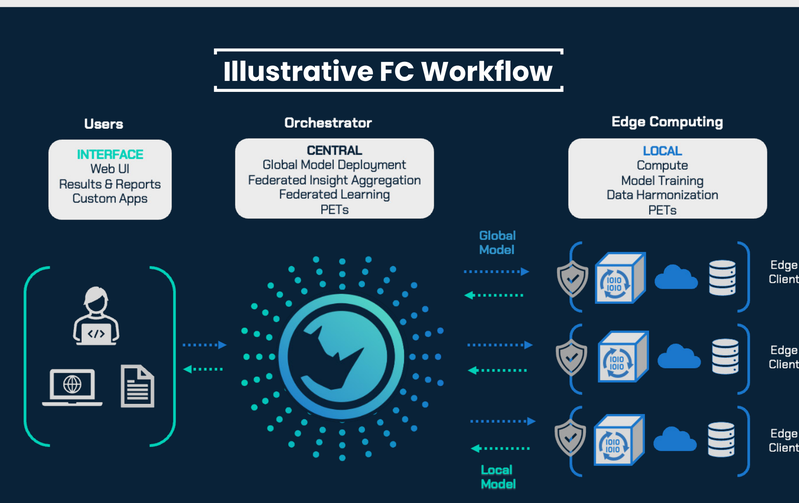
Including use cases, such as target identification and molecular design, advances in AI are accelerating drug discovery. A published analysis of AI-derived molecules found they have an 80-90% success rate in Phase 1.[1] However, the effectiveness of AI models and applications is dependent on access to vast amounts of sensitive, decentralized data that is protected by IP, privacy, and regulatory constraints - preventing its centralization, requiring resource-intensive legal and technical overhead, and ultimately resulting in AI models being trained on narrow or siloed datasets, lacking diversity and depth.

Introduction to Federated Computing

Bringing together Federated Learning (FL), edge computing, and privacy-preserving technologies (PETs), Federated Computing (FC) offers the missing architecture layer, which enables AI model training, fine-tuning and inferencing across sensitive datasets without transferring raw data. Models, algorithms and code are sent to the data. The compute happens where the data resides, at the edge; only aggregated results are shared to a centralized location.

Key points:
- Data never leaves the institution or site where data resides
- Models train and run where data resides
- Only governed outputs or model updates are shared
- Works across cloud, on-prem, and multi-party collaborations

Illustrative FC Workflow



Federated Architecture - Privacy Preserving Technology is a Critical Aspect of FC

FC can be complemented with industry best-practice security standards (e.g. encryption at rest, in transit, and in-processing aka confidential computing; as well as remote hardware attestation) along with privacy enhancing technologies (PETs), such as differential privacy, k-anonymization, and homomorphic encryption. Data never leaves its secure environment; only analytical output to encrypted parameters or model updates are exchanged.

SECURITY

- Data, code, and model parameters are encrypted at rest, in-transit, and in processing (with support for confidential computing across clouds - ex: AWS, Azure, and GCP)
- Role based access controls designate user access to data, models, code, and logs
- Attestation service enables validating and authorizing code containers
- Inbound ports in firewalls remain closed
- Full audit logs verify appropriate usage of data

PRIVACY

- Differential Privacy - a specified level of noise is added to data before calculating metrics
- Partial Weight Sharing - a fraction of model weights are dropped and not shared with the FL server
- Secure Multi-Party Computation (SMPC) - the model updates from federated clients can be aggregated in a privacy-preserving

CONFIDENTIALITY

- Users have no direct access to collaborators' data
- Row-level data never leaves its site - only aggregate statistics and model weights travel to the FL Server
- Projects and their underlying datasets and code objects are visible only to collaborators in that project

Federated AI Architecture Enables Collaborative AI Systems in Drug Discovery

Local Nodes

Each AI system participant (e.g., a Pharma company, biotech or healthcare org) operates as a separate node - maintaining its own secure server. This server may hold:

- Multi-modal data: Including clinical, genomic, and omic.
- Proprietary Compound Libraries: Millions of virtual molecules.
- Assay Results: Data on how those molecules interacted with specific proteins or diseases in the past.
- Local Training Code: A specialized AI model (often a Graph Neural Network or Transformer) that "reads" chemical structures.

Orchestrator

A central cloud server acts as the neutral "referee." It doesn't hold any data. Instead, it manages the workflows, additional privacy enhancements and the version control of the AI model. It sends the latest "blank slate" model to every node and waits for their mathematical updates to come back.

The orchestrator allows for additional app, cloud, and agent integrations in a secure way.

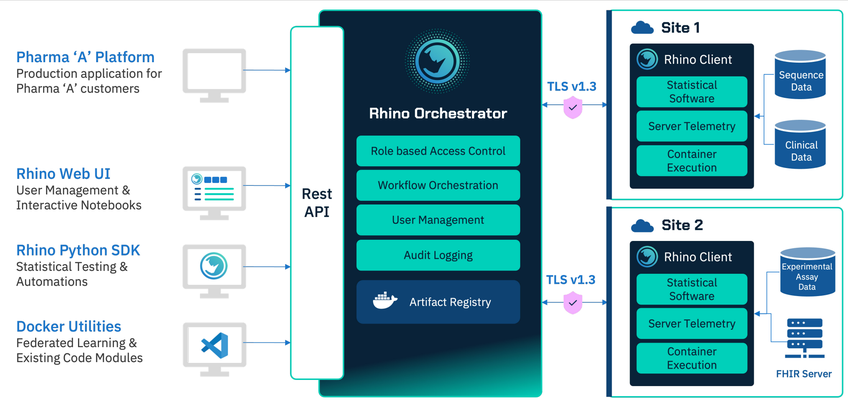
Learnings (Weights)

When a node trains the model on its sensitive or regulated data, it sends back the weights, allowing the central model to gain all of the insights from the data without having to actually see it. For example, signals that say, "I found that molecules with this specific nitrogen bond tend to be less toxic."

Key Capabilities Enabled

Phase	Traditional AI	Federated AI
Data Access	Limited to what one company owns.	Access to the "Collective Intelligence" of the other participants.
IP Protection	Risk of leaks if data is shared.	Risks mitigated; proprietary structures stay behind firewalls.
Model Robustness	Too little, lack of diverse data in one node to train a model.	Combines data slices from 'endless' nodes to find patterns.
Regulatory	High legal barrier to moving patient data.	Naturally compliant (GDPR/HIPAA) as data stays local.
Architecture	Build for each use case. Difficult to expand.	Build once for many uses; Easily deploy new apps or agents or add new nodes.

Illustrative Collaborative Federated AI Infra Layer



Examples of Flagship FC Programs

Lilly TuneLab

Federated AI program where dozens of biotechs securely run inference and fine-tune Eli Lilly models on their own data - without moving data or IP. (Tunelab.Lilly.com)

Impact:
- Enables AI models to learn from external partner datasets
- Partners keep full control of IP and data
- Continuous improvement of discovery models

FAITE CONSORTIUM

A federated consortium enabling AbbVie, AstraZeneca, Amgen, Johnson & Johnson, and UCB to co-train and fine-tune AI for predicting properties of biologics. (www.FAITEConsortium.com)

Impact:
- Shared model development
- Privacy-preserving collaboration
- Advancing prediction of biologic properties

Cancer AI Alliance

Federated consortium focused on connecting siloed cancer research from different treatment centers for training AI models to accelerate treatment development and improve care. (www.CancerAlliance.AI)

Impact:
- Multi-institution oncology AI collaboration
- NCI-designated cancer centers
- Multi-modal clinical + imaging datasets
- Federated learning across institutions

Conclusion

Federated computing enables access to data in ways previously not possible, where data and IP stay private, secure, and compliant. Federated computing provides the enterprise AI architectural layer needed to enable the development of collaborative AI systems, where AI models and agents can be built and deployed and shared insights are amplified. FC is already showing promise with established drug discovery and research programs.

Rhino Federated Computing | www.RhinoFCP.com | Contact: Elke Nelson-Nichols: elke@rhinofcp.com

Sources: [1] How successful are AI-discovered drugs in clinical trials? A first analysis and emerging lessons. https://www.sciencedirect.com/science/article/pii/S135964462400134X