*Proceedings of Machine Learning Research* **318** (2026)
The 39th Canadian Conference on Artificial Intelligence

Canadian AI 2026
25-29 May 2026,Vancouver,BC

# Fed-Universe: A Semantic-Geometric-Topological-Human (S-G-T-H) Stack for Negotiated Alignment in Federated Systems

Ting Xu[†,*] Henry Leung[‡]
[†] Department of Biomedical Engineering, University of Calgary, Calgary, Canada
[‡] Department of Electrical and Software Engineering, University of Calgary, Calgary, Canada

**Abstract**

Conventional federated learners implicitly optimize for global means, creating a "one-size-fits-all" paradigm that inevitably suppresses the minority under cross-site heterogeneity. To bridge this critical gap, we present *Fed-Universe*, a generalizable Semantic–Geometric–Topological–Human ($\mathcal{S}$–$\mathcal{G}$–$\mathcal{T}$–$\mathcal{H}$) architecture that transforms passive averaging into an active negotiation for decentralized alignment. The $\mathcal{S}$-layer deploys an edge-capable LLM as a semantic surrogate to translate heterogeneous inputs into standardized patient profile prompts. The $\mathcal{G}$-layer enforces geometric quality assurance (GQA) via a mapping $f : \Theta \to Q$ from the statistical parameter space to a physically interpretable quality domain, employing a cosine similarity gate to retain minority updates based on physical geometry rather than statistical proximity. The $\mathcal{T}$-layer executes topological Pareto control, identifying a sensitivity-based knee point ($\lambda_k$) to balance multi-objective aggregation without majority degradation. Finally, the $\mathcal{H}$-layer operationalizes a cognitive twin dynamic to mirror the user's real-time psychological state, dynamically adapting human-computer interaction modes via active intent alignment. To ground this theoretical architecture empirically, we validated the $\mathcal{S}$-$\mathcal{G}$-$\mathcal{T}$ core on a binational synthetic clinical network simulation. This proof-of-concept demonstrated a "zero-sum escape": a minority node (<3% data volume) achieved utility parity (loss reduction from 0.857 to 0.340) alongside dominant hubs. Our *Fed-Universe* framework is being expanded along two focused pillars: institutional multimodal alignment (*Fed-Ultra*) and individual cognitive intent alignment (*Fed-Human*).

**Keywords:** Multimodal Federated Learning, Large Language Model, Privacy Preservation, Parameter-Efficient Fine-Tuning, Pareto Optimization, Human-in-the-Loop.

## 1. Introduction

Standard federated learning paradigms (e.g., FedAvg [1]) operate under a "one-size-fits-all" objective. By implicitly optimizing for the global mean, they act as passive systemic averagers [2, 3]. In non-IID real-world settings, the global model tends to overfit dominant hubs while neglecting local heterogeneity [4, 5], such as cross-site physiological variation, rare patient subpopulations, or human emotional variance. Furthermore, in human-computer interaction (HCI), rigid models often discard cognitive variance as noise. For instance, if an AI treats a user's frustration as noise and resets to a blank slate (the "stateless user paradox"), the misunderstood user grows more frustrated. This drives the user to abandon the interaction, eventually leading to "feedback collapse" where the system also learns nothing from the user [6].

To bridge this critical gap, we propose *Fed-Universe*. Named for its universal adaptability, it serves as a foundational meta-architecture capable of deployment across diverse dimensions: **Fed-Ultra** (context-aware medical imaging) and **Fed-Human** (human-aware HCI intent alignment). Unlike personalized federated learning (PFL) that isolates local models, *Fed-Universe* maintains a collaborative knowledge commons by acting as an active negotiator, NOT an averager. It operationalizes a Semantic–Geometric–Topological–Human ($\mathcal{S}$-$\mathcal{G}$-$\mathcal{T}$-$\mathcal{H}$) stack to semantically translate heterogeneous contexts, geometrically gate misaligned updates, topologically balance minority representations [7], and dynamically adapt

* tina.xu1@ucalgary.ca

HCI via cognitive mirroring [8]. By shifting from rigid aggregation to dynamic negotiation, *Fed-Universe* establishes a symbiotic ecosystem where algorithmic intelligence and human cognition continuously co-evolve.

## 2. Methods

*Fed-Universe* is a robust framework that reframes federated learning as a structured negotiation rather than a naive averaging process. Concretely, the architecture operates across four synergistic layers as shown in Figure 1:

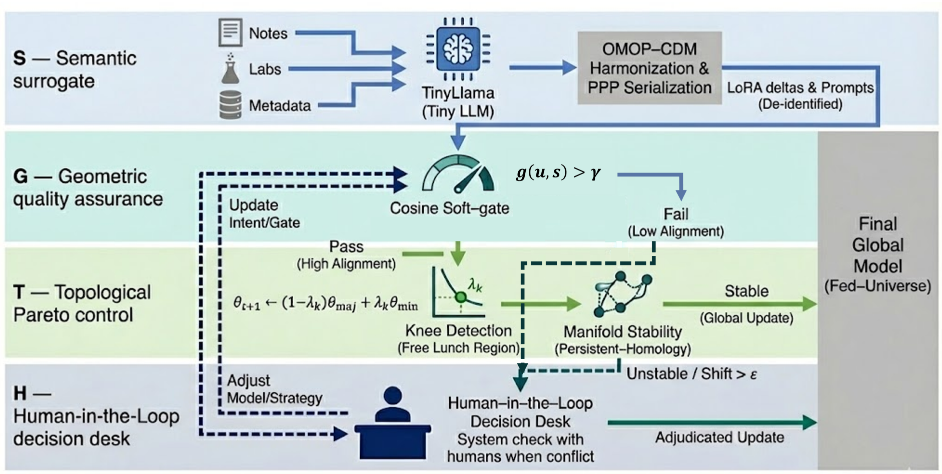


*Figure 1.* **The *Fed-Universe* architecture:** A federated $\mathcal{S}$-$\mathcal{G}$-$\mathcal{T}$-$\mathcal{H}$ negotiation stack for decentralized intent alignment and Pareto-optimal fairness, equipped with localized geometric safety gates.

- **$\mathcal{S}$-Layer: Semantic surrogate.** A lightweight LLM (e.g., TinyLlama [9]) operates inside each client site as a local data translator. Heterogeneous raw inputs (clinical notes, lab results, metadata) are harmonized to the OMOP-CDM standard [10] and serialized into **patient profile prompts (PPP)**. To ensure stringent privacy and minimize compute overhead, raw files remain strictly localized; only de-identified intents and small LoRA parameter deltas [11] are shared with the global server.
- **$\mathcal{G}$-Layer: Geometric quality assurance (GQA).** We propose a mapping $f : \Theta \to Q$ from the statistical parameter space $\Theta$ to a physically interpretable quality domain $Q$. Conventional statistical distance-based aggregators (e.g., Krum [12], RFA [13]) often conflate valid minority features with adversarial noise. By evaluating directional alignment within $Q$ via a continuous cosine similarity gate:
$$g(u, s) = \sigma(\beta \cos\theta(u, s) - \tau)$$
GQA isolates structural validity from statistical frequency. This ensures updates from localized minorities (e.g., high-BMI cohorts) are preserved based on their underlying physical geometry, rather than rejected as anomalies in $\Theta$.
- **$\mathcal{T}$-Layer: Topological Pareto control (TPC).** To systematically balance majority ($\theta_{\text{maj}}$) and minority ($\theta_{\text{min}}$) representations, we formulate the global update as a multi-objective negotiation:
$$\theta_{t+1} \leftarrow (1 - \lambda_k)\theta_{\text{maj}} + \lambda_k\theta_{\text{min}}$$

To enforce topological control across the parameter space, we map the empirical Pareto frontier. Specifically, the optimal aggregation coefficient ($\lambda_k$) is algorithmically determined via nonlinear sensitivity analysis. By identifying the point of maximum curvature along the utility trade-off curve (the topological "knee point"), this approach mathematically targets the "free lunch region," maximizing minority utility without degrading the established majority baseline.

- **$\mathcal{H}$-Layer: Human-in-the-Loop (HITL) decision desk.** This layer operationalizes the **cognitive twin dynamic (CTD)** to mirror the user's real-time context [8], enabling the adaptation of localized modes based on captured states. When algorithmic alignment ($g(u, s) < \gamma$) or Pareto stability ($\lambda_k$) fails, the $\mathcal{H}$-layer resolves the resulting "interaction divergence" by synchronizing systemic updates with the mirrored human intent. By treating users as dynamic entities rather than static data sources, this alignment prevents feedback collapse and fosters a virtuous cycle of human-AI co-evolution, ultimately ensuring human sovereignty as the failsafe.

## 3. Results

We evaluate the $\mathcal{S}$–$\mathcal{G}$–$\mathcal{T}$ pipeline using a binational simulation integrating Canadian and U.S. synthetic clinical registries. The architectural proof-of-concept is explicitly validated: the $\mathcal{S}$-layer standardizes heterogeneous inputs into de-identified LLM prompts, while the $\mathcal{G}$-layer safeguards updates that align with the "golden direction". Finally, the $\mathcal{T}$-layer identifies the Pareto-optimal knee point ($w = 2.0$, calculated using $\lambda_k \approx 0.67$) to balance the multi-objective aggregation and accommodate underrepresented nodes.

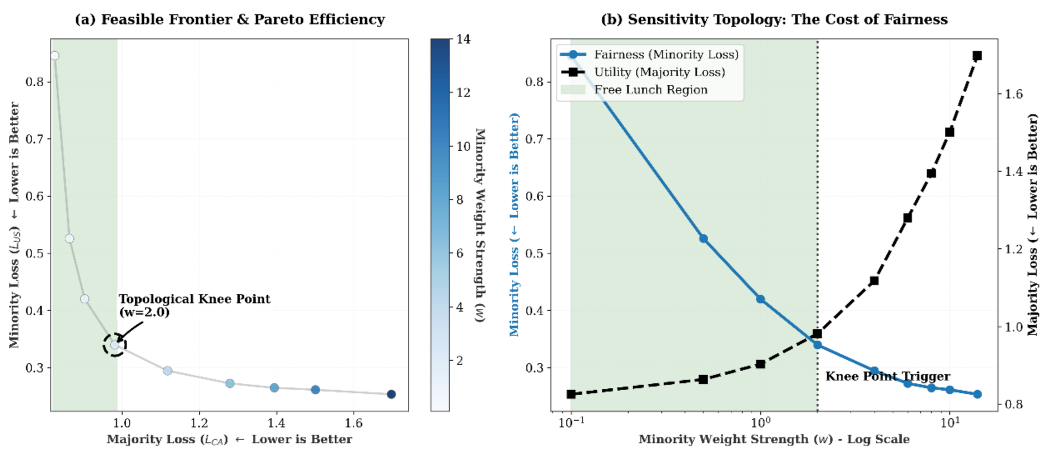


*Figure 2.* **Pareto frontier sensitivity analysis:** The identified knee point at $w = 2.0$ establishes an optimal equilibrium, maximizing minority spoke utility without degrading the majority hub baseline.

As shown in Figure 2, the $\mathcal{S}$–$\mathcal{G}$–$\mathcal{T}$ trajectory deviates from the suboptimal FedAvg equilibrium. Decoupled from volume dominance, the minority node (<3% of total data) achieves utility parity, with its loss decreasing from 0.857 to 0.340. Notably, this gain occurs without degrading the data-rich majority baseline (0.981 vs. 0.985). These findings suggest that the framework effectively captures the "free-lunch region," ensuring that global alignment incorporates rather than suppresses underrepresented patterns. Compared to simple averaging, the pipeline ensures that global generalizability does not come at the cost of local representation, providing the necessary algorithmic foundation for our two focused application pillars: *Fed-Ultra* and *Fed-Human*.

## 4. Discussion

Having validated the $\mathcal{S}$-$\mathcal{G}$-$\mathcal{T}$ core on a binational synthetic clinical network, the next phase of our thesis is to rigorously test the framework's generalizability. We structure our framework around two representative pillars of alignment failure in medical AI:

- **Fed-Ultra:** This pillar integrates epidemiological priors (NHANES, SEER) [14] with ultrasound datasets (DDTI, TN3K) via $\mathcal{S}$-layer translation and $\mathcal{T}$-layer fusion. By employing GQA in the $\mathcal{G}$-layer, updates from physiologically distinct cohorts (e.g., high-BMI) are retained based on valid physical structures in $Q$, avoiding mis-classification as statistical anomalies in $\Theta$. We hypothesize this approach will improve early thyroid cancer screening accuracy for underrepresented populations compared to standard statistical gating.
- **Fed-Human:** Addressing human-computer interaction, this pillar mitigates interaction divergence—where model optimization drifts from user needs. The $\mathcal{G}$-layer isolates core user intent from high-entropy emotional variance to provide a reliable alignment vector. Concurrently, the $\mathcal{H}$-layer utilizes cognitive twin dynamics [8] to adapt system responses based on the user's real-time psychological state, effectively resolving the stateless user paradox [6] and establishing a virtuous cycle of adaptive human-AI collaboration.

Looking ahead, the $\mathcal{T}$-layer will leverage graph Laplacian regularization as a computationally efficient proxy for mutual information. By embedding client-graph topology directly into the Pareto objective, this formulation aims to minimize edge processing overhead while ensuring robust, deterministic intra-cluster cohesion without relying on heuristic adaptive thresholds.

## 5. Conclusion

*Fed-Universe* reframes federated learning as active negotiation via a four-layer $\mathcal{S}$-$\mathcal{G}$-$\mathcal{T}$-$\mathcal{H}$ stack designed to break the "one-size-fits-all" paradigm in federated learning. By transforming passive averaging into structured negotiation, the framework integrates semantic translation ($\mathcal{S}$), geometric quality assurance ($\mathcal{G}$), topological Pareto control ($\mathcal{T}$), and dynamic intent alignment via cognitive twin mirroring ($\mathcal{H}$), accommodating the underrepresented while emphasizing human-centric application. Validated on a binational clinical simulation, the $\mathcal{S}$-$\mathcal{G}$-$\mathcal{T}$ core achieved a "zero-sum escape": a minority node ($<3\%$ data volume) reduced loss from 0.857 to 0.340 without degrading the majority baseline. These results ground the forthcoming *Fed-Ultra* and *Fed-Human* pillars toward a collaborative, equitable knowledge commons and a robust human-AI partnership.

## Acknowledgements

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC). The authors acknowledge the Canadian Network for Observational Drug Effect Studies (CNODES) and the MITRE Corporation (SyntheticMass) for the synthetic datasets used in this research. The authors also thank Baihui Chen from Beijing Normal University for her expertise in psychological counseling on the design of the $\mathcal{H}$-layer.

## Declaration of Generative AI and AI-assisted Technologies

During the preparation of this work, the authors used Large Language Models (including Claude and Gemini) in order to polish the language and assist with LaTeX structural formatting. After using these tools, the authors thoroughly reviewed and edited the content as needed and take full responsibility for accuracy and integrity of the final content.

## References


[1] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas. "Communication-Efficient Learning of Deep Networks from Decentralized Data". In: *Artificial Intelligence and Statistics*. PMLR. 2017, pp. 1273–1282.

[2] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith. "Federated optimization in heterogeneous networks". In: *Proceedings of Machine Learning and Systems* 2 (2020), pp. 429–450.

[3] Q. Yang, Y. Liu, T. Chen, and Y. Tong. "Federated machine learning: Concept and applications". In: *ACM Transactions on Intelligent Systems and Technology (TIST)* 10.2 (2019), pp. 1–19.

[4] M. Mohri, G. Sivek, and A. T. Suresh. "Agnostic federated learning". In: *International Conference on Machine Learning*. PMLR. 2019, pp. 4615–4625.

[5] P. Kairouz, H. B. McMahan, B. Avent, A. Bellet, M. Bennis, A. N. Bhagoji, K. Bonawitz, Z. Charles, G. Cormode, R. Cummings, et al. "Advances and open problems in federated learning". In: *Foundations and Trends® in Machine Learning* 14.1–2 (2021), pp. 1–210.

[6] S. Amershi, D. Weld, M. Vorvoreanu, A. Fourney, B. Nushi, P. Collisson, J. Suh, S. Iqbal, P. N. Bennett, K. Inkpen, et al. "Guidelines for human-AI interaction". In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 2019, pp. 1–13.

[7] Z. Hu, Y. Wang, and X. Liu. "Federated learning in non-IID settings: A topological perspective". In: *IEEE Transactions on Neural Networks and Learning Systems* 34.5 (2022), pp. 2341–2355.

[8] B. Shneiderman. "Human-centered artificial intelligence: Reliable, safe & trustworthy". In: *International Journal of Human–Computer Interaction* 36.6 (2020), pp. 495–504.

[9] P. Zhang, G. Zeng, T. Wang, and W. Lu. "TinyLlama: An Open-Source Small Language Model". In: *arXiv preprint arXiv:2401.02385* (2024).

[10] G. Hripcsak, J. D. Duke, N. H. Shah, C. G. Reich, V. Huser, M. J. Schuemie, M. A. Suchard, R. W. Park, I. C. K. Wong, P. R. Rijnbeek, et al. "Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers". In: *Studies in Health Technology and Informatics* 216 (2015), pp. 574–578.

[11] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen. "LoRA: Low-Rank Adaptation of Large Language Models". In: *International Conference on Learning Representations*. 2022.

[12] P. Blanchard, R. El Mghari, R. Guerraoui, and J. Stainer. "Machine learning with adversaries: Byzantine tolerant gradient descent". In: *Advances in Neural Information Processing Systems*. Vol. 30. 2017.

[13] K. Pillutla, S. M. Kakade, and Z. Harchaoui. "Robust aggregation for federated learning". In: *IEEE Transactions on Signal Processing* 70 (2022), pp. 1142–1154.

[14] V. K. Nguyen et al. "Harmonized US National Health and Nutrition Examination Survey 1988-2018 for high throughput exposome-health discovery". In: *Scientific Data* 10.1 (2023), p. 96.