International Journal on Science and Technology (IJSAT)
E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

Low-Resource Adaptive Pretraining using Knowledge-Infused Curriculum Learning with Trusted Federated Explainability for Integrity, Accountability, and Trade-off Control

Mohan Siva Krishna Konakanchi
mohansivakrishna16@gmail.com

***Abstract*—Pretrained language models have become foundational for modern NLP, yet pretraining remains computationally expensive and data-hungry. Many organizations and research teams operate under low-resource constraints—limited compute, limited labeled data, restricted access to large corpora, or governance constraints that prevent data centralization. In these environments, *adaptive pretraining* (continuing pretraining on domain or task-relevant data) can yield strong gains, but selecting what to pretrain on and in what order becomes critical to efficiency. Curriculum learning offers a framework for ordering training examples, while knowledge infusion offers a way to steer learning using structured signals such as ontologies, knowledge graphs, and rule-derived annotations. However, the combination introduces practical risks: knowledge sources vary in quality across silos, the resulting training loop can be fragile under non- IID data, and explainability requirements increase operational overhead.**

This paper proposes *KICL-APT* (Knowledge-Infused Curriculum Learning for Adaptive PreTraining), a low-resource frame- work that integrates (i) knowledge-infused curriculum design to prioritize high-yield training examples, (ii) a trust metric- based federated learning (FL) governance layer to enable cross- silo collaboration without centralizing raw data, and (iii) an explicit controller to quantify and optimize the explainability–performance trade-off. KICL-APT introduces a trust metric that scores each participant using provenance, update consistency, evaluation reliability, and policy compliance for knowledge usage. Trust-aware robust aggregation limits poisoning and reduces the influence of low-quality knowledge sources. Explainability budgeting provides stable, actionable rationales for both curriculum decisions and model behaviors while controlling overhead.

We evaluate KICL-APT through a controlled prototype simulation that emulates low-resource adaptive pretraining across heterogeneous silos with variable knowledge quality and non- IID domain distributions. Results show that knowledge-infused curricula improve sample efficiency and reduce energy proxy cost compared to unguided adaptive pretraining, while trust-aware aggregation improves robustness under faulty and adversarial contributors. Moderate explanation budgets achieve stable explanations with limited degradation in downstream performance. We conclude with deployment guidance for knowledge-infused curricula in resource-constrained, federated, and auditable settings.

***Index Terms*—adaptive pretraining, low-resource NLP, curriculum learning, knowledge infusion, federated learning, trust metrics, explainable AI, integrity, accountability**

I. Introduction

Large-scale pretraining has dramatically improved NLP performance, enabling transfer across tasks with minimal super- vision. Transformer architectures and attention-based training have become core enablers [6]. However, full-scale pretraining is costly; low-resource settings must instead rely on *adaptive pretraining*, where a base pretrained model is continued on smaller domain corpora or specialized text. Adaptive pretraining can be effective, but it is also sensitive: small datasets, domain noise, and poor example selection can lead to overfit- ting or wasted compute.

Curriculum learning—training on examples in a purposeful order—can improve convergence and sample efficiency in various contexts [**?**]. Although the earliest curriculum literature predates 2010, curriculum and related self-paced learning methods have been actively explored through the 2010s, including large-scale learning and representation settings [1]–[3]. In parallel, *knowledge infusion* aims to incorporate structured knowledge (ontologies, lexicons, knowledge graphs, rules) into learning, improving robustness and data efficiency. Knowledge infusion is particularly valuable when labeled data is limited or when domain constraints (e.g., medical, finance, compliance) require grounding in structured semantics.

A. Low-Resource Constraints and Cross-Silo Reality

In many organizations, domain data and knowledge are split across silos:

- region- or tenant-specific text corpora cannot be centralized,
- proprietary glossaries, ontologies, and labeling rules differ across teams,
- governance rules restrict sharing of raw text due to privacy and compliance.

Federated learning (FL) offers a path to train shared representations without centralizing raw data [7], [8], [12]. Yet, federated adaptive pretraining introduces new risks: non-IID distributions, heterogeneous knowledge quality, and integrity failures that can corrupt shared models.

B. Explainability and Governance Requirements

Adaptive pretraining and knowledge infusion can affect model behavior in opaque ways. High-stakes domains require explainable decisions and auditable lineage: why a corpus segment was prioritized, which knowledge sources were used, and how they influenced behavior. Explainability methods provide tools for understanding model decisions [13]–[15], but explanations add overhead and can reduce performance if they constrain model choices [17]. Therefore, a low- resource framework must manage an explicit explainability–performance trade-off.

C. Problem Statement

We address three coupled problems:

P1 (Sample-efficient adaptive pretraining). How can low-resource adaptive pretraining select and order training examples to maximize gains per unit compute?

P2 (Trusted cross-silo learning). How can multiple silos collaborate via FL while ensuring integrity and accountability, especially with heterogeneous knowledge quality?

P3 (Explainability–performance trade-off). How can we quantify and optimize the trade-off between explainability and downstream performance under low-resource budgets?

D. Contributions

This paper proposes *KICL-APT* and contributes:

- A practical knowledge-infused curriculum design that ranks training examples using a mix of difficulty, knowledge coverage, and domain relevance signals, designed for low-resource efficiency.
- A trust metric-based federated governance layer that binds learning influence to provenance, update consistency, evaluation reliability, and knowledge-policy compliance.
- A trust-aware robust aggregation approach to reduce poisoning and mitigate low-quality knowledge effects [10], [11].
- An explainability controller using budgets and stability checks to manage the explainability–performance trade-off operationally.
- A prototype evaluation under non-IID silos, variable knowledge quality, and adversarial/faulty contributors.

II. Related Work

A. Transformers and Representation Learning

Transformer architectures and self-attention are foundational for modern pretraining and transfer learning [6]. While this paper does not propose a new architecture, our framework targets how to *adapt* pretrained representations efficiently under resource constraints.

B. Curriculum and Self-Paced Learning

Self-paced learning and curriculum strategies select training instances in a staged manner, often from easier to harder or from high-confidence to low-confidence examples [1], [2]. Curriculum approaches have also been studied in reinforcement learning and broader learning contexts [3]. These methods motivate our curriculum scoring approach but do not address knowledge infusion combined with federated governance.

C. Knowledge Infusion and Knowledge Graphs

Knowledge graphs and structured knowledge have been used to improve representations and reasoning. A major direction is learning embeddings for entities and relations [4], and integrating structured knowledge with neural models. In low-resource settings, knowledge can act as a prior that reduces the need for large labeled datasets.

D. Federated Learning and Robust Aggregation

Federated learning enables collaborative training without centralizing data [7], [8]. Secure aggregation preserves privacy of updates [9]. Robust aggregation aims to tolerate adversarial updates [10], [11]. However, federated pretraining with knowledge infusion faces additional integrity risks: a silo may provide low-quality or biased knowledge rules that subtly corrupt the shared model. This motivates explicit trust metrics and accountability controls.

E. Explainable AI

Model-agnostic explanations such as LIME and SHAP, and gradient-based attributions such as Integrated Gradients, provide tools to interpret model outputs [13]–[15]. Concerns about post-hoc explanations in high-stakes settings motivate interpretable models or constrained explanation processes [17]. We use explanation budgets to operationalize these concerns.

F. Auditability and Accountability

Permissioned blockchain systems and secure logging infrastructures support auditable provenance and non-repudiation [18], [19]. KICL-APT adopts an audit plane that records commitments for curricula decisions, trust reports, and model lineage.

III. KICL-APT Framework Overview

KICL-APT integrates three layers: (i) knowledge-infused curriculum learning for low-resource adaptive pretraining, (ii) trusted federated governance, and (iii) explainability trade-off control.

A. Layer 1: Knowledge-Infused Curriculum

Each silo constructs a curriculum that orders examples for adaptive pretraining. The curriculum uses three categories of signals:

- **Difficulty signals:** model uncertainty, loss on examples, or proxy perplexity.
- **Domain relevance signals:** similarity to target domain terms, recent incidents, or key intents.
- **Knowledge coverage signals:** coverage of ontology terms, entity relations, or rule triggers.

In low-resource settings, the curriculum aims to maximize improvement per unit compute by prioritizing high-yield examples.

B. Layer 2: Trusted Federated Learning

Each silo trains locally and shares model updates. A trust metric determines influence and robust aggregation reduces anomalies.

C. Layer 3: Federated Explainability and Trade-off Controller

Explainability budgets allocate explanation effort where needed (e.g., curriculum choices and model behaviors affecting critical concepts). The controller selects explanation methods and stability checks based on budgets and governance requirements.

IV. Trust Metric-Based Federated Governance

A. Threat Model

We consider:

- **Faulty silos:** misconfigured curricula, noisy labels, unstable evaluations.
- **Adversarial silos:** poison updates by injecting biased knowledge rules or corrupted corpora segments.
- **Accountability evasion:** missing provenance, unverifiable knowledge sources, or inconsistent reporting.

B. Trust Metric (Operational Definition)

Each silo i receives a trust score $T_i \in [0, 1]$ computed from normalized components:

- **Provenance and reproducibility (P_i):** verifiable lineage for corpora snapshots, rule sets, and training configuration.
- **Update consistency (U_i):** anomaly checks for abrupt update shifts and drift relative to historical rounds.
- **Evaluation reliability (E_i):** stability of downstream metrics and consistency across reruns.
- **Knowledge-policy compliance (K_i):** adherence to approved knowledge sources, bias checks, and rule validation procedures.
- **Curriculum sanity (C_i):** signals that curriculum is not pathological (e.g., not always selecting the same narrow subset).

Guardrails. Severe violations sharply reduce trust:

- missing provenance attestations for knowledge sources,
- repeated evaluation inconsistencies,
- rule sets that fail validation checks,
- updates flagged as outliers across multiple rounds.

C. Trust-Aware Robust Aggregation

Standard FedAvg aggregates updates weighted by data size [8]. KICL-APT uses:

Aggregation influence = data weight × trust weight.

After trust gating, robust filtering (trimmed or selection-based) reduces the impact of remaining anomalous updates [10], [11]. Secure aggregation can be used when privacy is critical [9].

D. Audit Plane

KICL-APT records commitments for:

- global model lineage per round,
- trust score rationale summaries,
- curriculum configuration fingerprints (hashes of curriculum scoring parameters),
- knowledge source version identifiers (hashed references),
- aggregation metadata (number of gated silos).

A permissioned ledger or append-only log can support integrity and non-repudiation [18], [19].

V. Knowledge-Infused Curriculum Design

This section describes how curricula are built in a way suitable for low-resource constraints.

A. Curriculum Scoring Signals

Each candidate training example receives a *curriculum score* computed from interpretable components:

- **Learning yield proxy:** expected improvement if trained on this example (estimated from current loss/uncertainty).
- **Knowledge yield proxy:** how much the example covers important entities, relations, or rules.
- **Diversity proxy:** penalize redundancy to avoid repeatedly training on near-duplicate samples.
- **Risk/importance proxy:** boost examples tied to critical intents or compliance concepts.

The scoring is implemented as a weighted combination of normalized proxies (kept simple for auditability).

B. Curriculum Stages

KICL-APT uses staged training:

- **Stage 1 (Anchor knowledge):** prioritize examples with high knowledge coverage and low ambiguity to anchor entity/term representations.
- **Stage 2 (Domain adaptation):** prioritize high domain relevance and diverse examples for broader adaptation.
- **Stage 3 (Hard cases):** include more difficult examples to improve robustness.

This structure resembles "easy to hard" curricula but is enhanced with explicit knowledge coverage signals.

C. Knowledge Infusion Mechanisms

KICL-APT supports lightweight knowledge infusion without complex modeling:

- **Rule-derived tags:** annotate examples with ontology concepts or relation indicators.
- **Entity linking signals:** mark entity spans and types using internal dictionaries or knowledge graphs.
- **Contrastive pairs:** construct simple positive/negative pairs based on knowledge consistency.

These mechanisms are compatible with low-resource settings and can be computed locally.

VI. Explainability–Performance Trade-off Framework

A. Explainability Targets

KICL-APT explains:

- **Curriculum decisions:** why certain examples were prioritized (knowledge coverage, domain relevance, risk).
- **Model behaviors:** why the adapted model behaves a certain way on critical inputs.
- **Federated governance actions:** why certain silos were down-weighted or gated.

B. Explanation Methods

We use established explanation primitives:

- **LIME/SHAP-style attributions** for token/feature influence [13], [14].
- **Integrated Gradients** for differentiable models [15].
- **Rule-like anchors** for compact rationales when feasible [16].

Explanations are computed locally and shared as summaries with commitments logged.

C. Explainability Quality Metrics

Operational measures:

- **Stability:** top-k agreement under perturbations.
- **Actionability:** mapping to ontology concepts, entities, or known domain intents.
- **Fidelity:** local alignment with model behavior.

D. Budgeted Trade-off Controller

An *explanation budget* per round determines:

- how many curriculum decisions to explain deeply,
- how many model behaviors to audit with full explanations,
- whether stability checks are enforced.

The controller selects a configuration that maximizes a simple utility notion:

> *Utility increases with downstream performance and explanation quality, and decreases with explanation cost and training cost.*

This makes trade-offs explicit for governance.

VII. Methodology

A. Prototype Evaluation Setup

We evaluate using a controlled simulation representing:

- $N = 20$ silos with non-IID domain corpora,
- variable corpus size (low-resource) per silo,
- heterogeneous knowledge quality (some silos have incomplete or noisy ontologies),
- faulty and adversarial participants.

B. Adaptive Pretraining Protocol

Each silo continues pretraining from a shared base model for a limited number of steps under strict compute budgets. We compare:

- **Unguided APT:** random sampling from local corpus.
- **Curriculum APT:** difficulty-based curriculum without knowledge signals.
- **KICL-APT:** knowledge-infused staged curriculum.

C. Federated Protocol Variants

We compare:

- **FedAvg** [8]
- **Robust-only** [11]
- **Trust-only** (trust-weighted averaging)
- **KICL-APT Federated** (trust gating + robust filtering + knowledge-infused curriculum)

D. Downstream Evaluation

We evaluate adapted models on downstream tasks representative of domain intents. Metrics:

- **Task score:** normalized accuracy/F1 (abstracted).
- **Compute proxy:** normalized training steps and sequence processing cost.
- **Robustness drop:** degradation under adversarial/faulty silos.
- **Explanation stability:** top-k stability for selected samples.

VIII. Experiments

A. Integrity Failure Injection

We include:

- **Faulty silos (4):** unstable evaluation and noisy curricula (over-selecting narrow subsets).
- **Adversarial silos (2):** biased knowledge rules to skew representations of sensitive concepts.

B. Budget Regimes

We test explanation budgets:

- **E1 Low:** deep explanations for top 5% most critical curriculum and model events.
- **E2 Medium:** deep explanations for top 20% with stability checks.
- **E3 High:** deep explanations for all selected events.

IX. Results

To avoid formatting issues, tables are minimal in columns.

A. Sample Efficiency and Performance

Table I compares downstream task score and compute proxy. Knowledge-infused curricula improve both task score and compute proxy efficiency by prioritizing high-yield examples and reducing wasted steps on redundant or low-signal text.

TABLE I
DOWNSTREAM SCORE AND COMPUTE PROXY (LOWER IS BETTER FOR COMPUTE)

Method	Task Score	Compute Proxy
Unguided APT	0.78	1.00
Curriculum APT	0.81	0.94
KICL-APT (local)	**0.84**	**0.88**

TABLE II
FEDERATED OUTCOMES UNDER INTEGRITY FAILURES

Method	Task Score	Robust Drop
FedAvg	0.80	0.06
Robust-only	0.82	0.04
Trust-only	0.83	0.03
KICL-APT Federated	**0.86**	**0.01**

B. Federated Robustness Under Integrity Failures

Table II reports federated outcomes under faulty/adversarial silos.

Trust-aware robust aggregation reduces the impact of biased knowledge rules and unstable silos, improving both performance and robustness.

C. Explainability–Performance Trade-off

Table III shows how explanation budgets affect stability and task score.

TABLE III
EXPLAINABILITY BUDGET TRADE-OFF (KICL-APT FEDERATED)

Budget	Task Score	Expl. Stability
E1 Low	**0.87**	0.60
E2 Medium	0.86	**0.75**
E3 High	0.85	0.78

A moderate budget provides substantial stability improvements with minimal performance loss, supporting a practical governance policy: audit and explain the most critical curriculum and model behaviors, not every event.

X. DISCUSSION

A. Why Knowledge Infusion Helps in Low-Resource Adaptive Pretraining

In low-resource settings, random sampling wastes compute on redundant or low-signal data. Knowledge infusion provides structured priors:

- anchor representations around key entities and relations,
- emphasize rare but important concepts,
- improve robustness to domain noise by using ontology constraints.

Curriculum staging ensures these priors are learned early, improving sample efficiency.

B. Cross-Silo Knowledge Quality Variance

Different silos may have conflicting or incomplete ontologies. Trust metrics and knowledge-policy compliance help:

- prevent low-quality knowledge from dominating global updates,
- encourage provenance and reproducibility of knowledge sources,
- support post-hoc auditing of knowledge influence.

C. Interpretable-First vs Hybrid Approaches

In high-stakes domains, post-hoc explanations may be insufficient [17]. KICL-APT supports:

- **Interpretable-first mode:** stronger reliance on rule-derived tags and transparent curriculum scoring.
- **Hybrid mode:** higher-capacity models with budgeted explanations and stability checks.

D. Limitations

Prototype simulation. Our evaluation is simulated; real-world corpora and knowledge graphs may introduce additional challenges.

Knowledge bias. Knowledge sources can encode bias. Knowledge-policy compliance and audits mitigate but do not eliminate bias risks.

Curriculum complexity. Overly complex curricula can be fragile. KICL-APT keeps scoring simple to remain auditable.

Trust gaming. Participants may attempt to optimize trust scores. Guardrails and periodic audits mitigate this risk.

XI. CONCLUSION

This paper proposed KICL-APT, a low-resource adaptive pretraining framework that combines knowledge-infused curriculum learning with trusted federated explainability. KICL-APT improves sample efficiency by prioritizing high-yield examples using difficulty, domain relevance, and knowledge coverage signals. It enables cross-silo collaboration without centralizing raw data through a trust metric-based federated learning layer that ensures integrity and accountability, us- ing trust-aware robust aggregation and an auditable lineage plane. Finally, KICL-APT introduces a practical controller to quantify and optimize the explainability–performance trade-off using explanation budgets and stability checks. Prototype sim- ulation results demonstrate improved downstream performance and compute efficiency, robustness under integrity failures, and stable explanations under moderate budgets. Future work includes real-world deployments, richer knowledge validation, and privacy-preserving explanation sharing across regulatory boundaries.

ACKNOWLEDGMENT

The author thanks the research community for foundational work on curriculum learning, knowledge representation, federated learning, and explainable AI that informed this framework perspective.

REFERENCES

[1] M. Kumar, B. Packer, and D. Koller, "Self-paced learning for latent variable models," in *Proc. NeurIPS*, 2010.
[2] L. Jiang, D. Meng, S. Yu, Z. Lan, S. Shan, and A. Hauptmann, "Self-paced learning with diversity," in *Proc. NeurIPS*, 2015.
[3] A. Graves, M. G. Bellemare, J. Menick, R. Munos, and K. Kavukcuoglu, "Automated curriculum learning for reinforcement learning," in *Proc. ICML*, 2017.
[4] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko, "Translating embeddings for modeling multi-relational data," in *Proc. NeurIPS*, 2013.
[5] M. Nickel, K. Murphy, V. Tresp, and E. Gabrilovich, "A review of relational machine learning for knowledge graphs," *Proc. IEEE*, vol. 104, no. 1, pp. 11–33, 2016.
[6] A. Vaswani *et al.*, "Attention is all you need," in *Proc. NeurIPS*, 2017.
[7] J. Konec˘ny´, B. McMahan, and D. Ramage, "Federated optimization: Distributed optimization beyond the datacenter," *arXiv preprint arXiv:1511.03575*, 2015.
[8] H. B. McMahan *et al.*, "Communication-efficient learning of deep networks from decentralized data," in *Proc. AISTATS*, 2017.
[9] K. Bonawitz *et al.*, "Practical secure aggregation for privacy-preserving machine learning," in *Proc. ACM CCS*, 2017.
[10] P. Blanchard, E. Mhamdi, R. Guerraoui, and J. Stainer, "Machine learning with adversaries: Byzantine tolerant gradient descent," in *Proc. NeurIPS*, 2017.
[11] D. Yin, Y. Chen, K. Ramchandran, and P. Bartlett, "Byzantine-robust distributed learning: Towards optimal statistical rates," in *Proc. ICML*, 2018.
[12] P. Kairouz *et al.*, "Advances and open problems in federated learning," *arXiv preprint arXiv:1912.04977*, 2019.
[13] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you?: Explaining the predictions of any classifier," in *Proc. ACM KDD*, 2016.
[14] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proc. NeurIPS*, 2017.
[15] M. Sundararajan, A. Taly, and Q. Yan, "Axiomatic attribution for deep networks," in *Proc. ICML*, 2017.
[16] M. T. Ribeiro, S. Singh, and C. Guestrin, "Anchors: High-precision model-agnostic explanations," in *Proc. AAAI*, 2018.
[17] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nature Machine Intelligence*, vol. 1, no. 5, pp. 206–215, 2019.
[18] E. Androulaki *et al.*, "Hyperledger Fabric: A distributed operating system for permissioned blockchains," in *Proc. EuroSys*, 2018.
[19] B. Putz, F. Pernul, and G. Kablitz, "A secure and auditable logging infrastructure based on a permissioned blockchain," *Computers & Security*, vol. 87, 2019.