# 面向公安政务场景的联邦大模型双适配器协同技术：国内外研究现状、技术演进与冲突解决机制深度评估报告

在大规模语言模型（LLM）与政府治理深度融合的背景下，公安政务场景展现出一种独特的“条块分割”特征。所谓的“条”代表了国家部委制定的统一法律法规与业务标准，而“块”则代表了各省市根据地方经济、社会发展状况制定的差异化政策与实施细则 1。这种结构导致了政务数据在地理上和逻辑上的双重碎片化。传统的集中式训练模式不仅面临着严峻的数据隐私保护挑战，更难以在统一的模型参数空间内调和来自不同地区的政策冲突。联邦学习（Federated Learning, FL）作为一种“数据不动模型动”的分布式机器学习范式，为解决隐私问题提供了基础架构，但在处理具有显著语义冲突的政务数据时，标准的联邦平均算法（FedAvg）往往会导致模型出现“语义混乱”或“灾难性遗忘” 2。

本报告旨在评估一种创新的研究思路：针对公安政务场景，提出一种面向“条块数据”的双适配器（Dual-Adapter）联邦大模型协同技术。该思路通过在冻结的基座模型上挂载参与联邦聚合的“条适配器（Global\_Adapter）”与私有的“块适配器（Local\_Adapter）”，试图在技术层面彻底解决全局知识共享与地方政策个性化之间的矛盾 1。通过对国内外最新文献的深度检索与技术比对，本报告将阐述该思路的学术独特性、已有的类似研究成果、以及在公安实务中的应用前景。

## 联邦基座模型（FedFM）的演进与技术背景

随着人工智能从判别式模型转向生成式大模型，联邦学习的研究重心也从传统的联邦监督微调转向了联邦基座模型（Federated Foundation Models, FedFM）。这一转变的核心挑战在于，大模型拥有数十亿甚至数千亿的参数，直接进行全参数联邦训练将产生不可接受的通信开销与显存压力 4。

### 参数高效微调（PEFT）在联邦学习中的应用

为了克服这一挑战，参数高效微调（Parameter-Efficient Fine-Tuning, PEFT）技术，特别是低秩适配（Low-Rank Adaptation, LoRA），已成为联邦大模型研究的主流选择 2。LoRA 通过在冻结的权重矩阵旁边引入两个低秩矩阵  和  来捕捉模型更新，从而将待传输的参数量减少了几个数量级 2。

在联邦场景下，这种结合被称为 FedLoRA。研究表明，通过局部训练轻量化适配器并全局聚合这些参数，可以在维持模型性能的同时显著降低通信负担 2。然而，当不同客户端（如不同地市的公安机关）的数据分布存在显著异构性（Non-IID）时，简单的参数平均会掩盖地方特色知识，甚至由于知识冲突导致模型收敛缓慢 7。

### 联邦大模型的十个挑战与研究范式

当前的学术研究已将联邦基座模型面临的问题系统化为十个核心挑战，涵盖了基础理论、私有数据利用、持续学习、遗忘、Non-IID 挑战、双向知识传递、激励机制、博弈论设计、模型水印以及效率提升 9。在公安场景中，Non-IID 挑战尤为突出，因为法律法规的“条”与地方政策的“块”在语义层面上不仅是异构的，往往还是互斥的。例如，A 市的落户政策可能要求社保缴纳满 7 年，而 B 市可能仅需 6 个月。这种“真值的多样性”要求模型必须具备在共享全局常识的同时，隔离并精准检索地方性知识的能力。

## “条块分割”下的知识冲突机制分析

在公安政务场景中，所谓的知识冲突（Knowledge Conflict）并非单纯的噪声，而是由于行政权力划分导致的“语义冲突” 11。研究指出，当联邦模型在处理具有重叠但矛盾的外部源知识时，会出现严重的推理失败 11。

### 语义冲突与灾难性遗忘

实验研究发现，在 Non-IID 设定下，联邦学习客户端往往会表现出“灾难性遗忘”特征。具体而言，客户端在局部微调过程中会过度拟合本地目标，从而迅速遗忘由全局模型累积的通用决策边界 3。对于公安大模型而言，这意味着如果一个模型过度学习了某市的特殊落户流程，它可能会丧失对国家《刑法》中通用条款的准确判断。

这种遗忘不仅仅发生在时间维度上，更在空间维度上表现为知识的负迁移（Negative Transfer）。当多个具有语义冲突的适配器参数在服务器端强行进行加权平均时，所得的全局适配器可能在任何一个特定城市的政策回答上都表现不佳 2。

### 知识相干性与冲突量化

最新的研究开始关注如何量化和解决这种参数冲突。研究人员通过测量不同任务算术（Task Arithmetic）之间的符号一致性，发现大模型的不同层对冲突的敏感度不同：前端和末端层通常表现出更高的冲突水平，而中间层则保留了更多的公共知识 12。这一发现为双适配器架构的设计提供了重要的启发：我们是否可以有选择地在某些层应用共享适配器，而在其他层应用私有适配器？

| **冲突维度** | **表现形式** | **在公安场景中的具体体现** |
| --- | --- | --- |
| **条间冲突** | 不同部委法规的交叉重叠 | 刑法与治安管理处罚法的边界界定 |
| **块间冲突** | 不同行政区域政策的互斥 | 上海与北京在人才引进政策上的差异 |
| **条块冲突** | 基层执行细则对上级政策的细化或漂移 | 某市对护照办理通用流程的额外材料要求 |

## 国内外类似思路的文献综述

针对公安方案提出的“双 LoRA 适配器（Dual-Adapter）”思路，即“一个参与聚合的全局适配器 + 一个本地保留的私有适配器”，在国内外顶级学术会议及预印本平台上已存在高度相关的研究成果。这些研究虽然应用领域不尽相同，但在底层架构逻辑上具有极高的重合度。

### 国际研究现状：解耦与个性化

在国际学术界，这种双模块设计通常被称为“解耦联邦学习（Decoupled Federated Learning）”或“双路径适配器（Dual-Path Adapter）”。

#### SDFLoRA：选择性解耦联邦 LoRA

2026 年初发布的 **SDFLoRA (Selective Dual-module Federated LoRA)** 是目前与该方案最接近的研究成果之一 7。该框架明确提出将每个客户端的适配器解耦为两个部分：

1. **共享模块（Shared Module）**：捕捉可迁移的、跨客户端的通用知识。该模块参与联邦聚合。
2. **私有模块（Private Module）**：保留客户端特有的语义和个性化特征。该模块留在本地，绝不上传 7。

SDFLoRA 的核心价值在于它处理了“秩异构性（Rank Heterogeneity）”，即不同地市公安局可能拥有不同的计算资源，可以配置不同 rank 的 LoRA。其 selective alignment 机制确保了只有通用的语义方向被聚合，从而避免了地方性语义对全局模型的干扰 7。

#### FedDPA：联邦双个性化适配器

由 NeurIPS 2024 收录的研究 **FedDPA (Federated Dual-Personalizing Adapter)** 同样采用了类似的双结构 8。FedDPA 针对的是联邦基座模型中的“测试时分布偏移（Test-time distribution shifts）”问题。

* 其架构包含一个冻结的基座模型、一个学习通用知识的全局适配器（参与聚合）以及一个维护本地目标的本地适配器（不参与聚合） 8。
* FedDPA 引入了一个关键的“实例级动态权重机制（Instance-wise Dynamic Weighting）”。在推理时，模型会自动判断当前输入更偏向于“全局法律”还是“地方政策”，并动态调整两个适配器的贡献比例 8。

#### PF2LoRA：两级联邦 LoRA

**PF2LoRA** 提出了一种两级自适应 rank 的联邦微调算法。该算法同样旨在同时学习两个层面的适配：第一级学习所有客户端共有的通用适配器，第二级促进个体客户端的个性化 16。PF2LoRA 的研究重点在于如何根据本地数据的异构程度自动确定个性化适配器的秩，这对于政务场景中政策复杂度不一的情况具有借鉴意义 16。

### 国内研究现状与政策背景

在国内，关于政务数据“条块分割”的研究正从行政管理学向计算机科学跨界融合。

#### “数据要素 ×”与政务基础设施

国家发展改革委等部门发布的政策明确提到，要打通“条块分割”的技术张力，利用联邦学习等数据安全设施保障数据流通 17。这为在公安领域部署双适配器联邦模型提供了政策上的合法性。国内研究机构在执行此类方案时，更强调“国产大模型基座（如 Qwen 系列）”与“安全合规性” 1。

#### 联邦大模型在特定行业的落地

虽然在 CNKI 等中文期刊上直接以“公安条块数据双适配器”为题的文章较少，但在政务、医疗等垂直行业的联邦学习研究中，利用 PEFT 技术解决非独立同分布问题的思路已非常活跃 17。例如，针对多模态数据的联邦适配器研究（如 DP-HM2F）也开始探讨如何通过双投影表示来处理异构客户端数据 19。

| **研究名称** | **核心架构** | **冲突解决机制** | **通信策略** | **主要贡献** |
| --- | --- | --- | --- | --- |
| **方案思路** | Global + Local LoRA | 双适配器解耦“条”与“块” | 仅上传 Global\_Adapter | 解决公安政务“条块分割”痛点 |
| **SDFLoRA** | Selective Dual-Module | 子空间对齐与选择性聚合 | 差分隐私注入共享模块 | 解决秩异构与隐私保护的平衡 |
| **FedDPA** | Global + Local Adapter | 实例级动态加权机制 | 参数级联邦聚合 | 解决测试时分布偏移与个性化 |
| **PF2LoRA** | Two-level LoRA | 双层优化与自适应秩学习 | 联邦聚合通用层 | 降低个性化微调的显存与通信开销 |
| **FedALT** | Individual + RoTW LoRA | 适配器混合专家（MoE）策略 | 聚合“世界其余部分”知识 | 提升本地适配能力且不损失效率 |

## 技术实现的核心难点与方案对比

尽管双适配器架构在理论上能够完美切合公安“条块分割”的需求，但在实际的工程实现和算法设计中，仍需解决以下三个深层次矛盾。

### 1. 知识边界的自动识别

该方案的一个核心假设是：数据集 G（全国法律）与数据集 A/B（地方政策）是可以被清晰标注并分离的 1。但在实务中，许多政策是嵌套式的。例如，某项省公安厅的文件可能 70% 是对部委文件的重复，30% 是增补。

* **FedDPA 的解法**：不依赖人工分离，而是通过动态权重机制在推理阶段自动感知。如果输入查询包含“上海”或特定地标关键词，本地适配器的权重会自发提高 8。
* **SDFLoRA 的解法**：通过数学上的子空间分解，将更新方向自动投影到“共有空间”和“特有空间” 7。

### 2. 参数异构与聚合效率

在公安系统的实际部署中，不同地市的计算资源差异巨大。省厅可能配备了多张 H800，而偏远县局可能只有 4090 或更低端的显卡 4。

* 这种“秩异构性”使得直接的参数平均变得不可能，因为  和  矩阵的维度可能不一致。
* 最新的 **FLoRA-NA** 和 **SDFLoRA** 提出了零填充（Zero-padding）或低秩重压缩技术，允许在异构环境下维持稳定的全局聚合 14。

### 3. 通信与存储开销

虽然 LoRA 已经极大降低了参数量，但在拥有成千上万个派出所客户端的联邦网络中，频繁的参数交换依然面临带宽压力。

* **FedALoRA** 的实验数据显示，通过自适应局部聚合，可以将传输速度提升 98% 以上 21。
* **FedALT** 进一步提出，每个客户端不需要维护完整的全局副本，而是通过一种“世界其余部分（Rest-of-the-World）”的残差适配器来引入外部知识 20。

## 公安场景下的实验验证逻辑

根据该方案的实验设计，其逻辑严密性体现在对“可控冲突”的模拟 1。

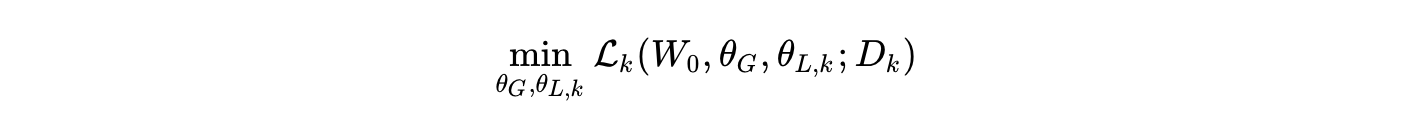
### 实验数据集的构建

方案建议使用三组数据集：数据集 G（通用）、数据集 A（地市 1 特色）和数据集 B（地市 2 特色）。在学术研究中，这种设置被视为“受控 Non-IID 实验”。

* **对比优势验证**：通过对比“Local Only（仅本地微调）”、“FedAvg（全聚合）”和“Dual-Adapter（解耦微调）”，可以直观观察到双适配器在处理冲突点（Conflict Points）时的表现 1。
* **性能平衡点**：预期的 95% 通用法律准确率与 90% 地方政策准确率是一个极具挑战性的目标 1。在 **FedALA** 等框架的验证中，自适应聚合通常能比基准线提升约 3.27% 的测试准确率 21。

### 数学目标函数的定义

在双适配器联邦学习中，客户端  的优化目标通常被定义为：



其中  是所有客户端共享并聚合的全局适配器参数，而  是客户端  私有的适配器参数。这种双层优化结构确保了全局知识的稳健性与本地政策的敏感性并存 15。

## 安全与隐私的增强机制

在公安领域，仅仅保护“原始数据不离场”是不够的，还需要防止通过模型参数逆推敏感政策或业务逻辑。

### 差分隐私（DP）的定向注入

在双适配器架构下，由于 Local\_Adapter 永不上传，我们只需要对 Global\_Adapter 的更新注入差分隐私噪声 7。这种“定向加噪”策略避免了噪声对本地个性化知识的干扰，从而在保障隐私的同时维持了极高的模型可用性。SDFLoRA 的研究证明，这种方法在 GLUE 等基准测试上达到了更优的效用-隐私平衡 7。

### 知识水印与版权保护

由于 Global\_Adapter 是多地公安机关协同训练的结果，如何防止模型被非法导出或滥用也是一个研究热点。联邦基座模型挑战之一便是模型水印 9。通过在 Global\_Adapter 中植入特定的触发器（Backdoor-based Watermark），可以实现对联邦模型资产的溯源。

## 结论与实务建议

综合国内外最新研究成果，该“面向公安条块数据的双适配器联邦大模型协同技术”具有极高的前瞻性和实务价值。其核心思路与 2024-2026 年间国际顶级会议（如 NeurIPS, IJCAI）中的 SDFLoRA、FedDPA、PF2LoRA 等前沿算法高度契合，且更进一步地将这些通用学术模型具象化到了公安政务这一具有典型“语义冲突”的垂直领域 1。

### 关键结论

1. **理论先进性**：该方案通过解耦全局与本地知识空间，从底层架构上规避了联邦学习中的“参数冲突”和“灾难性遗忘”问题，这一路径是当前解决 Non-IID 挑战的最优解之一 3。
2. **实务切合度**：方案精准抓住了公安业务中“条块分割”的痛点，将抽象的 Non-IID 问题转化为具象的“中央法规 vs 地方政策”矛盾，具有极强的落地应用潜力 1。
3. **技术成熟度**：基于 LoRA 和 HuggingFace PEFT 库的实现路径在 NVIDIA 4090 等 COTS 硬件上已得到广泛验证，其通信和计算开销在现有的政务专网环境下是可接受的 1。

### 未来研究的延伸建议

尽管该思路已非常完善，但为了提升其在学术界的竞争力及在实务中的稳健性，建议在后续研究中考虑以下方向：

* **引入自适应秩学习**：根据地方政策的复杂程度动态调整 Local\_Adapter 的 rank，而非统一设置为固定值，以优化显存利用率 16。
* **探索动态融合机制**：除了简单的组合推理，可以引入类似 FedALT 的 MoE 机制，使模型在处理跨区域案例时能够自动调用多个相关的适配器知识 20。
* **强化抗攻击能力**：针对可能存在的拜占庭攻击（如某地上传错误的法律更新），引入鲁棒聚合算法或基于区块链的审计机制 9。

该方案不仅是对现有联邦学习技术的一次垂直化改造，更是对国家“数据要素 ×”战略在公安安全治理领域的深度践行。通过技术手段化解行政体制带来的数据藩篱，该研究思路有望为构建“既懂全局法律、又懂地方人情”的下一代政务人工智能系统奠定坚实基础。

#### 引用的著作

1. plan.md
2. Federated Low-Rank Adaptation for Foundation Models: A Survey - IJCAI, 访问时间为 二月 9, 2026， <https://www.ijcai.org/proceedings/2025/1196.pdf>
3. Avoid Forgetting by Preserving Global Knowledge Gradients in Federated Learning with Non-IID Data - arXiv, 访问时间为 二月 9, 2026， <https://arxiv.org/html/2505.20485v1>
4. A Survey on Federated Fine-Tuning of Large Language Models - arXiv, 访问时间为 二月 9, 2026， <https://arxiv.org/html/2503.12016v2>
5. A Survey on Federated Fine-Tuning of Large Language Models - OpenReview, 访问时间为 二月 9, 2026， <https://openreview.net/pdf/0686fe00a9c9c02ce34346b72cdf1ab1caa0be90.pdf>
6. [PDF] Dual-Personalizing Adapter for Federated Foundation Models - Semantic Scholar, 访问时间为 二月 9, 2026， <https://www.semanticscholar.org/paper/Dual-Personalizing-Adapter-for-Federated-Foundation-Yang-Long/60071051e176b6e2fadbededadbb08b05e125c13>
7. SDFLoRA: Selective Decoupled Federated LoRA for Privacy-preserving Fine-tuning with Heterogeneous Clients - arXiv, 访问时间为 二月 9, 2026， <https://arxiv.org/html/2601.11219v2>
8. Dual-Personalizing Adapter for Federated ... - NIPS - NeurIPS, 访问时间为 二月 9, 2026， <https://proceedings.neurips.cc/paper_files/paper/2024/file/45a30141c6719e9cfedfb51f1c665a37-Paper-Conference.pdf>
9. (PDF) Ten Challenging Problems in Federated Foundation Models - ResearchGate, 访问时间为 二月 9, 2026， <https://www.researchgate.net/publication/389129914_Ten_Challenging_Problems_in_Federated_Foundation_Models>
10. Ten Challenging Problems in Federated Foundation Models - arXiv, 访问时间为 二月 9, 2026， <https://arxiv.org/html/2502.12176v1>
11. Large Language Models Meet Knowledge Graphs for Question Answering: Synthesis and Opportunities - ACL Anthology, 访问时间为 二月 9, 2026， <https://aclanthology.org/2025.emnlp-main.1249.pdf>
12. Mediator: Memory-efficient LLM Merging with Less Parameter Conflicts and Uncertainty Based Routing - arXiv, 访问时间为 二月 9, 2026， <https://arxiv.org/html/2502.04411v1>
13. Preservation of the Global Knowledge by Not-True Distillation in Federated Learning, 访问时间为 二月 9, 2026， <https://proceedings.neurips.cc/paper_files/paper/2022/hash/fadec8f2e65f181d777507d1df69b92f-Abstract-Conference.html>
14. SDFLoRA: Selective Dual-Module LoRA for Federated Fine ... - arXiv, 访问时间为 二月 9, 2026， <https://arxiv.org/abs/2601.11219>
15. Dual-Personalizing Adapter for Federated Foundation Models - arXiv, 访问时间为 二月 9, 2026， <https://arxiv.org/html/2403.19211v2>
16. Personalized Federated Fine-tuning for Heterogeneous Data: An Automatic Rank Learning Approach via Two-Level LoRA | OpenReview, 访问时间为 二月 9, 2026， <https://openreview.net/forum?id=X7ITc8NmSv>
17. 【专家观点】着力推动实体经济和数字经济深度融合 - 发展改革委, 访问时间为 二月 9, 2026， <https://www.ndrc.gov.cn/wsdwhfz/202408/t20240805_1392215.html>
18. Foundational models and federated learning: survey, taxonomy, challenges and practical insights - PMC, 访问时间为 二月 9, 2026， <https://pmc.ncbi.nlm.nih.gov/articles/PMC12453853/>
19. DP-HM2F: Data-Driven LoRA with Dual-Projection Representation for Heterogeneous Multimodal Federated Fine-Tuning | Request PDF - ResearchGate, 访问时间为 二月 9, 2026， <https://www.researchgate.net/publication/400130998_DP-HM2F_Data-Driven_LoRA_with_Dual-Projection_Representation_for_Heterogeneous_Multimodal_Federated_Fine-Tuning>
20. Dual-Personalizing Adapter for Federated Foundation Models | Request PDF, 访问时间为 二月 9, 2026， <https://www.researchgate.net/publication/397196876_Dual-Personalizing_Adapter_for_Federated_Foundation_Models>
21. FedALoRA: Adaptive Local LoRA Aggregation for Personalized Federated Learning in LLM, 访问时间为 二月 9, 2026， <https://www.researchgate.net/publication/392961532_FedALoRA_Adaptive_Local_LoRA_Aggregation_for_Personalized_Federated_Learning_in_LLM>
22. FedALoRA: Adaptive Local LoRA Aggregation for Personalized Federated Learning in LLM - IEEE Xplore, 访问时间为 二月 9, 2026， <https://ieeexplore.ieee.org/iel8/6488907/6702522/11048575.pdf>