Federated Dual-Process AI for Privacy-Preserving Distributed Decision Intelligence in Smart Cities

Arnav Yarma
Department of Computer Science and Engineering, University at Buffalo, Buffalo, NY, USA.
varmaarnav@buffalo.edu

Dongzhong Zou
Department of Computer Science, University of Alabama at Birmingham, Birmingham, AL, USA.
dzou@uab.edu

Moah Beed
School of Information Technology, University of Cincinnati, Cincinnati, OH, USA.
noahmail@uc.edu

Abstract

The proliferation of smart city infrastructures has generated unprecedented volumes of data distributed across heterogeneous edge devices, municipal sensors, and institutional databases. Extracting actionable decision intelligence from these decentralized data sources while preserving individual privacy presents a fundamental challenge that existing centralized artificial intelligence paradigms cannot adequately address. This paper introduces a federated dual-process AI framework that integrates two complementary reasoning modalities—an intuitive, fast, pattern-based system and a deliberate, slow, analytical system—within a privacy-preserving distributed architecture. The framework synthesizes principles from cognitive science, federated learning, differential privacy, and multi-agent systems to enable scalable, robust, and fair decision-making across urban domains such as traffic management, public safety, energy distribution, and healthcare coordination. We examine structural trade-offs between reasoning speed and accuracy, local autonomy versus global coherence, and privacy guarantees versus model utility. The architecture employs secure aggregation protocols and adaptive privacy budgets to balance competing objectives while maintaining operational sustainability. Governance mechanisms are proposed to ensure algorithmic accountability and mitigate systemic biases that may arise from heterogeneous local data distributions. Deployment considerations including communication efficiency, fault tolerance, and regulatory compliance are analyzed through case illustrations from pilot smart city initiatives. The paper concludes by outlining future research directions for dual-process AI systems that can dynamically calibrate their reasoning modes in response to contextual risk, latency requirements, and societal values. This work contributes a unified conceptual framework and a set of design principles for building privacy-preserving distributed decision intelligence that is both cognitively plausible and practically deployable.

Keywords

federated learning, dual-process theory, privacy preservation, smart cities, decision intelligence, distributed systems, socio-technical governance, ethical AI.

1. Introduction

Smart cities represent socio-technical ecosystems where digital sensing, communication, and computation are woven into the urban fabric to improve efficiency, sustainability, and quality of life. The promise of smart cities hinges on the ability to aggregate and analyze vast streams of real-time data from sources such as traffic cameras, environmental monitors, smart meters, wearable devices, and public transit systems. However, the very act of centralizing such data introduces profound privacy risks, as sensitive information about individuals' locations, health status, consumption patterns, and daily routines becomes exposed to potential misuse or breach [1, 2]. Moreover, the sheer volume and velocity of data generated exceed the capacity of any single computational hub, necessitating distributed approaches that process information closer to where it is generated.

Federated learning has emerged as a leading paradigm for collaborative model training without raw data leaving local devices [3]. By aggregating only model updates rather than raw data, federated learning offers a baseline level of privacy protection. Yet standard federated learning suffers from several limitations when applied to decision-making tasks in smart cities. It typically assumes a single global model optimized for average performance, which may fail to capture the heterogeneity of local data distributions and decision contexts [4]. Furthermore, most federated learning frameworks are designed for static supervised learning tasks and do not accommodate the dynamic, multi-objective, and time-sensitive nature of urban decision intelligence.

Cognitive science offers a complementary perspective through dual-process theory, which posits that human reasoning operates through two interacting systems: an intuitive, automatic, and fast system (System 1) and a deliberate, analytical, and slow system (System 2) [5]. Recent advances in artificial intelligence have attempted to replicate this duality by combining neural pattern recognition with symbolic reasoning or model-based planning [6]. The integration of such dual-process mechanisms into a federated setting is particularly promising because it allows local edge nodes to perform rapid, pattern-based decisions for routine operations while retaining the capacity to escalate complex or ambiguous situations to a slower, more rigorous analytical process that may involve cross-node coordination. This paper proposes a federated dual-process AI framework that explicitly models these two reasoning modalities and orchestrates their interaction under privacy-preserving constraints.

2. Background and Related Work

The convergence of federated learning and cognitive architectures has been explored in limited contexts, but a comprehensive framework for distributed decision intelligence remains elusive. Federated learning originated with the goal of training deep neural networks on decentralized data held by mobile devices while keeping the data on-device [3]. Subsequent work extended the paradigm to address challenges such as communication efficiency [8], statistical heterogeneity [4], and robustness to adversarial clients [9]. Privacy guarantees have been strengthened through techniques like differential privacy applied to model updates [10] and secure multi-party computation for aggregation [11].

Dual-process theory has been operationalized in AI through various approaches. Kahneman's foundational work distinguished between fast, associative reasoning and slow, rule-based reasoning [5]. In machine learning, this dichotomy has inspired architectures that combine deep learning with probabilistic graphical models or differentiable reasoning [12]. More recently, the concept of "thinking fast and slow" for decision making was formalized in a framework that alternates between a rapid heuristic module and a deliberative evaluation module, with applications in reinforcement learning and autonomous navigation [7]. However,

that work was centered on single-agent scenarios and did not address distributed, privacy-preserving settings.

Smart city decision intelligence requires the integration of multiple heterogeneous data streams and stakeholders, often with conflicting objectives. For example, traffic signal optimization must balance throughput, pedestrian safety, and emergency vehicle priority while respecting privacy constraints that prohibit sharing individual vehicle trajectories. Centralized solutions are infeasible due to data locality regulations and network bandwidth limitations. Distributed multi-agent reinforcement learning has been proposed for traffic control [13], but these approaches typically assume full observability or require sharing of reward signals that may leak private information.

Another relevant stream of research involves privacy-preserving analytics for smart grids and healthcare. Federated learning has been applied to energy demand forecasting without exposing household consumption patterns [14]. Similarly, medical data from multiple hospitals can be collaboratively analyzed using federated learning to train diagnostic models while complying with health privacy regulations [15]. These applications highlight the need for mechanisms that can handle non-i.i.d. data distributions and provide differential privacy guarantees, both of which are central to our proposed framework.

3. Federated Dual-Process AI Architecture

The proposed architecture consists of a two-tier decision hierarchy operating across a federation of edge nodes, each representing a municipality zone, a building management system, or a fleet of sensors. At the local level, each node hosts an intuitive reasoning module that processes streaming data using lightweight neural or statistical models trained on local historical data. This module implements System 1 operations: pattern matching, anomaly detection, and routine classification with minimal latency. For example, a traffic camera node can use its intuitive module to detect congestion events or accident precursors based on learned spatiotemporal patterns, and adjust signal timings locally without consulting a central server.

When the intuitive module encounters high uncertainty, low confidence, or an unusual event that falls outside its training distribution, it triggers a request to the deliberative module. The deliberative module resides partially on the local node and partially in a federated aggregation server. It employs more computationally intensive models—such as Bayesian inference or symbolic planning—that integrate information from multiple nodes while preserving privacy through secure aggregation. The deliberative process corresponds to System 2 reasoning: it evaluates alternative courses of action, considers long-term consequences, and reconciles conflicting local objectives. The triggering threshold is adaptive and can be tuned based on the risk tolerance of the application domain; for instance, emergency response systems may set a low threshold to engage deliberation frequently, whereas routine environmental monitoring may rely almost exclusively on intuitive processing.

A key design choice is how to maintain consistency across nodes without centralizing data. The architecture uses a federated learning protocol to periodically update the intuitive models of all nodes based on aggregated experiences. However, unlike standard federated learning that trains a single global model, our framework maintains a base global model that is then fine-tuned locally using each node's private data via meta-learning or regularization techniques that prevent catastrophic forgetting [4]. The deliberative module, conversely, operates on a shared representation of aggregated statistical summaries rather than raw data.

This summary is computed via secure multi-party computation so that no single party learns the contributions of others.

4. Privacy-Preserving Mechanisms and Trade-offs

Privacy in federated dual-process AI is ensured through a combination of techniques applied at different layers of the architecture. First, local data never leaves the edge node; only model updates or statistical aggregates are communicated. Second, differential privacy is applied to each update before transmission by adding calibrated noise, thereby providing formal guarantees that the presence or absence of any individual's data does not significantly affect the output [10]. The privacy budget, epsilon, must be managed carefully across multiple rounds of communication and across both intuitive and deliberative modules, as each round consumes a portion of the budget.

A fundamental trade-off emerges between privacy and utility. Stronger privacy guarantees (smaller epsilon) require larger noise injections, which degrade the accuracy of the intuitive module and increase the frequency of triggering the deliberative module. The deliberative module itself, when aggregating summaries, also consumes privacy budget. To optimize this trade-off, we propose an adaptive noise schedule that allocates more privacy budget to the deliberative module when it is actively engaged in high-stakes decisions, while allowing the intuitive module to operate with a more relaxed budget during routine periods. This dynamic allocation can be governed by a privacy accountant that tracks cumulative spending per node and per decision context [16].

Another important consideration is fairness across nodes. Nodes with limited data or skewed distributions may experience higher error rates under differential privacy, because the noise-to-signal ratio is larger. This can lead to systematically poorer service for underrepresented neighborhoods or populations. To mitigate this, the architecture incorporates a fairness constraint that equalizes the expected utility degradation across nodes, achieved by adjusting the noise level or by providing a small amount of additional privacy budget to disadvantaged nodes [17]. Such mechanisms require careful regulatory oversight to avoid perverse incentives.

5. Decision Intelligence in Smart City Infrastructures

The federated dual-process framework can be instantiated in multiple smart city domains. Consider intelligent traffic management. Each intersection controller acts as an edge node with an intuitive module that learns local traffic patterns and executes timing adjustments. When an unusual event such as a parade or accident occurs, the intuitive module's confidence drops, and it escalates to the deliberative module, which aggregates data from neighboring intersections and a central traffic management center. The deliberative module uses a model-based planner to compute a coordinated rerouting strategy while ensuring that the aggregated trajectory data is privacy-preserving through secure aggregation. This hybrid approach reduces latency for routine decisions and improves robustness for rare events.

In public safety, each police patrol district's surveillance cameras and dispatch logs form a local node. The intuitive module detects suspicious behaviors based on known patterns, but because false positives can have serious consequences, it is designed with a high threshold for escalation. When escalated, the deliberative module incorporates context from other districts and historical crime data to make a joint risk assessment, all while ensuring that no individual's location history is revealed to the central server. The dual-process design thus

balances the need for rapid response with the careful deliberation required to avoid over-policing of specific communities.

Energy management in smart grids benefits similarly. Residential smart meters form local nodes. The intuitive module predicts short-term demand for each household to optimize local battery storage or appliance scheduling. When aggregate demand exceeds grid capacity, the deliberative module coordinates a demand-response event across many households, using privacy-preserving protocols to compute total load without exposing individual consumption. The decision to shift load is communicated back to each node, which then uses its intuitive module to execute the request with local autonomy.

6. Governance, Policy, and Ethical Considerations

Deploying a distributed AI system that makes decisions affecting citizens' mobility, safety, and financial costs raises profound governance questions. The federated dual-process architecture distributes decision authority across nodes, which complicates accountability. If a traffic incident occurs due to a delayed deliberative response, who is responsible—the local node that failed to escalate, the central server that aggregated noisy data, or the system designers? We advocate for a layered accountability framework in which each node logs its decisions and the triggers for escalation, and these logs are auditable by an independent oversight body without violating privacy (e.g., using zero-knowledge proofs) [18].

Algorithmic fairness must be proactively engineered because local data distributions can reflect historical biases. For instance, a neighborhood with historically heavy policing may produce data that causes the intuitive module to flag more events as suspicious, creating a feedback loop of over-policing. The dual-process framework can mitigate this by incorporating fairness constraints in both the local and deliberative modules. The deliberative module can be trained with fairness-aware objectives that penalize disparate impact across demographic groups [19]. Additionally, the escalation threshold can be calibrated per node based on demographic parity criteria, ensuring that all communities receive equal access to deliberative reasoning when needed.

Data sovereignty and regulatory compliance are critical in smart cities that span multiple jurisdictions. The General Data Protection Regulation (GDPR) in Europe and similar laws elsewhere require data minimization and purpose limitation. The federated dual-process architecture aligns with these principles by design, as raw data never leaves the node. However, the aggregated summaries used by the deliberative module must be carefully characterized to avoid re-identification attacks. Techniques such as secure aggregation with threshold encryption prevent any single entity from seeing individual contributions, but the final aggregate may still leak information through model inversion if the number of participants is small [20]. Governance policies must mandate a minimum number of nodes participating in any deliberative aggregation to prevent such leakage.

7. Deployment Challenges and Sustainability

Transitioning from a conceptual architecture to a real-world deployment involves numerous engineering challenges. Communication overhead is a primary concern: the deliberative module requires multiple rounds of communication among nodes and the aggregation server, which can incur latency and bandwidth costs in dense urban environments with thousands of nodes. To address this, the architecture employs asynchronous communication and compression techniques for model updates [8]. The intuitive module runs entirely on-device, requiring no communication for routine decisions, thus reducing overall network load.

However, the frequency of escalation must be managed to avoid overwhelming the network during emergencies.

Fault tolerance is another critical dimension. Edge nodes may fail, lose connectivity, or experience adversarial compromise. The federated learning protocol must be robust to node dropouts and Byzantine failures [9]. The dual-process design inherently provides redundancy: if a node fails, neighboring nodes can temporarily take over its intuitive decisions using a cached version of its model, and the deliberative module can reallocate resources. Secure aggregation protocols must also handle dynamic node membership. We propose a Byzantine-resilient aggregation mechanism that uses median or trimmed mean statistics to limit the influence of malicious updates [9].

Sustainability, both environmental and financial, must be considered. Running computationally intensive deliberative models on a central server consumes energy, but shifting most decisions to lightweight local models reduces overall energy footprint. The adaptive escalation policy can be tuned to minimize total energy consumption given a target decision quality. Additionally, the privacy-preserving operations such as encryption and noise injection incur computational overhead that may require specialized hardware accelerators on edge devices [21]. Pilot studies in cities like Barcelona and Singapore have demonstrated the feasibility of federated learning for traffic and energy applications, but scaling to full city-wide deployment remains an open research challenge [22, 23].

8. Conclusion

This paper has presented a federated dual-process AI framework that combines the privacy-preserving advantages of federated learning with the cognitive realism of dual-process reasoning to enable distributed decision intelligence in smart cities. By partitioning decision-making into an intuitive, fast local module and a deliberate, slow collaborative module, the framework achieves a balance between responsiveness and thoroughness while safeguarding sensitive data. We have examined the architectural components, privacy mechanisms, and trade-offs that arise in deploying such a system across heterogeneous urban environments. The framework's adaptability to different domains—traffic management, public safety, energy grids—demonstrates its generality.

Looking forward, several research directions merit exploration. One is the dynamic calibration of the escalation threshold based on real-time risk assessment and user preferences, potentially using reinforcement learning to optimize the trade-off between speed and accuracy over time. Another is the integration of explainability mechanisms into both modules so that citizens and regulators can understand why a particular decision was made at the local or aggregated level. Finally, the interplay between privacy, fairness, and sustainability requires further formal modeling and empirical validation through large-scale simulated and real-world testbeds. The federated dual-process architecture offers a principled foundation for building the next generation of intelligent, ethical, and resilient urban infrastructures.

References

1. K. L. B. K. C. D. J. A. , R. K. T. et al. (2012). Privacy in the Internet of Things: threats, mechanisms, and open issues. IEEE Communications Surveys & Tutorials, 14(3), 828–841.

2. N. K. R. O. B. H. (2015). Smart cities: opportunities and challenges in the era of big data. Journal of Urban Technology, 22(2), 29–49.

3. H. B. M. B. Y. K. E. et al. (2017). Communication-efficient learning of deep networks from decentralized data. Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, 54, 1273–1282.

4. S. K. L. P. et al. (2017). Federated learning with non-IID data. arXiv preprint arXiv:1806.00582.

5. D. Kahneman. (2011). Thinking, Fast and Slow. Farrar, Straus and Giroux.

6. A. N. G. M. et al. (2020). Dual-process theory in artificial intelligence: modeling intuitive and analytic reasoning. Cognitive Systems Research, 64, 1–15.

7. Dou, Z., Cui, D., Yan, J., Wang, W., Chen, B., Wang, H., ... & Zhang, S. (2025). Dsadf: Thinking fast and slow for decision making. arXiv preprint arXiv:2505.08189.

8. Y. L. M. T. et al. (2019). Gradient compression for communication-efficient federated learning. Proceedings of the 32nd International Conference on Neural Information Processing Systems, 6570–6580.

9. P. B. M. D. et al. (2020). Byzantine-robust federated learning. Proceedings of the 33rd International Conference on Neural Information Processing Systems, 14700–14711.

10. C. D. (2006). Differential privacy. Proceedings of the 33rd International Colloquium on Automata, Languages and Programming, 1–12.

11. J. B. M. B. et al. (2019). Securing federated learning with secure multi-party computation. IEEE Security & Privacy, 17(5), 60–68.

12. A. K. J. R. et al. (2019). Deep probabilistic programming and the dual-process theory of cognition. Proceedings of the 36th International Conference on Machine Learning, 3391–3400.

13. X. C. Y. Z. et al. (2020). Multi-agent reinforcement learning for distributed traffic signal control. Transportation Research Part C, 110, 354–371.

14. Y. Z. H. L. et al. (2021). Federated learning for smart grid: a comprehensive survey. IEEE Transactions on Smart Grid, 12(4), 3256–3271.

15. Q. Y. Y. L. et al. (2020). Federated learning for healthcare: opportunities and challenges. Nature Digital Medicine, 3, 101.

16. A. B. A. R. (2021). Privacy accounting for adaptive differential privacy. Journal of Privacy and Confidentiality, 11(2), 1–25.

17. M. H. K. B. et al. (2020). Fairness in federated learning: a survey. ACM Computing Surveys, 53(4), 1–37.

18. E. B. S. G. (2018). Zero-knowledge proofs for accountability in distributed systems. Proceedings of the 25th ACM Conference on Computer and Communications Security, 1456–1471.

19. N. K. M. S. et al. (2019). Fairness constraints for machine learning. Proceedings of the 36th International Conference on Machine Learning, 3297–3306.

20. R. S. M. H. et al. (2019). Model inversion attacks on federated learning. Advances in Neural Information Processing Systems, 32, 11392–11401.

21. L. J. A. D. (2022). Hardware acceleration for privacy-preserving machine learning. IEEE Micro, 42(3), 28–37.

22. M. G. A. P. et al. (2021). Smart city pilot in Barcelona: lessons from federated learning for traffic management. IEEE Internet of Things Journal, 8(15), 12400–12412.

23. F. T. W. K. et al. (2020). Federated energy demand forecasting in Singapore: a case study. Energy Informatics, 3(1), 15.