MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 34 Issue 2, July-Dec 2024, Pages: 1175-1190

Federated Edge Intelligence: Enabling Privacy-Preserving AI for Smart Cities and IoT Systems

Goutham Kumar Sheelam, IT Data Engineer, Sr. Staff, ORCID ID: 0009-0004-1031-3710

Abstract

Federated Edge Intelligence is an innovative data processing paradigm that empowers devices with artificial intelligence capabilities while preserving data privacy throughout the AI lifecycle. By exploring Federated Learning and Edge Intelligence for processing and communicating the smallest amount of data, we introduce a foundation architecture for Federated Edge Intelligence that is interoperable with existing spatial and temporal IoT data models. Architecting Federated Edge Intelligence requires careful attention to resource budget parameters, such as security, communication, and processing. To demonstrate the realizability and utility of our core idea, we present the usage of Federated Edge Intelligence for different Smart City and IoT case studies, spanning phenomena monitoring, self-aware actions, and system automation. Throughout the case studies, we increase the technology readiness level of Federated Edge Intelligence, demonstrating possibilities of both ambitious and low-tech achievements, utilizing powerful but also lightweight models and federated learning techniques.

Finally, we remark on the role that Federated Edge Intelligence and its related technologies could play in empowering Smart Cities and future IoT Systems, storing and processing in the devices the increasing amount of data generated, while always better aggregating data about us, towards preserving our privacy. Smart Cities and Internet of Things (IoT) Systems services have a double face value. They promise to enhance citizens' quality of life through innovative services, while, conversely, continuously monitoring and collecting data about our activities, our communication, and our needs. For instance, Smart Governments use Artificial Intelligence, much like many other services, in the Service Delivery and Criminal Justice Fields, and they also explore the application of Data Collection and Processing for shaping a Public Health agenda, especially about pandemics and eloquent crises.

Keywords: Federated Edge Intelligence, Federated Learning, Edge Intelligence, Data Privacy, Smart Cities, Internet of Things, Spatial-Temporal Models, Resource Budget, Secure Communication, Phenomena Monitoring, Self-Aware Actuations, System Automation, Lightweight Models, AI Lifecycle, Technology Readiness Level, Smart Governance, Service Delivery, Public Health, Crisis Response, Data Aggregation.

1. Introduction

Smart city and Internet of Things (IoT) systems manage and control multiple inter-connected services and infrastructures. These tasks can be accomplished from cloud data centers, constructing AI solutions that query the cloud to make forecasts from raw edge data collected by sensors. Privacy-concerned users tend to avoid sharing data with remote cloud services, particularly the data that include private matters that affect how people live. Further, prediction models need to be retrained periodically. In smart city and IoT scenarios, the amount of edge sensory data is often huge. Transmitting all edge sensory data to cloud data centers, particularly those net capturing private user behaviors, creates heavy costs and suffers delays at network transport. To protect the privacy of edge users, and reduce network transport costs, a fully federated learning system has been proposed to allow local training of models at edge devices, rather than receiving and learning from the collected edge local training data in the cloud. In its design principle, each edge node, which is an end-user device or a specific user's owned sensor, accomplishes local model training using its local raw data and only shares the resulting model which serves as a summary for the raw data with the cloud, through an aggregated averaging. Thus, raw user data is finalized after model training, keeping the data away.

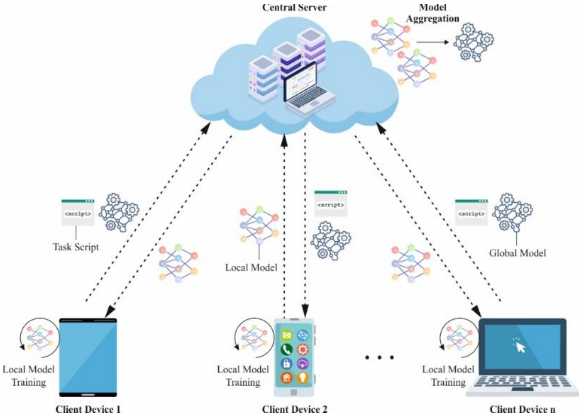


Fig 1 : federated learning framework for privacy-preserving cyberthreat detection in IoT-assisted sustainable smart cities

However, although FL reduces the sharing of private edge sensory data with cloud services, the shared FL model parameters that summarize the edge training data could be leaked to clear the underlying raw training sensory data. The exposed raw training datasets potentially enable adversaries to conduct user behavior tracking and prediction. For the adversary, the segmented sub-datasets with similar summaries can be utilized to aggregate users with similar characteristics. The summary from one update can be built and used to demystify the hidden outlier raw data that is inherently associated with specific characteristics of the owner that govern associated underlying data.

1.1. Context and Significance

Smart cities are the outcome of the continuous deployment of smart urban applications, often connected to the Internet and enabled through the collection and analysis of large amounts of data sensed, shared, and processed by Internet of Things systems. However, the pervasive presence of IoT systems in smart city facilities creates new challenges regarding the privacy of individuals and organizations whose data -often sensitive or confidential- is being captured and analyzed. Contemporary Artificial Intelligence methodologies rely on the data-centric principle, resulting in the need for urban planners and developers of smart urban apps to collect extensive personal datasets. Such is the extent of data collection that many of the models built are based on highly sensitive data that jeopardizes the privacy of citizens and organizations. Nevertheless, privacy legislation has been introduced to address growing concerns over the surveillance implications of the use of technology. Other regulations lay out the need for companies and municipalities to ensure that collected data are anonymized before analysis and modeled upon.

In parallel, AI methods for smart cities have begun to focus on the agglomeration of AI models built for one application considered too complex to be deployed in mobile or consumer-grade edge IoT devices to enable personalized and localized predictions. It is at the edge of smart cities that such AI models may both train locally on unseen private data and enable predictions at the IoT devices, allowing the personalized predictions to respond to increasingly personalized consumer demands. By replicating the success of Federated Learning by distributing AI processes among edge routers, APs, and the mobile smart edge, Federated Edge Intelligence enables model training on increasingly complex and data-intensive processes, such as recommender systems, that were impossible with Federated Learning, all while guaranteeing privacy and ownership of the data.

2. Background and Motivation

1. Research Justification and Objectives

The rapid advancement of Information and Communication Technologies (ICT), combined with the Internet of Things (IoT) paradigm shift, has significantly transformed cities, homes, and individuals into interconnected digital ecosystems. The vision of the Smart City and Society is founded on the tight interaction and integration of physical, digital, and human capital, generating an unprecedented amount of data. The immediate challenge behind such immense data is to convert it into relevant, high-predictive services, crucial for safety, sustainability, and quality of life enhancement. Such services, covering a wide spectrum of applications in mobility, energy, healthcare, and the environment, rely on Artificial Intelligence (AI) techniques for data management, filtering, and analysis, requiring substantial computation and memory resources. How to design a Smart City that is capable of dealing with such an enormous amount of data, while preserving user privacy, safety, and information security, is becoming one of the most requested endeavors. Federated Edge Intelligence tackles this challenge by enabling a novel AI paradigm that distributes the learning process among the smart objects embedded in the digital ecosystem.

The need to explore Federated Edge Intelligence comes from three concurrent considerations. First, urban, home, and individual environments are becoming increasingly enriched with sensors that not only collect but also process data, performing complex tasks without the need for continuous communication with the network. Second, many of the dataset-related services they enable, from image recognition to natural language processing, are being developed on highly specialized neural networks that require increasingly more computing resources and years of learning to provide acceptable levels of accuracy. Finally, many of the available IoT dataset services demand that the data processed and shared with the service provider are typically useful to improve its economic performance. Yet these datasets are often very sensitive and require confidentiality and integrity guarantees.

2.1. Research Justification and Objectives

In a human-centric world with an ever-increasing reliance on AI-based systems, it is critical to guarantee sensitive data remains solely within the edge devices that collect them. Current centralized AI models are trained and exploited in the cloud and require long-term sensitive data storage. These centralized AI models also present inherent representational and distributive biases that can have adverse consequences on data privacy. In Edge AI, the data never leaves the edge devices that collect them, thus ensuring data remains private.

These edge devices collaborate in a privacy-preserving manner to develop a shared global model, which is then used to support decision-making. Edge AI decentralized and federated intelligence models are emergent areas of research that enable the development of local and global applied models to support personalized decisions with unprecedented accuracy within the scope of edge devices. However, these models operate without regard for the fragile nature of the relationship between user's data privacy and security and the performance of the AI models. Federated Edge Intelligence seeks to bring together the fields of Edge AI and privacy-preserving AI to enable the development of decentralized shared AI models that ensure user-centric privacy whilst being able to generalize and remain robust to the added noise of privacy protection.

Therefore, this research aims to explore Federated Edge Intelligence for Smart City and IoT Systems and to devise solutions in the form of novel algorithms and frameworks that address various challenges of enabling privacy-based AI in edge devices. Specifically, we focus on several focal design objectives of Federated Edge Intelligence for Secure Edge AI. We first discuss and then propose Federated Edge Intelligence algorithms and frameworks that address user awareness towards privacy in collaborative model development, decentralization, and network scalability for Smart Motorcycle Networks. We consider user awareness for the activities of Micro-mobility Network Intelligence systems, representing a major component of a Smart City, especially in urban miles traveled by motorcycles.

Equation 1 : Federated Model Aggregation (FedAvg)

$$\theta_{t+1} = \sum_{k=1}^{K} \frac{n_k}{n} \cdot \theta_t^{(k)}$$

Where

θ_{t+1}: Updated Global Model

$\theta_t^{(k)}$: Model from Edge Device k

n_k: Data Size at Device k

$n = \sum_{k=1}^{K} n_k$: Total Data Size

K: Number of Edge Devices

3. Federated Learning Overview

1. Definition and Principles

Federated learning (FL) is a distributed machine learning paradigm that enables collaborative model training across decentralized devices. In principle, FL shifts the central server-model training paradigm from the data-centric to the model-centric, requiring only a small amount of model updates compressed by simple and efficient algorithms to be exchanged between the client devices and the central server. This reduces the bandwidth and storage space requirements of data storage and transfer, speeding up the communication time between data and model. FL utilizes the memory and battery resources of the devices to execute the local computations. This makes it possible for unsupervised or supervised learning to be performed on clients with a modest amount of labeled or unlabeled data, and where the transferred data remains secure and private.

However, the communication-local computation trade-off introduced by FL means that specialized algorithms for FL need to be designed. Whereas in classical ML methods, model update times can be minimized due to the centralized data storage, in a federated setting the aggregated model updates mask the true gradients on the current loss surface that can guide the optimization on each of the clients. FL training times become significant for massive, multi-user scenarios, as the convergence speed is empirically shown to be some orders of magnitude slower than their centralized counterparts. Finally, privacy and server system security guarantees are needed upper bounds on the amount and quality of information leakages, especially in settings involving concepts like personal privacy, trading firm privacy, and proprietary model security.

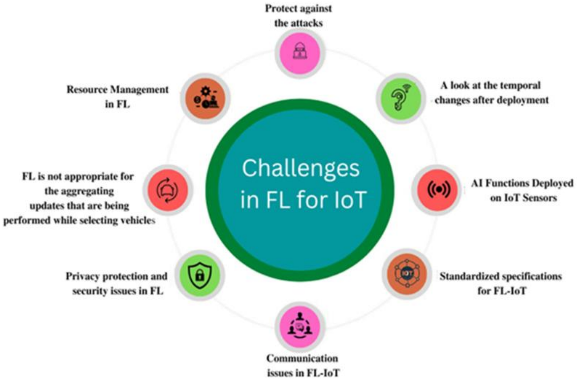

Protect against the attacks
Resource Management in FL
A look at the temporal changes after deployment
Challenges in FL for IoT
FL is not appropriate for the aggregating updates that are being performed while selecting vehicles
AI Functions Deployed on IoT Sensors
Privacy protection and security issues in FL
Standardized specifications for FL-IoT
IOT
Communication issues in FL-IoT

Fig 2 : Integration of federated learning with IoT for smart cities applications

2. Comparison with Centralized Learning

The FL paradigm can be distinguished from important and related extremes and subtypes of distributed learning. FL can be viewed as a decentralized learning algorithm, a special case in the family of distributed ML algorithms, where the data is split across multiple nodes. This is a principal superset of FL, which can leverage communication-efficient algorithms allowing the participation of devices with limited storage and power resources. Distributed learning assumes that the training data is distributed across multiple nodes, each with local storage of a smaller data size than the total dataset. In practice, data is often partitioned in a way that is not i.i.d., which degrades performance. Additionally, FL differs from domain adaptation methods that divide data to perform local updates on heterogeneous datasets.

3.1. Definition and Principles

We define FL as a distributed AI model training paradigm emerging from parallel and collaborative local model training processes performed on edge devices featuring local data storage capability. Specifically, under the assumption of limited data availability, each edge device trains a local AI model on its device-specific data, in parallel and locally to the data storage, while relying on periodic communication and collaborative interaction with a remote controlling entity, commonly called the central server. The local model training processes at the edge devices enable a federated learning system to efficiently exploit the personalized knowledge embedded in the local datasets for training a more accurate global AI model deployed in the server and increasingly exploiting the benefits of decentralization, virtualization, and edge computing that characterize the current and next-generation IoT ecosystems. A major novelty of FL concerning traditional distributed AI model training approaches is the fact that local AI model training is performed with an approach that is less data-hungry, that is, to perform training on restricted device-specific datasets.

In FL, the AI model training process is a collaborative endeavor aiming at taking into account edge devices featuring diverse user/device/activity models by exploiting the availability of distributed localized data, while pursuing user privacy protection by limiting user data exchange and consolidating edge user model personalization at the device level. In addition to device-localized data, personalization of user models has also been investigated in the context of server-based training processes based on pooling together dataset fragments that have been created for specific user entities by one or more servers of terrestrial infrastructure. In this case, the servers, after performing privacy-preserving transfer learning processes, become selection bodies on which similar user models are consolidated into a common unit that is periodically refreshed to reduce the training drift over time, and where the edges perform model inference.

3.2. Comparison with Centralized Learning

Standard machine learning often relies on the centralized learning paradigm, in which a powerful model using statistically sufficient data is trained in a single location by aggregating data generated at multiple locations. Centralized deep learning of DNNs for many tasks has achieved impressive performance. However, it has notable challenges that inspired new approaches such as federated learning for edge intelligence in the context of IoT. The key limitations of centralized learning, from the viewpoint of IoT and smart cities, include the following.

1. Privacy and security of data. The DNN model in centralized learning is trained with the private, personal, and sensitive data of individual users. As the model is trained, there are inherent privacy risks of exposure to monitoring or adversarial parties, which may infer private patterns from the model behavior. Also, sensitive data in the cloud is vulnerable to data leaks by data breaches.
2. Bandwidth cost of user-cloud data upload. Data upload to the cloud involves the cost of transferring high-volume, continuous, raw sensor data. Mobile devices with high battery consumption may suffer from having to upload frequent, huge data files over wireless links to the cloud location. The high volume of data traffic may over-utilize the performance of a network whose bandwidth is limited for all users. The requirement for user-cloud data upload may lead to significant infrastructure costs.
3. Latency of model improvement. Centralized learning requires every user to wait for the completion of the cloud model training, which uses the batch data from all users, before getting an improved model for his/her local task. Users with fast-transient tasks that need quick response may find centralized learning infeasible.

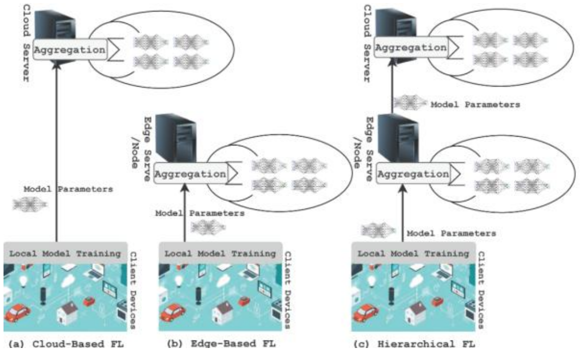
4. Edge Computing in IoT

Edge computing is considered a promising architectural paradigm to solve the unique challenges set forth by the IoT ecosystem, which generally consists of numerous resource-constrained devices emitting massive amounts of data in real-time. Conventional cloud computing, where all data from remote devices is sent to a centralized data center for processing and analysis, is unsatisfactory for some IoT applications that need very low latency since the centralized data center is typically far from the data source. Edge computing performs the computation tasks closer to the remote data sources by distributing the computation resources at the network edge, which provides great computational capabilities compared to resource-constrained IoT devices while also ensuring low communication latency and bandwidth burden.

An edge computing system is usually designed as a two-tier hierarchical framework consisting of cloud computing and edge computing for facilitation, where resource-constrained IoT devices are deployed at the first layer and edge servers are installed at the second layer. Traditionally, edge servers are composed of resource-rich devices like micro-datacenters, gateways, or service proxies installed at places relatively closer to the endpoint devices, such as base stations, access points, or even routers. Because the overhead of communication is much lower between local edge servers and remote IoT devices than between cloud centers and remote IoT devices at a larger distance, edge servers can perform tasks like filtering, analyzing, and processing the data coming from the IoT devices in the first place before relaying the processed results to the centralized cloud data center. By efficiently coordinating tasks between edge servers and cloud data centers, edge computing can provide major benefits for resource-constrained IoT systems, including low latency, less network burden, fault tolerance, and privacy preservation.

4.1. Architecture of Edge Computing

Edge computing is a paradigm that offers benefits such as reduced latency, local storage, and improved bandwidth for applications. It incorporates computing capabilities at the extreme ends of the network, close to the data source, and thus offers services that are partially cloud-like but also partially networking-centric and device-centric. The cloud controls everything but its reach, which is limited due to latency, privacy, and bandwidth issues, while local devices are connected through fast networks and handle their tasks without a cloud connection. The infrastructure for edge computing consists of micro data centers or small clusters of computers located near IoT devices, capable of performing localized AI functions and running both data-storing and data-processing analytics workflows, as well as distributed deep learning frameworks. This is a quite new and promising computing model.



(a) Cloud-Based FL (b) Edge-Based FL (c) Hierarchical FL

Fig 3 : Federated Learning in Edge Computing

As IoT systems come into maturity, their architectures move from the centralized server to a more distributed cloud-edge-IoT architecture, where edge and IoT devices share their processing load. Many applications, such as real and augmented virtual reality, smart cities, smart homes, self-driving cars, and ambient intelligence, will benefit from such an architecture, as it brings local intelligence closer to the user. Furthermore, workloads provided by IoT devices for inference at the edge are more predictive in time and space concerning workloads not derived from IoT sensors, which are mainly provided by desktop and mobile devices, such as laptops and smartphones. For example, for an augmented reality video frame being rendered by an edge or a cloud server, there will be in the next milliseconds several frames that will also need rendering. In this case, the last two predicted frames can be rendered ahead of time, or the rendering workload can be distributed and shared in parallel on several edge or cloud servers.

4.2. Benefits of IoT Systems

This section summarises the main benefits that Edge Computing brings to the design and implementation of IoT Systems. Concerning latency requirements, offloading the execution of some tasks from the cloud to the network edge reduces the delays introduced by network propagation delays, making user interactive services such as video surveillance responsive to user requests. The processing of data performed by the edge also allows for a more rapid response of the IoT-to-user interaction paths, which can be critical for time-sensitive applications.

In addition, Edge Computing reduces the bandwidth needed for the collection of data by the cloud platform provider, reducing cloud processing/storage costs and increasing the economic sustainability of the IoT. The network traffic volume reduction, together with the application of edge filtering techniques, also helps with scaling issues that could be raised by the synchronous access of large sets of IoT devices to cloud computing functions. Edge Computing eliminates the need to access the cloud if the application does not require centralized storage of data and when both IoT devices and end-users are physically close to each other, allowing for truly local data processing, without which many IoT applications would not be feasible. The continuity of connectivity between the IoT devices and the cloud is another important issue. In fact, for those applications where connectivity issues may arise, the offloading of some processing functions to the edge is an effective way to mask the connectivity issues. The download continues to support data processing at a local level and, if processing cannot be localized for any reason, the cloud can be used intermittently, transferring data only when connectivity is available.

5. Privacy-Preserving Techniques

Privacy has become essential in modern AI systems, especially in sensitive fields, such as healthcare. Privacy-preserving techniques are available at various stages of data life cycles, designed to conceal sensitive information in data, using trusted third parties to work with sensitive data, and using proper access controls and auditing. Generally, there are three major categories of techniques from the salvage and data analysis aspect: data encryption methods, differential privacy, and secure multi-party computing. Data encryption methods target to protect the content of data itself, while differential privacy and secure multi-party compute mechanisms aim to minimize the risk of revealing sensitive information about participants in a collective analysis.

Data Encryption Methods

Data encryption methods are the most primitive techniques. Their purpose is to make data unusable unless it is decrypted. As an example, symmetric-key cryptography methods, such as Data Encryption Standard and Advanced Encryption Standard, are widely used in practice. In a symmetric-key method, a secret key is shared by those parties allowed to access the content of data. Other parties cannot decrypt the content even after acquiring the data. However, in practice, there are two drawbacks to symmetric-key methods in the AI system context. The trust assumption on sharing a secret key is rather strong. For example, in the case of training an AI model using data from multiple sites, it is usually impossible for different data holders to share a symmetric key. Additionally, it is sometimes impossible to continue training the AI model when the model has been trained and a new site wants to join the process, as it is very difficult for the new site to have the same secret key. The second concern is that symmetric-key methods can only provide security for the data content and not the data context. Therefore, the AI model needs to be appropriately modified for different use cases.

Equation 2 : Privacy-Preserving Noise Addition (Differential Privacy)

$$\tilde{g}_k = g_k + \mathcal{N}(0, \sigma^2)$$

Where

$\tilde{g}_k$: Noisy Gradient Sent from Device k

g_k: Original Gradient

$\mathcal{N}(0, \sigma^2)$: Gaussian Noise for Privacy Protection

σ: Noise Scale Parameter

5.1. Data Encryption Methods

Encrypting data should be one of the first design decisions for a smart city project that includes IoT systems. Although the various drafts of the report do not include any data encryption best practices. Several commercial Clouds provide encryption solutions including hardware-level encryption. Other cloud services rely on service provider-level security without an option or data encryption. Therefore, projects in cities that are not utilizing clouds should provide their encryption policies. Data encryption solutions range from the simplest software-based file-level encryption to hardware-level keys providing the most sophisticated security. This latter solution does not allow anyone but the city to own the data and access the information. Software-level encryption constitutes the most widely applied scheme for anyone providing services around data generation or transmission in a less secure environment. New algorithms such as homomorphic encryption and fully homomorphic encryption enable data to be manipulated without decrypting it first. These methods use multiple keys which lessen the need for fast key access security. Encrypting the data flow is also critical to reducing the chances of a hacker stealing a key or some subsequent numerical values. Proper regulations require that data is properly hashed whenever it is transmitted despite the speed limitations it applies to the IoT edge. Other methods such as random insertions or hash masking are other solutions to speed up queries without exposing sensitive data. The hashing method is not reversible to recover the original value meaning that encrypted data is necessary for encrypted queries and a public data structure is not corruptible if it must be simultaneously maintained. This issue also applies to non-residential applications that women and men use to learn about urban aspects or aspects related to specific public or private entities.

5.2. Differential Privacy

Differential privacy enables trusted third parties to analyze data and extract useful information while assuring that the privacy of individuals in the dataset has not been compromised. Therefore, differential privacy guarantees that the presence or absence of a specific individual's privacy in the dataset does not have much effect on the output of the analysis. This technique can be applied to multiple domains, such as analysis of census data, public records, query-intensive services, medical studies, and so on. Its major feature is the use of a privacy budget, typically noted ε. It is a measure of how much privacy loss is tolerable. A small value of ε results in more privacy, while a large value of ε results in more accuracy.

Unlike traditional approaches in which privacy is ensured for an entire dataset, differential privacy provides individual-level privacy protection. In practical terms, the power of differential privacy is realized through a two-step process, which consists of perturbation followed by analysis. The key idea is that, before analyzing the data, the trusted third party perturbs the original data using a sensitive mechanism, usually a Laplace or Gaussian mechanism, to distort its original information. The perturber assumes that, with a certain probability, during the perturbation process an individual's original information will be added with noise, which discloses the individual's information, resulting in the risk of privacy violation. This risk can be controlled by a parameter ε. Adding the noise, however, also distorts the dataset, introducing errors into the analysis; this error is greater if ε is small.

Differential privacy is an attractive approach to performing statistical inference over private data, given its supply of strong privacy guarantees. More generally, differential privacy allows a trusted third party to generate differentially private data releases, based on collecting sensitive information from the users. These mechanisms can be put to many uses; for example, launching a differentially private statistical analysis at a trusted server.

6. Applications in Smart Cities

Smart cities heavily rely on the Internet of Things (IoT). In smart city paradigms, distributed sensors, camera systems, and their associated networks collect and analyze huge volumes of dynamic heterogeneous data and offer services that improve the quality of life for its inhabitants. Advancements in deep learning methods and developments in computer vision principles show promise in creating new opportunities for automation and the creation of intelligent systems that could provide benefits to numerous research and application areas, including public safety, traffic management, and energy services.

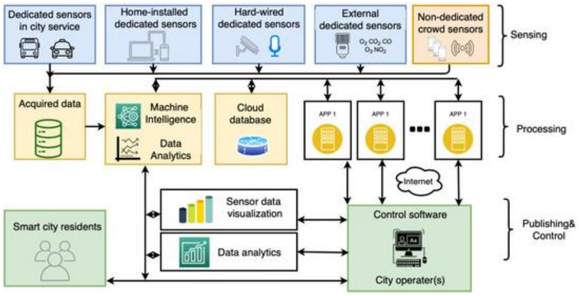

Dedicated sensors in city service
Home-installed dedicated sensors
Hard-wired dedicated sensors
External dedicated sensors
O_2 CO_2 CO O_3 NO_2
Non-dedicated crowd sensors
Sensing
Acquired data
Machine Intelligence
Data Analytics
Cloud database
APP 1
APP 1
APP 1
Processing
Internet
Sensor data visualization
Data analytics
Control software
City operater(s)
Smart city residents
Publishing& Control

Fig 4 : Federated Learning in Smart City Sensing

Existing smart city models rely heavily on always-on camera systems, which aggregate immense amounts of crowd data in centralized data repositories, resulting in poor performance due to lack of bandwidth, increased privacy risks, and compromised quality. Enabling an intelligent data processing approach at edge devices using federated edge intelligence will allow for real-time analytics capabilities, creating automatic systems for important public services, and only useful information will be centralized, while data privacy will be respected. We next describe several applications for traffic management, public safety, and energy use in smart cities where the intelligent edge will either have a strong impact or could change the way we use technology for these services.

Traffic management is one of the most popular applications of computer vision by creating systems to increase automobile, public transportation, and pedestrian traffic flow and safety by providing anonymized real-time information. New congestion and transit time patterns were observed during the pandemic. As cities reopen and restrictions are eased, it is expected that people will become more mindful of social distancing and many will choose to switch modes of travel, for example, avoid public transportation. Real-time use of intelligent edge services to manage and optimize traffic flow from traditional automobiles and new modes of alternative methods of travel such as ride-sharing and food delivery is an area ripe for new research.

6.1. Traffic Management Systems

Introduction to ITS and Traffic Surveillance at the Edge

Traffic is an inherent aspect of modern cities, and traffic patterns reveal important information about the operation of a city. Urban transportation relies heavily on roadways and vehicles, and one key feature of their operation is the fact that they are simple, yet present an amazing variety of behaviors, including travel to and from work, leisure and tourism, and transport of goods and services. Anomalies in the normal patterns of vehicles may also indicate relevant city-wide events, such as sporting events or other occasions that attract large crowds to specific locations. Infrastructure in intelligent transportation systems (ITS) enables the development of algorithms to monitor traffic, detect anomalies, model and forecast usual patterns, and analyze events when required. State-of-the-art ITS relies heavily on traffic surveillance applications that take advantage of the rapid development of intelligent camera sensors embedded in the environment. These surveillance cameras provide high-bandwidth images of multiple traffic objects and events, which allow the real-time detection of traffic objects, incident and accident detection, and consequent contingency measures.

However, current centralized surveillance systems are in most cases accountable neither for the protection of the privacy of oblivious users nor for the possible generation of huge amounts of traffic behind the scenes. Moreover, when traffic analysis is needed, it is still common practice to transfer all recorded data to workable processing and storing backend located in the cloud for analytics and inference purposes. Centralized approaches rely on the complete collection of data in the cloud for analysis and do not intrinsically protect the privacy of all the actors involved. Putting aside possible camera intrusiveness issues, these approaches typically need to collect data from the cloud backend for training and inference. Centralized processing is not capable of providing privacy preservation for ongoing activities, drawing attention to users identified within a camera's field of view who are not intentionally generating activities of interest.

6.2. Public Safety and Surveillance

The problems of public safety, crime, and social disorder have been a recurring theme in the development of modern cities. Technology plays an increasingly important role in designing solutions to improve an individual's feeling of personal safety in both public and private spaces. Smart cities embrace security and surveillance technologies that capture video and audio to analyze potential security breaches, the discharging of firearms, and speech about violence and aggression - the contrary to peace. These technologies require advanced artificial intelligence capabilities to detect the problems for local authorities to be alerted in real-time, rather than simply storing huge volumes of video and audio data for possible analysis at a later time. AI and machine learning are now providing the ability to identify people and vehicles of interest, and then continually track the movement of the individuals, with multi-modal data sources providing multiple forms of sensor fusion to enhance the data accuracy.

Privacy issues of biometrically recognizing individuals that may be involved in malicious or criminal activity mean that locally processing these sensors, especially video cameras, generates large volumes of video data. Federated edge intelligence allows local AI to be performed against forecasted analysis models to reduce the amount of video sent to AI services in other federated nodes. This enables privacy-preserving video analysis to be performed at the edge of the urban environment. Less some challenges of few-shot learning when new facial biometric databases of suspected criminals need to be incorporated if the individual has not had prior documented interactions with authorities. Federated edge intelligence platforms can perform video analysis in public spaces, keeping the video and images of potentially implicated individuals secure and private.

6.3. Energy Management

Smart cities are examples of complex cyber-physical systems requiring a high level of interaction between physical processes and computational resources performed on the edge. Furthermore, the distributed nature of IoT-based smart cities mandates an on-device intelligence to guarantee fast response to system dynamics and limited bandwidth usage. Edge Intelligence enables Core and Edge Clouds design and development by offering a unique platform-agnostic approach to dealing with such a distributed architecture.

When it comes to power and energy management in smart city scenarios empowered by modern Edge Cloud architectures, powering modern cities with clean energy sources while ensuring the safety of vulnerable populations is a major challenge. In this regard, various contributions model energy exchanges among electricity customers as products sold and purchased at different prices, creating market-like dynamics. The design of price models with spike, decreasing, and increasing effects, is a hot research topic in this area. Moreover, dynamically and appropriately varying the prices of energy given to customers by a power company offers an appealing alternative to locally manage and control the energy to be exchanged, without massive investments in energy infrastructures by the power company. Then, smart power companies could accumulate to their customer portfolio different groups of customers sharing a common objective.

7. Challenges and Limitations

As with any AI-based system, FEdI cannot be employed in a solution without correctly recognizing and assessing its limitations concerning others, being either centralized or FC-based. We consider that three key aspects are affecting the FEdI model: data heterogeneity, scalability issues, and latency concerns, due to the Federated Learning model being based on the assumption that nodes are homogeneous. Additionally, many devices have limits to the amount of data that can be processed and stored in memory at a given time. This may limit the ability of federated learning to converge to a global model if the required communication between nodes and the FEdI model owner cannot be realized, or, at least, requires additional communication between peers. Different devices may also have a different capability when it comes to participating in the model aggregation phase, due to the models being trained differently at each node with different amounts of data. However, it should be acknowledged that some of the limitations derived from aggregating communication from heterogeneous devices with Quality-of-service considerations may undergo amelioration by generalizing models to account for several classes of nodes rather than specific classes.

Concerning more extreme heterogeneity scenarios, it must also be said that aggregating parametrized models trained with FL at some nodes may not be desired when the FL process is thought at different hierarchical levels and FEdI model owners at some levels assume different operational conditions. However, for these more difficult scenarios, centralized learning may not be a reliable option either, due to the enormous data amount coming from the IoT systems permeating smart city operations and the privacy of private enterprises providing data for the smart city ecologies. In these extreme cases, decentralized learning could be considered as a solution, although it has still many open research issues that have not been solved yet.

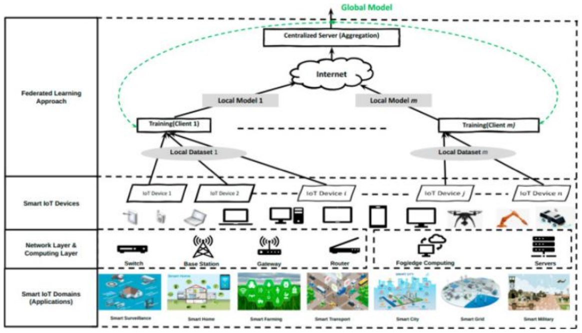


Fig 5 : Security of federated learning with IoT systems: Issues, limitations, challenges, and solutions

7.1. Data Heterogeneity

Federated Edge Intelligence is envisioned to aggregate decentralized Internet of Things (IoT) systems with edge intelligence, creating platforms for addressing the challenges encountered in Smart Cities. However, the integration of numerous data sources is hampered by the heterogeneity associated with the nature and characteristics of collected data, particularly its structure and representation. For example, in a road traffic monitoring system, environmental sensors collect diverse data indicative of varied traffic conditions, such as CO levels, humidity, temperature, particulate matter, or wind intensity; highway cameras capture images or video sequences depicting undesirable scenarios; connected vehicles report accounts revealing information on breakdowns or accidents; and social media posts and geolocalized tweets raise alerts regarding traffic or indicate road closures. The input parameters used, and data formats adopted by each IoT system are distinct, leading to a non-homogeneous distributed data environment across the Edge. The aggregation process, thus, confronts semantic and syntactic mismatches during model development.

Creating an Edge IoT-enabled ecosystem is important, as the contribution of each IoT system can vary over time due to the specific traffic conditions and disorder possibilities. For example, during an emergency scenario, social media would provide valuable knowledge; while in the absence of any accidents, environmental sensors or connected vehicle accounts could be the main contributors. The local data associated with the IoT systems are rare as any disorder scenario comprises only a small number of events compared with the overall traffic volume. As a consequence, the diversity of the minority data classes associated with these systems in a global model acts together as very limited and critical support for the learning process, being able to affect and undermine the global model’s performance. Thus, to ensure the overall activity of the edge-based intelligence system, the global framework must be capable of dealing with the model learning process, even in the presence of limited or unavailable data from a particular model contributor.

7.2. Scalability Issues

Scalability represents a cardinal attribute for the increasingly deployed Edge and Fog Intelligence Computing paradigm. Although data generated and transmitted from smart city and IoT systems may accelerate the development of a myriad of AI models to be applied in a multitude of uncorrelated domains, while requesting a budget on a large scale, the online training and plan execution of resulting models remains a challenge for scalable deployment of the Edge and Fog, increasing and creating overwhelming traffic, congestion, latency and additional load in the smart city sensor setup. Our recent works address this model scalability by harnessing Edge supervision and edge-to-cloud cooperation. Our ideas fall into two categories: (1) Edge supervision of federated distillation: we address the communication bottlenecks of collaborative learning by deploying a local model in the computing node or edge access point shortly supporting some IoT clients. Supported clients upload feature and label batches, which are used by the local model for one-step training. Local capability is thus optimally increased without soliciting the upload of the heavy full gradients, and (2) Dynamic edge network: we deploy a hierarchical cloud-edge network, where different edge organizations (at different tiers) with different capabilities coexist. Other clusters of motes, mobile nodes with insufficient capability or with stringent battery timings can seek support from nearby more powerful participating clusters or the edge access point by switching their transmission mode during some sessions or clusters.

7.3. Latency Concerns

Federated learning relies on transferring local model updates periodically, which incurs high latency, especially with large numbers of local updates. The communication cost of uploading updates typically throttles the overall task execution time in a federated learning setting. Moreover, the expected latency of FL may be unacceptable in some real-life applications. To mitigate this issue, existing work primarily focuses on reducing the size and frequency of model updates sent to the coordinating entity from the local nodes. Compression techniques, such as low-precision update quantization or sparsification, either decrease the update rate or reduce its scale. For example, local updates via quantized gradients also need to have a small enough mean square error; otherwise, the global model can diverge. In addition, once nodes transmit their local models to the global coordinator, the models need to be aggregated, which adds further latency. To limit the burden on the coordinated entity, an update-pruning scheme can be used, so that the global model can be fine-tuned at a reduced cost. However, pruning also tends to worsen the obtained model’s accuracy. Adaptive aggregation rules could help. However, it is not clear how or when to optimally use them or if it is reasonable to assume that the nodes have constant performance during the entire procedure.

Furthermore, previous studies on federated learning have mainly concentrated on minimizing expected latency based on the size of the communication and training, cache reuse, and sparse group lasso model representations. In contrast, it is often more helpful to look at the worst correspondence necessary latency to facilitate given learning and statistical accuracy guarantees, in particular, assuming non-identically distributed covariates at the different nodes or assuming some agents are sending updates much more likely than others, e.g., to aggregate data from non-ergodic edges. There has been little work on this aspect. Moreover, beyond the aforementioned expected game-theoretic results, there have been no explorations of coordination policies, such as explicitly modeling interaction towards distributed training via zero-sum games.

8. Future Directions

This chapter discussed how Federated Edge Intelligence provides AI and ML-based automatic services that millions of users in smart cities and IoT systems would use. In comparison to conventional cloud computing services, FEdI preserves privacy by design, scales efficiently in response to user demand, and is implemented on highly heterogeneous edge environments. However, we can envision many more potential applications of FEdI, as well as how it can be integrated with future advances in AI infrastructures and urban technology environments. Therefore, in this chapter, we will discuss some of these future directions, including enhanced 5G network integration, protocols that provide more privacy, security, and heterogeneous capabilities for FEdI, and its integration with future AI algorithmic advances.

As wireless communication keeps evolving, new generations of networks are capable of handling and transferring data at an ever-increasing pace. Although it currently serves mostly consumer and mobile service with applications such as video on demand and augmented reality sound and data transfer capabilities, 5G will enable new opportunities in business, urban environments, and industrial services. Besides private networking in the form of campus-local 5G networks, network slicing will allow for multipurpose business and urban services to be executed with different quality of service guarantees. The advancements in the underlying physical layer will enable improved services for datagram transmission, latency, and jitter performance, which will enable IoT for mission-critical applications. Such are emergency response, autonomous driving such as V2X, and remote robotic controlling. Therefore, FEdI services can benefit these mission-critical applications requiring real-time responsiveness and guaranteed communication channels in the design of their underlying algorithms. Furthermore, one can envision novel FEdI services and algorithms that automatically cross-trigger FEdI algorithms or mesh together multiple FedI services in the case of some triggered event or specific conditions.

8.1. Integration with 5G Networks

In this section, we will explore some future research directions regarding federated edge intelligence, with a specific focus on 5G integration. Federated edge intelligence relies on low-latency, high-bandwidth communication among IoT edge nodes for its performance in a smart city scenario. With recent advancements in smart city planning, such as low-power link and backhaul infrastructure support, as well as the use of millimeter wave or free-space optics for high-throughput low-latency data exchange, 5G communication can be easily integrated into federated learning. Specifically, how would the need for energy-efficient IoT edge devices, utilizing advanced compression approaches and optimized communication schedules with the federated server, fit into the basic operation of federated learning and distributed optimization in general? How would the network bandwidth and resource allocation uncertainty impact the design and robustness of federated edge intelligence?

As an important piece of future communications technology, 5G is expected to enable the development of smart city services. Indeed, a large body of relevant studies demonstrates the ability of small cells and millimeter waves to solve capacity and access network backhauling and reliability constraints faced by communication on which advanced services, such as smart city services, are built. Such services will require ultra-low latency, low power, and ultra-high reliability in created smart city environments. Examples of commercial deployments and testing can be found addressing traditional communication services, such as public safety, video streaming, and gaming, as well as advanced ambient intelligence and/or automation functions enabled by human-human, human-machine, and machine-machine interactions.

8.2. Advancements in AI Algorithms

We discuss some fundamental advancements in AI algorithms for FEdI. FEdI follows a centralized approach for model aggregation. During model aggregation, to achieve better ensemble, FEdI must use advanced AI algorithms that can work with and control the diverse models from the FL process.

FEdI stores multiple models for different AI tasks. Federated Meta and Few-Shot Learning are promising AI toolboxes to deal with extreme distributed Inference–one for few-shot learning from novel users, and the other for learning task distributions from a meta-level. We expect FEdI will realize the potential power in FL–both for transferring and sharing knowledge across tasks/users.

A famous AI principle says: "Over-parametrization of neural networks is necessary but not sufficient". Deep models must also learn to quantify the prediction uncertainty, e.g., as in probabilistic deep models. This property is strongly related to human's intuitive learning process in the children's developmental stage, which realizes the physical laws that are invariant across space and time and uses these laws to guide their learning process to limit the function spaces of the observed uncertainties. Learning predictive uncertainties can not only control overfitting Curves but also give precious insights into negative (distrustful) prediction outputs. FEdI's multi-model within a single-edge device boosts the demands for uncertainty understanding and for supporting end-to-end inference control.

AI consists of a number of various cognitive modalities. For edge AI systems, the AI engine in FEdI will need to support various modalities (both in the training of such models and the deployment). For example, data pre-processing and low-level data understanding (video/image/event/audio synthesis) are the basis of the agile promotion of high-level decision AI modeling. The close-cooperation of these two phases is natural. The information provided by one phase can help improve outcomes in the latter phase.

8.3. Policy and Regulation Implications

Smart cities present a new set of complexities for top-down governance models, including issues of equitability, transparency, and security. Multiple stakeholders participate in the generation of urban data, whose interpretation and utilization must be ensured to develop services that are meaningful for residents and improve their well-being. The partnerships between the private and public sectors that are a feature of many smart city projects can lead to a lack of accountability when it comes to the ethical rules that govern data analyses and utilization, as well as skewing data interpretations for a profit-based agenda. Prolonged overload of monitoring is associated with negative effects for city residents yet there are no established cards or auditors for smart city projects, nor are there standardized operational procedures. Some smart city projects may contribute to gentrification processes or be unilaterally focused on efficiency and convenience for users at the expense of impacts in terms of pollution and crime. We argue that the emphasis on the public good posed by the smart city postulate requires building the capacity to respond to or monitor such unexpected consequences, as well as producing data usage diffusion policies.

The fact that such externalities are often difficult to manage calls for resources to be set aside to address negative impacts after the fact. If privatized services seek to offset the costs of these potential negative effects on cities' taxpayers through profit- and efficiency-motivated crunching of services, tightly binding long-term concession contracts must be publicly debated to build user trust. We argue that the emergence of edge intelligence opens a new avenue addressing the trade-offs between differential public and private data usage regulations versus the collective resonance and wellbeing effects associated with local decision. New accountability frameworks may also be needed to expand the fundamental rights established by data protection and digital privacy regulation to cover other areas of the digital economy and society, such as public sector services and public goods, where such risks may be heightened.

Equation 3 : Edge Utility Optimization

$$U_e = \frac{A_m}{T_c + \lambda \cdot P_e}$$

Where

U_e: Edge Utility Score

A_m: Model Accuracy

T_c: Communication Time

P_e: Edge Power Usage

λ: Weight Balancing Energy vs Performance

9. Conclusion

Federated learning fills an important yet frequently encountered gap—data annotation. Indeed, while a large amount of unlabeled data is available, it is necessary to transfer a small amount of data to an annotation provider to train a suitable deep neural network for any downstream tasks. Knowledge distillation has been provided as federated model personalization from a central server, but has faced privacy and availability problems: (i) privacy would be compromised if trained models are shared with the central server, which is unavoidable under knowledge distillation; and (ii) model updates can be very large in dimension, which breach the consent of vulnerable users. Furthermore, the ability of any human annotator is limited to only a few domains, but it is possible to lend the expertise of a few trusted annotators. Fully decentralized federated learning has not yet accommodated data inherent to cross-domain federated learning tasks, possibly due to the complexity of the communication networks.

Additionally, the pros of the collaborative model update sharing mechanism for decentralized federated learning of cross-domain tasks over the cons of a small trusted group of annotators providing high-quality annotations for careless human labelers are clear, and collaborative domain-competent annotators will address this need. Moreover, we have connected our method to federated learning and diffuse information. We have shown that it is possible to benefit from collaborative model updates in the presence of distributed-related distributions while fully alleviating privacy problems, and have proposed a PAC-like scenario wherein available models serve as human annotators, paving the way for future investigations into hybrid human/human annotator models and human/machine interactivity. Our method is flexible, allowing available models that are different or similar to the creating model for annotation.

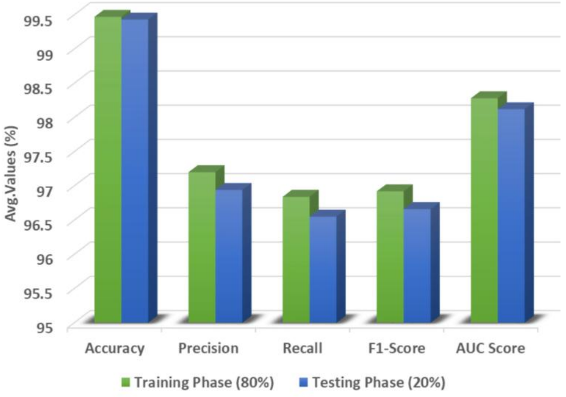


Fig 6 : IoT-assisted sustainable smart cities

9.1. Final Thoughts and Key Takeaways

In this book, we explored the notion of Federated Edge Intelligence and how it contributes to the next generation of intelligent edge computing within Smart Cities and the Internet of Things (IoT) systems and applications. We explained how Federated Edge Intelligence enables collaborative learning of machine and deep learning models over distributed and edge-constrained IoT devices without disclosing the privacy-sensitive data it is locally generated; and, importantly, how Federated Edge Intelligence can preserve the user and societal privacy while enabling the automation and optimization of processes and decision-making within Smart Cities and IoT. To that end, we first provided the key technical details of Federated Edge Intelligence. Then, we discussed approaches on how techniques, solutions, and frameworks designed for Federated Edge Intelligence could benefit intelligent edge computing and bridge its research gaps with Federated Learning.

We concluded the book with a set of practical examples of how Federated Edge Intelligence could harness the optimization and automation of processes and information management in several use cases within Smart Cities and IoT Systems. These use cases included how Federated Edge Intelligence can monitor critical infrastructures at the city level and optimize people's mobility, and, at the same time, optimize air quality at the city level; automate the detection of construction workers' safety violations; facilitate the development of better human-centric smart eyeglasses; facilitate privacy-preserving smart medical monitoring systems; facilitate the development of effective and actionable contact tracing and health monitoring systems; help with privacy-preserving facial recognition; help avoid data hoarding attacks in smart critical infrastructures; and finally, facilitate the management of vulnerability in the local edge devices that are situated within Smart Cities and IoT Systems.

10. References

[1] Polineni, T. N. S., Ganti, V. K. A. T., Maguluri, K. K., & Rani, P. S. (2024). AI-Driven Analysis of Lifestyle Patterns for Early Detection of Metabolic Disorders. Journal of Computational Analysis and Applications, 33(8).

[2] Sondinti, K., & Reddy, L. (2024). Financial Optimization in the Automotive Industry: Leveraging Cloud-Driven Big Data and AI for Cost Reduction and Revenue Growth. Financial Optimization in the Automotive Industry: Leveraging Cloud-Driven Big Data and AI for Cost Reduction and Revenue Growth (December 17, 2024)

[3] Sambasiva Rao Suura. (2024). Integrating Generative AI into Non-Invasive Genetic Testing: Enhancing Early Detection and Risk Assessment. Utilitas Mathematica, 121, 510–522. Retrieved from https://utilitasmathematica.com/index.php/Index/article/view/2046

[4] Venkata Narasareddy Annapareddy. (2024). Harnessing AI Neural Networks and Generative AI for Optimized Solar Energy Production and Residential Battery Storage Management. Utilitas Mathematica, 121, 501–509.Retrievedhttps://utilitasmathematica.com/index.php/Index/article/view/2045

[5] Harish Kumar Sriram. (2024). Leveraging AI and Machine Learning for Enhancing Secure Payment Processing: A Study on Generative AI Applications in Real-Time Fraud Detection and Prevention. Utilitas Mathematica, 121, 535–546. Retrieved from https://utilitasmathematica.com/index.php/Index/article/view/2048

[6] Karthik Chava. (2024). Harnessing Generative AI for Transformative Innovations in Healthcare Logistics: A Neural Network Framework for Intelligent Sample Management.

Utilitas Mathematica, 121, 547–558. Retrieved from https://utilitasmathematica.com/index.php/Index/article/view/2049

[7] Komaragiri, V. B. Harnessing AI Neural Networks and Generative AI for the Evolution of Digital Inclusion: Transformative Approaches to Bridging the Global Connectivity Divide

[8] Chaitran Chakilam. (2024). Revolutionizing Genetic Therapy Delivery: A Comprehensive Study on AI Neural Networks for Predictive Patient Support Systems in Rare Disease Management. Utilitas Mathematica, 121, 569–579. Retrieved from https://utilitasmathematica.com/index.php/Index/article/view/2051

[9] Murali Malempati. (2024). Generative AI-Driven Innovation in Digital Identity Verification: Leveraging Neural Networks for Next-Generation Financial Security. Utilitas Mathematica, 121, 580–592. Retrieved from https://utilitasmathematica.com/index.php/Index/article/view/2052

[20] Challa, K. (2024). Artificial Intelligence and Generative Neural Systems: Creating Smarter Customer Support Models for Digital Financial Services. Journal of Computational Analysis & Applications, 33(8).

[21] Nuka, S. T. (2024). Exploring AI and Generative AI in Healthcare Reimbursement Policies: Challenges, Ethical Considerations, and Future Innovations. International Journal of Medical Toxicology and Legal Medicine, 27(5), 574-584.

[22] Burugulla, J. K. R. (2024). The Future of Digital Financial Security: Integrating AI, Cloud, and Big Data for Fraud Prevention and Real Time Transaction Monitoring in Payment Systems. MSW Management Journal, 34(2), 711-730.

[23] Intelligent Supply Chain Optimization: AI Driven Data Synchronization and Decision Making for Modern Logistics. (2024). MSW Management Journal, 34(2), 804-817.

[24] Pamisetty, V. (2024). AI Powered Decision Support Systems in Government Financial Management: Transforming Policy Implementation and Fiscal Responsibility. Journal of Computational Analysis & Applications, 33(8).

[21] Revolutionizing Automotive Manufacturing with AI-Driven Data Engineering: Enhancing Production Efficiency through Advanced Data Analytics and Cloud Integration . (2024). MSW Management Journal, 34(2), 900-923.

[22] Leveraging Deep Learning, Neural Networks, and Data Engineering for Intelligent Mortgage Loan Validation: A Data-Driven Approach to Automating Borrower Income, Employment, and Asset Verification. (2024). MSW Management Journal, 34(2), 924-945.

[23] Lahari Pandiri, Subrahmanyasarma Chitta. (2024). Machine Learning-Powered Actuarial Science: Revolutionizing Underwriting and Policy Pricing for Enhanced Predictive Analytics in Life and Health Insurance . South Eastern European Journal of Public Health, 3396–3417. https://doi.org/10.70135/seejph.vi.5903

[24] Mahesh Recharla, (2024). The Role of Agentic AI in Next-Generation Drug Discovery and Automated Pharmacovigilance for Rare and Neurological Diseases. Frontiers in Health Informatics, Vol. 13(8), 4999-5014

[25] Botlagunta Preethish Nandan. (2024). Revolutionizing Semiconductor Chip Design through Generative AI and Reinforcement Learning: A Novel Approach to Mask Patterning and Resolution Enhancement. International Journal of Medical Toxicology and Legal Medicine, 27(5), 759–772. https://doi.org/10.47059/ijmtlm/V27I5/096

[26] Challa, S. R., Challa, K., Lakkarasu, P., Sriram, H. K., & Adusupalli, B. (2024). Strategic Financial Growth: Strengthening Investment Management, Secure Transactions, and Risk Protection in the Digital Era. Journal of Artificial Intelligence and Big Data Disciplines, 1(1), 97-108.

[27] Intelligent Technologies for Modern Financial Ecosystems: Transforming Housing Finance, Risk Management, and Advisory Services Through Advanced Analytics and Secure Cloud Solutions. (2024). MSW Management Journal, 34(2), 953-971.

[28] Pallav Kumar Kaulwar,. (2024). Agentic Tax Intelligence: Designing Autonomous AI Advisors for Real-Time Tax Consulting and Compliance. Journal of Computational Analysis and Applications (JoCAAA), 33(08), 2757–2775. Retrieved from https://eudoxuspress.com/index.php/pub/article/view/2224

[29] AI-Powered Revenue Management and Monetization: A Data Engineering Framework for Scalable Billing Systems in the Digital Economy . (2024). MSW Management Journal, 34(2), 776-787.

[30] Paleti, S., Pamisetty, V., Challa, K., Burugulla, J. K. R., & Dodda, A. (2024). Innovative Intelligence Solutions for Secure Financial Management: Optimizing Regulatory Compliance, Transaction Security, and Digital Payment Frameworks Through Advanced Computational Models. Journal of Artificial Intelligence and Big Data Disciplines, 1(1), 125-136.

[31] Singireddy, J. (2024). Deep Learning Architectures for Automated Fraud Detection in Payroll and Financial Management Services: Towards Safer Small Business Transactions. Journal of Artificial Intelligence and Big Data Disciplines, 1(1), 75-85.

[32] Sneha Singireddy. (2024). Leveraging Artificial Intelligence and Agentic AI Models for Personalized Risk Assessment and Policy Customization in the Modern Insurance Industry: A Case Study on Customer-Centric Service Innovations . Journal of Computational Analysis and Applications (JoCAAA), 33(08), 2532–2545. Retrieved from https://eudoxuspress.com/index.php/pub/article/view/2163

[33] Challa, S. R. (2024). Behavioral Finance in Financial Advisory Services: Analyzing Investor DecisionMaking and Risk Management in Wealth Accumulation. Available at SSRN 5135949.

[34] Maguluri, K. K., Ganti, V. K. A. T., & Subhash, T. N. (2024). Advancing Patient Privacy in the Era of Artificial Intelligence: A Deep Learning Approach to Ensuring Compliance with HIPAA and Addressing Ethical Challenges in Healthcare Data Security. International Journal of Medical Toxicology & Legal Medicine, 27(5).

[35] Danda, R. R., Nampalli, R. C. R., Sondinti, L. R. K., Vankayalapati, R. K., Syed, S., Maguluri, K. K., & Yasmeen, Z. (2024). Harnessing Big Data and AI in Cloud-Powered Financial Decision-Making for Automotive and Healthcare Industries: A Comparative Analysis of Risk Management and Profit Optimization.

[36] Suura, S. R. (2024). Generative AI Frameworks for Precision Carrier Screening: Transforming Genetic Testing in Reproductive Health. Frontiers in Health Informa, 4050-4069.

[37] Annapareddy, V. N., & Sudha Rani, P. (2024). AI and ML Applications in RealTime Energy Monitoring and Optimization for Residential Solar Power Systems. Available at SSRN 5116062

[38] Kannan, S., & Seenu, A. (2024). Advancing Sustainability Goals with AI Neural Networks: A Study on Machine Learning Integration for Resource Optimization and Environmental Impact

[39] Chava, K., & Saradhi, K. S. (2024). Emerging Applications of Generative AI and Deep Neural Networks in Modern Pharmaceutical Supply Chains: A Focus on Automated Insights and Decision-Making

[40] Komaragiri, V. B. (2024). Generative AI-Powered Service Operating Systems: A Comprehensive Study of Neural Network Applications for Intelligent Data Management and Service Optimization. Journal of Computational Analysis & Applications, 33(8).

[41] Chakilam, C., & Seenu, D. A. (2024). Transformative Applications of AI and ML in Personalized Treatment Pathways: Enhancing Rare Disease Support Through Advanced Neural Networks. Frontiers in Health Informa, 4032-4049..

[43] Malempati, M. (2024). Leveraging cloud computing architectures to enhance scalability and security in modern financial services and payment infrastructure. European Advanced Journal for Science & Engineering (EAJSE)-p-ISSN 3050-9696 en e-ISSN 3050-970X, 1(1).

[44] Nuka, S. T. (2024). The Future of AI Enabled Medical Device Engineering: Integrating Predictive Analytics, Regulatory Automation, and Intelligent Manufacturing. MSW Management Journal, 34(2), 731-748.

[55] Singireddy, S., Adusupalli, B., Pamisetty, A., Mashetty, S., & Kaulwar, P. K. (2024). Redefining Financial Risk Strategies: The Integration of Smart Automation, Secure Access Systems, and Predictive Intelligence in Insurance, Lending, and Asset Management. Journal of Artificial Intelligence and Big Data Disciplines, 1(1), 109-124.

[46] Kalisetty, S., & Lakkarasu, P. (2024). Deep Learning Frameworks for Multi-Modal Data Fusion in Retail Supply Chains: Enhancing Forecast Accuracy and Agility. Journal of Artificial Intelligence and Big Data Disciplines, 1(1), 137-148.

[47] Venkata Krishna Azith Teja Ganti ,Kiran Kumar Maguluri ,Dr. P.R. Sudha Rani (2024). Neural Network Applications in Understanding Neurodegenerative Disease Progression. Frontiers in HealthInformatics, 13 (8) 471-485

[48] Venkatasubramanian, K., Yasmeen, Z., Reddy Kothapalli Sondinti, L., Valiki, S., Tejpal, S., & Paulraj, K. (2024). Unified Deep Learning Framework Integrating CNNs and Vision Transformers for Efficient and Scalable Solutions. Available at SSRN 5077827.

[49] Sambasiva Rao Suura. (2024). Artificial Intelligence and Machine Learning in Genomic Medicine: Redefining the Future of Precision Diagnostics. South Eastern European Journal of Public Health, 955–973. https://doi.org/10.70135/seejph.vi.4602

[50] Satyasree, K. P. N. V., & Kothpalli Sondinti, L. R. (2024). Mitigating Financial Fraud and Cybercrime in Financial Services with Security Protocols and Risk Management Strategies. Computer Fraud and Security, 2024(11).

[51] Suura, S. R. (2024). The role of neural networks in predicting genetic risks and enhancing preventive health strategies. European Advanced Journal for Emerging Technologies (EAJET)-p-ISSN 3050-9734 en e-ISSN 3050-9742, 1(1).

[52] A comparative study of identity theft protection frameworks enhanced by machine learning algorithms. (2024). MSW Management Journal, 34(2), 1080-1101.

[53] Komaragiri, V. B. (2024). Data-Driven Approaches to Battery Health Monitoring in Electric Vehicles Using Machine Learning. International Journal of Scientific Research and Management (IJSRM), 12(01), 1018-1037.

[54] Reddy, J. K. (2024). Leveraging Generative AI for Hyper Personalized Rewards and Benefits Programs: Analyzing Consumer Behavior in Financial Loyalty Systems. J. Electrical Systems, 20(11s), 3647-3657.

[55] Singireddy, S., Adusupalli, B., Pamisetty, A., Mashetty, S., & Kaulwar, P. K. (2024). Redefining Financial Risk Strategies: The Integration of Smart Automation, Secure Access Systems, and Predictive Intelligence in Insurance, Lending, and Asset Management. Journal of Artificial Intelligence and Big Data Disciplines, 1(1), 109-124.

[56] Kalisetty, S., Pandugula, C., Sondinti, L. R. K., Mallesham, G., & Rani, P. S. (2024). AI-Driven Fraud Detection Systems: Enhancing Security in Card-Based Transactions Using Real-Time Analytics. Journal of Electrical Systems, 20, 1452-1464.

[54] Suura, S. R. (2024). Agentic artificial intelligence systems for dynamic health management and real-time genomic data analysis. European Journal of Analytics and Artificial Intelligence (EJAAI) p-ISSN 3050-9556 en e-ISSN 3050-9564, 1(1).

[55] Komaragiri, V. B., Edward, A., & Surabhi, S. N. R. D. Enhancing Ethernet Log Interpretation And Visualization

[57] Challa, K. (2024). Neural Networks in Inclusive Financial Systems: Generative AI for Bridging the Gap Between Technology and Socioeconomic Equity. MSW Management Journal, 34(2), 749-763.

[58] Moore, C., & Routhu, K. (2023). Leveraging Machine Learning Techniques for Predictive Analysis in Merger and Acquisition (M&A). Available at SSRN 5103189.

[59] Moore, C. (2023). AI-powered big data and ERP systems for autonomous detection of cybersecurity vulnerabilities. Nanotechnology Perceptions, 19, 46-64.

[60] Chinta, P. C. R., Katnapally, N., Ja, K., Bodepudi, V., Babu, S., & Boppana, M. S. (2022). Exploring the role of neural networks in big data-driven ERP systems for proactive cybersecurity management. Kurdish Studies.

[61] Katnapally, N., Chinta, P. C. R., Routhu, K. K., Velaga, V., Bodepudi, V., & Karaka, L. M. (2021). Leveraging Big Data Analytics and Machine Learning Techniques for Sentiment Analysis of Amazon Product Reviews in Business Insights. American Journal of Computing and Engineering, 4(2), 35-51.

[62] Maka, S. R. (2023). Understanding the Fundamentals of Digital Transformation in Financial Services: Drivers and Strategic Insights. Available at SSRN 5116707.

[63] Krishna Madhav, J., Varun, B., Niharika, K., Srinivasa Rao, M., & Laxmana Murthy, K. (2023). Optimising Sales Forecasts in ERP Systems Using Machine Learning and Predictive Analytics. J Contemp Edu Theo Artific Intel: JCETAI-104.