

Secure Data Engineering Pipelines For Federated Insurance AI: Balancing Privacy, Speed, And Intelligence

Balaji Adusupalli

Abstract

The insurance underwriting process assesses potential policies, and assigns risk ratings and premiums. In recent years, actuaries have become increasingly interested in leveraging machine learning algorithms to improve existing underwriting models that rely primarily on demographic data such as age, gender, and location for insurance applications such as auto, riders, and renters. Machine learning algorithms can be customized to include a wider array of potentially predictive features and can be more flexible in capturing non-linear relationships between these predictive features and policy loss risk. Moreover, technological advances now also make it feasible to use machine learning algorithms for the underwriting process itself and not only during the rate evaluation.

Traditional machine learning techniques train on pooled historical data. Federated learning is a technique to train machine learning models on separate, private data silos. Training data remains on the devices while information on the model is transferred. In federated learning, this is repeated until sufficient model performance is achieved. ¹Beyond privacy, the training of a machine learning model can also be a matter of speed and intelligence. A federated system that balances data privacy, speed, and intelligence is federated insurance AI. However, federated insurance AI has distinct implementation challenges around performance and security. We aim to address secure data engineering pipelines for practical federated insurance AI. The key is rapid, high-security transformations of data.

Keywords: *Insurance Underwriting, Risk Assessment, Premium Calculation, Machine Learning, Actuarial Science, Predictive Features, Non-Linear Relationships, Policy Loss Risk, Rate Evaluation, Federated Learning, Data Privacy, Distributed Training, Model Performance, Data Silos, Secure Data Engineering, Federated Insurance AI, Security Challenges, High-Security Transformations, Speed Optimization, Insurance Applications.*

1. Introduction

The lion's share of AI research and development lies in training models on vast datasets to deploy these models to perform useful tasks such as predicting the future or recommending good courses of action. However, opening sensitive data up for training models can come at a steep privacy price. Without mature strategies for secure deployment, it can seem that

combining the many benefits of centralization and breadth of training with the privacy and compliance benefits of local ownership of data is impossible.

We propose the concept of a Federated Insurance Data Engineering Pipeline as the natural solution to this seemingly irresolvable tension around privacy. It provides a blend of the necessary ingredients: centralized leadership about what the models should look like and how trusted data synthesis processes can be used to adjust privacy and learning weights, which are nudged to evolve in the direction the centralized trainers believe is most useful. However, this approach minimizes the risk of centralization in two crucial ways. Each data host can always implement privacy strategies entirely independent of the global ones. But in any case, the flowing back and forth of enriched inputs and super compressed derivative trust metrics eliminate most concerns about centralized data control. With this relatively lightweight and easy-to-understand understanding of how trust and correctness will flow throughout the federated pipeline, it's easy to adjust any unfair optimization syndrome while models are being developed. Data owners can thus be comfortable providing the kind of feed to central processing samplers that would normally be designed into a standard centralized training pipeline.

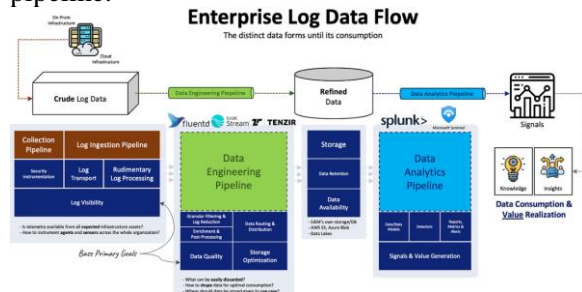


Fig 1 : Data Engineering Pipelines before an enterprise

1.1. Overview of the Study's Scope and Objectives

In this study, we provide and apply a secure data engineering pipeline to enable the creation of insurance AI models under a federated AI environment. A federated AI learning environment is where different local AI models, located in possibly different physical locations, learn only from their local datasets, and the models are aggregated together to build a strong global model. This study's data engineering pipeline separates raw personal information data into features and labels, applies transformations and encryption for privacy, which converts features and labels into training set data and validation set data, and further applies randomizations, fences, and secure information to enable AI model training and validation.

Table 1.1 represents the major steps of the data engineering pipeline. The steps in the pipeline include analysis, industry best practices, defining datasets, analysis of raw labels and features, label and feature derivations, developing benchmark AI models for validation, selection and further development of driven AI models, model building, and other related steps. The AI pipeline is applied to train risk AI models used in insurance contracts, risk prices, claim costs, and expected insurance distributions. AI models are typically built on actuarial tables to estimate the collective values of insurance contracts, and they need to satisfy local regulations and pass external actuarial audits. In a federated AI learning environment, model validation steps aim to validate the fairness and rationality of the individual local models and their aggregated global model. We are guided by the concept that we do socially responsible data engineering, observe industry best practices, and partner with responsible governments to build responsible insurance AI models where all stakeholders, including...

2. Background and Justification

The purpose of this manuscript is to contribute to the novel emerging area of federated AI. A federated AI system consists of partner organizations that aim to collaboratively build and maintain one or more AI models in a way that respects privacy, regulation, and corporate IP. The central enabling assumption behind this special form of multi-party AI is that partner-owned data together can yield intelligence that is economically larger than the sum of the parts. In virtually every reasonable parameter setting, the most naive way to learn the model, passing partner-owned data to one center, hereinafter the monolithic pipeline, is infeasible for many important reasons. Therefore, we and many others are researching the theoretically least-permissioned way to share data concerning computing AI models.

The idea, which we call AI engineering, is to share data across an organization-specified path or paths within or on the boundary of a trusted domain without violating governance. An AI pipeline is generally a service with several steps. Each step produces temporary artifacts such as images or derived features. Thus we talk of task T which produces a list of artifacts A. An AI task T produces a list of models, a task M. These models are then put into operation so the pipeline delivers a model. The set of pipeline-delivered models is what we call model store management, or MSM. The insurance industry is the place where the principles of unsupervised learning, optimization, proactive learning from pairs of bad-bad models, and federated learning from pairs of bad-good models are consummately applied.

2.1. Significance of Federated Learning in Data Protection

The need for greater efficiency and occasionally for intellectual property protection within organizations is often cited as the most compelling reason to move towards integrated systems that leverage models on private data. It is also widely acknowledged that, when the necessary legal approvals have been obtained, data sharing very publicly leads to better models through the representation of more diverse training data. Privacy-preserving decentralized systems like federated learning, therefore, allow data-driven research to continue developing cutting-edge AI systems while respecting the legal and ethical concerns of data subjects that led to the creation of data silos in the first place. It is expected that by enabling the collaborative advancement of security-hardened AI from analytics models on private data, tailored for every individual's protection needs, our work will enable historically competing needs to be met. In the federated learning environment, the data remains private, and training data are more distributed. Some predictive learning algorithms are run on each data holder's protected data to generate analytics models on their private data. The data holder can study and consider the use of these models for their specific protection needs. If they are useful, operators request permission to use them to query new data. These queries are run on each data holder's private data, and their results are collected and used for final model updating. Throughout the process, no data is directly moved between the individual collaborators or between the individual collaborators and their common data analytics operator.

Equation 1 : Privacy-Preserving Federated Learning Update

$$w_{t+1} = w_t - \eta \sum_{i=1}^n \frac{1}{n} \nabla L_i(w_t)$$

where

- w_t = Model weights at time t ,
- η = Learning rate,
- n = Number of participating nodes,
- L_i = Loss function for node i ,
- $\nabla L_i(w_t)$ = Gradient of the loss function.

3. Federated Learning in Insurance

In this section, we review the emerging use cases of federated learning in the insurance sector. Then, we present our thoughts on industry-level design patterns and open technical questions for using federated learning in building next-generation insurance AI data engineering pipelines. In federated learning, multiple parties each own private datasets and use their computation resources for model training. Parties only share their models, rather than their data, with each other via a secure aggregator. Therefore, federated learning can handle more private data sources compared to other model training approaches, while avoiding the time and costs of data sharing and integration across multiple parties. There are several emerging use cases where federated learning can provide strong privacy guarantees for training AI models. Federated learning avoids the costs and time for refining data access and creating a central data pipeline. These observations provide early indicators for potential use cases to use federated learning in harnessing insurance data. That is, different lines of business, different entities within or external to an insurance group, governing bodies, third-party service providers, and research entities can leverage the collective AI models for actionable insights. These use cases for federated learning can also potentially drive the adoption and sustainability of more advanced AI models, both over time and across multiple use cases.

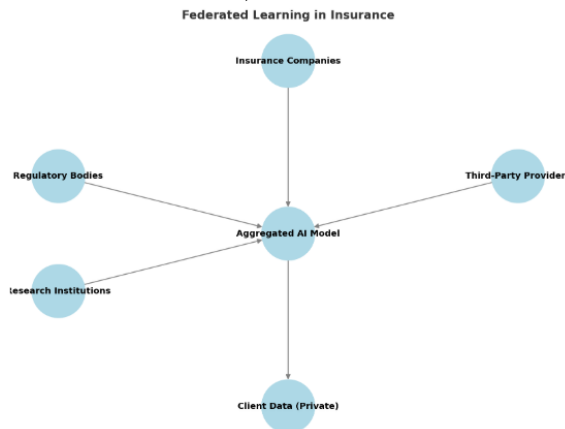


Fig 2 : Federated Learning in Insurance

3.1. Overview of Federated Learning

We review federated learning, a promising class of secure machine learning solutions. First, we frame our discussion within the context of decentralized data engineering pipelines to highlight the necessary steps that insurance actuaries and MVPs must perform in their experiments to adapt federated learning methods for their datasets. Then, we discuss a list of approaches relevant for the actuaries and MVPs to consider in their use of federated learning at these steps of their data engineering pipelines. With this content, we aim to bring the federated learning field within reach of a broader group of potential users and technical partners.

Suppose that N sites must cooperatively train an algorithm, starting from their datasets. Then, if no individual site may share its dataset or model with any other site due to privacy or security concerns, federated learning contains approaches to address these concerns. It often works as follows. First, we initialize the algorithm at each site. Then, for T rounds, each training example is consumed separately; the following example, provided with different labels, illustrates this. For each site, the current algorithm updates its weights using gradient descent, but it keeps its weights private. Only weights are exchanged between sites, not updates. After T rounds, a model is agreed upon at the end of the agreed-upon range for parameter averaging.

3.2. Applications in Insurance

Our use case is a global insurance company for which we publicly provide a package that can be used to collect the data (basic, agent, customer, risk, claim, and renewals). This data collection tool first retrieves customer count reports and filters the data based on eligibility conditions like region, age, and tariff plan. This tool also returns the minimum data needed for a given analysis and customer premiums. The main reason for providing this package is to allow researchers to test their hypotheses using real data and the reusability of our publicly available insights for other use cases in insurance. Furthermore, we provide a function that calculates basic statistics based on the outputs generated by the data collection function. Additionally, we have observed that for other service providers, these functions, like our data collection tool, could decrease the difficulties in working with such anonymized data.

In another use case, we show that insurance companies can identify and evaluate the performance of agents explicitly by comparing agent-level key performance indicators such as gross premium, incurred losses, technical prices, and social security claims with customer retention. Using state-of-the-art federated learning, social security fraud predictions can be significantly improved at the agent level with the help of machine learning models that are trained without sharing any private information between the insurer and the agent. These insights have a direct operational value and can be easily transformed into a better agent-retainer-based system, which in turn could lead to increased revenues for insurers.

4. Data Privacy Concerns

Organizations are sharing personal data electronically at an increasing rate. While opportunities for data sharing grow larger, the consequences of a breach become more severe as we increasingly digitize every aspect of our lives. As a result, privacy has become an issue of concern in the software engineering community. Fearing potential privacy breaches as well as legal liabilities, companies have been extremely conservative in releasing data. Furthermore, with the advent of more powerful machine learning models, it is an increasingly common practice for researchers to move away from curated self-produced data and rely on commercially available data that might be representative of real-world datasets for investigating data federation problems. The use of primary real-world data has raised significant data privacy concerns by industry researchers, as any privacy attack can cause direct and sometimes irreversible damage to individuals.

In the insurance and healthcare industries, customer data is the lifeblood of developing any data and AI algorithms. In addition, they are both required to comply with privacy rules. The five principles of fair information practices include: notice; choice and consent; access; security; and enforcement and alternatives. When dealing with fine-grained location data, strict privacy policies with appropriate physical, technical, and administrative safeguards need to be implemented. The out-of-compliance cost is haunting enough to scare companies straight. With the increasing risk of customer privacy when using real-world datasets, stringent privacy protection with timely evaluation of a project's impact is necessary. The difficulty is that the model losses traditionally do not take into account having a little bit of every customer's personal information.

4.1. Regulatory Frameworks

The regulatory landscape for regulating machine learning processes for insurance and other finance-sensitive applications is rapidly changing. The increasing requirements for explainability and accountability will drive machine learning outside of the realms of flexibility and complexity. However, procedures for validating models and datasets may also become stricter, which may give clear advantages to large incumbents that do not change their insurance and reinsurance governance strategies with proprietary data but will further inhibit and change the business models of startups and challenger models. These developments should be followed

closely and be part of public debates. In general, we are in favor of clear standards for validating insurance models, as they contribute to a common societal good and can help prevent scope creeping.

Strict regulation of model validation could lead to adverse effects such as unicorns with nation-states as markets, rather than the currently unlimited growth opportunities with nation-states as markets. In that scenario, models have a realistic option of adapting to vanishing margins and incomprehensible regulatory standards. Instead of cognitive standards that might never be reached and that will inadvertently distort the competitive landscape, adopting strict meta-standards of explainability that can be measured well is essential. Proper regulatory and governance frameworks do not only show desirable external effects such as sustainable, low AI risk insurance models for citizens; they also incentivize internal behavior by the data owner that results in transparent and understandable models and thus make regulatory pressure different.

4.2. Data Anonymization Techniques

Anonymization is the most fundamental operation of privacy-secure data processing. Data anonymization refers to the process of replacing identifying information of a data record with a representative value or other information so that the specific person or identity of the data will remain unrecoverable. The goal of data anonymization is to reduce the ability to detect and derive personal information, thereby protecting personal information security. Data anonymization is widely used in the sharing and publication of data and effectively prevents leaks of personal information. In many scenarios, the data anonymization technique meets the requirements of data providers.

An introduction to the process of data information preservation, including hash functions, deterministic encryption, searchable encryption, and processing, is provided. This model uses a more modular and compatible information preservation function, which can be used to construct a variety of data anonymization algorithms, such as K-anonymity, differential privacy, and L-diversity. The experimental analysis based on different datasets shows that the method can effectively maintain data characteristics while improving data security. This model generates diversification algorithm synopses through probability distribution methods, which can effectively protect sensitive data information while performing fast, efficient, and privacy-preserving private algorithms.

5. Designing Secure Data Pipelines

Data engineering pipelines are the first level of interfaces between data scientists and the data. The pipelines need to make sure that the produced AI systems have the desired legal and ethical proxies built in. At the same time, pipeline data should be as useful as possible for downstream AI node designers and should adhere to best engineering practices to assure maximum computation and communication efficiency as well as pipeline transparency and performance reproducibility. Our starting point is not to put label markers on data. Once data receives such markers, one is de facto in a government-style legal regulation mode and one does not leverage the intuition and knowledge of the end users of the data in the best possible way. There is also no guarantee that the produced models make legally safe and ethical predictions. Label marker-free techniques have appeared for natural language processing tasks.

A major issue to solve from a machine learning perspective is how much leverage to give to a prespecified data model. In the case of data augmentation in a private data engineering pipeline, for instance, the standard text data model may incorporate advanced a priori knowledge about the data model, disrupt privacy regulations, and be unable to generate a diverse set of private yet quite realistic corpora. We should generate proxy-prespecified models by generating adversarial networks that produce data-bend mini-batch level model differentiation. This is not

uncommon in the federated domain; classifications differ because divergent federated data concentrates on different identified clusters in the optimization phase. All such generated data models need to satisfy best industry practices, be subject to reproducibility checks, and be secure against gradient shape leaking.

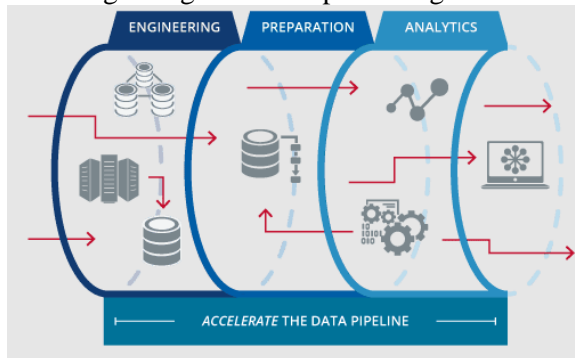


Fig 3 : Examples Of Data Pipelines

5.1. Architecture of Secure Pipelines

A federated architecture allows us to leverage horizontally distributed data. This becomes increasingly important when the inference is calculated as a function of various datasets. For example, to reduce the claim handling time, efficiently combining multiple heterogeneous datasets of a claim, including claim details, booking details, terrorism indices, politically exposed person lists, policies, and partner or medical assessor reports are used. As the source data increases, processing time increases. As both the data and processing speed are critical, we want to optimize this part as much as possible. However, traditional methodologies would only allow us to increase the number of resources that can be processed using multi-threading or cloud deployment. Contrarily, a federated architecture allows us to utilize all the resources in any network, within any organization. We can think of the architecture and applications using these datasets as a substrate, and many data models can be designed for these. With this in mind, the inference processing time is optimized.

We have a big data architecture to process the model training: but what about the model management, pre-processing, data validation, and onboarding? We first tackle this by creating a solution that uses a zero-knowledge proof to validate data in the format expected before it even enters the pipeline. Whether it is a complete dataset for model training, garbage collection, or human and Neuma-trained anti-chain models, it doesn't matter. If the correct data isn't in the right format, it gets sent back to the submitting system. We then cater for re-submission and the unique need parameters to protect the data locally. We include simple monitoring, helpfully guided by a lot of experience in previous projects. We further offer modes of zero-knowledge transfer learning through our product. This allows updates to the model to be trained in different regions, unseen by the others, and then passed to be declared, communicated, and inferred across all partners' GPUs. For most cases, lighter local updates of the model can constantly be validated if the injury is caused by the data format and use a richer, more frequent model.

5.2. Data Encryption Methods

Random Encryption (RE): In random encryption, random masking values may be applied in an even more flexible way – not only to identity attributes. Both alphanumeric and numeric formats can be encrypted using the same strategy that we define to protect the original data from exposure. Once more, the randomness of the generated cryptograms would hide the distribution, guaranteeing protection in case of compromise. Numeric fields are encrypted based on a 64-bit encoding, which results in a field with the same binary length. The presence

and length of decimal places allow users to infer a numeric field value. A parameter that multiplies the noise x_{field} is fixed and is used as a reference to the random masking limits.

Semantic Encryption (SE): Semantic encryption uses the values that identify the entity within a custom dictionary, aiming to protect the data from exposure. By applying the definition of this variation, standard attributes like identification fields can be encrypted, offering more confidentiality through the same approach. The semantic encryption method only works with fields of non-alphanumeric nature. The brute exposure of a custom dictionary is suggested because it does not present any type of sensitivity. The primary encrypting parameter adopts the techniques from random encryption, applied with a reference value that characterizes data dispersion for the entire norm. According to the minimum, maximum, and effective length of a specific field, we could classify four possible architectures that best fit the range limits. The main advantage of the architecture is the protection of the data, which is extended by the shared use of repositories. An additional possibility arises when user credentials or sensates need to authenticate stored information.

Equation 2 : Secure Multiparty Computation (MPC) for Data Aggregation

$$S = \sum_{i=1}^n Enc(d_i)$$

where
 S = Securely aggregated sum,
 $Enc(d_i)$ = Encrypted insurance data from node i ,
 n = Number of federated participants.

5.3. Access Control Mechanisms

As shown in the case study, implementing an access control system at the feature store level is advantageous. The feature store level controls the features that are shared across various models and numerous different consumer groups. Being a logical place allows for enhanced decision-making at feature consumption. Due to feature store compatibility with different models and the information on consumers built, access control can be flexibly shaped. Logical tiering can also control the number of layers within a feature store because sometimes more levels are detrimental. Still, the access control operation in the encrypted feature store always has the role of an authorizer. The number of copies of a particular feature is two—one encrypted in the feature store and one in clear text. The TPM also contains statistical data and incidents. The TPM can broadcast and authorize a request, but feature store logical tiering can help in relocating the feature store from the device to a centralized server. The authorizer can check for previously assigned authorization grants, which would mean that the feature has been calculated and stored in the feature store. Former consumers create behavior, and requests can be tracked in addition to previously assigned authorization. An attack might be exposed easily by a lack of compatible consumer ID correlation. If absent, it would mean that the information has to be contained in an ongoing request filed against that consumer, and has been provided to other consumers in the feature store at a prior time.

6. Performance Considerations

Insurance AI at scale is feasible because many characteristics of the structured data make the problem easier. For instance, fitting a well-specified generalized linear model to score a program of millions of risks takes a matter of minutes. The field could quickly move toward more sophisticated machine learning for scoring once technologies, such as infrastructure for federated learning, are developed and the resulting models are fielded for the varied stakeholders in the commercial insurance value chain. There are several disparate pieces involved in creating performance. The challenges with insurance data are many, but the good news is the low-to-modest amount of compute resources required for many of the stages.

Efficient model fitting of some good models with sound features can provide some powerful model-based action. There are a few particular technologies that do significantly complicate the insurance data challenge: 1) securing data and modeling when fitting a model, 2) AI/ML pipelines for ensuring model performance, and 3) local data engineering development scaling for an AI industry with a long left tail of small insurers. Finally, it is helpful to frame out the orders of magnitude of the stages people must engineer so that they know where to focus and invest, whether that investment is in-house team time and expertise or outsourcing to third-party vendors. Implementing some advanced analytics, even for very complex models like boosting, may only cost a few dollars per model at most traditional brokerages.

6.1. Latency and Throughput

Retraining models takes time and resources, which can be costly considerations. To minimize the time, the input data must consider many factors, among them: Dealing with historical data. If training data are historical, only a design with low time/latency, low batch size, or segment-based architecture will allow for sufficiently fast processing. Taking into account business cycle characteristics. For instance, if a business activity or a new regulation requires the modeling of vast instrumental variables or other economic behaviors, the business cycle speed constraint is likely to make a segment-based architecture required and a low batch size recommended. System structure constraints. The current state of servers, distributed computations, and parallelism characteristics can condition big/little batch size, segment-based process, speed, and latency. To maximize throughput, system structure constraints must be met, and these constraints will be determined by the system being built or used. Among the most common tasks that we find in the state-of-the-art data and AI in the insurance industry, usually computer vision, audio, text modeling, and regression or time series analysis, which we will now briefly review. Our goal is to determine the input data pipeline characteristics for each task in terms of speed and throughput to be able to compare them with the scaling characteristics of the segments of the problems to be solved. Then we will be able to design a general framework to compare them with the three main requirements of the SLAs.

6.2. Resource Allocation Strategies

The two main knobs on resource allocation concern allocating enough computing and memory to the next step in the pipeline, and not allocating so much that you suffer significant costs due to the need to wait at the next stage. How many resources are available for us to allocate, and how to better allocate towards these two goals are special challenges. In most classic deployments, the compute and memory trade-offs at each stage of the funnel are mostly pegged to the stage that presents the most difficult bottleneck: fortunately, this is not the case in the deep learning-based insurance AI pipeline. Instead, it is often some of the highest levels of feature engineering and data engineering that represent the computational resourcing pain point.

The resource allocation challenge is different at each stage in a complex data- and learning-centric pipeline such as the funnel. Each of our funnel's eleven main stages is at various times compute-starved and all set waiting to consume a pretty grandiose set of parallel hyper-threaded cores and significant memory; each of our organization's two engineering pools' two dozen or more human engineers (some of whom act quickly to prioritize, others less so); and each of our destinations' sliding windows of warehouse storage and memory arrays tailored to the confident generation of high-quality 'feature pipelines'. Making access to adequate computing and memory resources as simple as possible, but no simpler, is a special research and development focus for the insurance IoT.

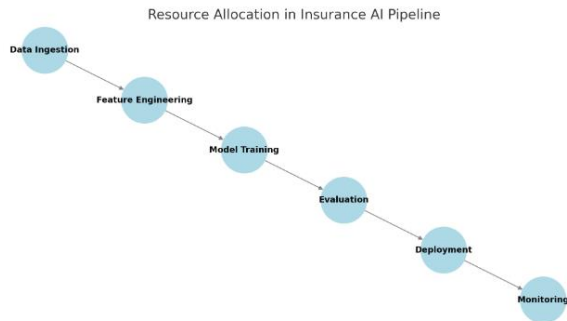


Fig 4 : Resource Allocation Strategies

7. Intelligent Data Processing Techniques

In addition to ensuring the privacy protection of the data, it is important that modern AI workloads are able to operate on data efficiently. To achieve that, it is necessary to develop advanced techniques to ensure that the AI work is done faster but also securely. Here, we discuss various techniques that can be used and which are used in our pilot experiments.

Secure data preprocessing. Depending on the type of the problem, different processing including data filters, transformation, reduction, etc. are performed. This is especially true for NLP tasks. Attention has to be paid to data preprocessing to ensure that the resulting data can ensure users' privacy. It is necessary to ensure that in a transformed representation of the data, critical characteristics to reveal sensitive information are not regained.

Training process acceleration. To speed up the training process while also balancing the workload, federated learning is performed. However, with the typical implementation of the algorithm, the training process can end up being slower compared to training on a non-federated dataset. Several emerging research works are looking at speeding up the training process, which often involves modifications to the machine learning model, such as sparsification, compression, and quantization. Also, in federated learning, after performing several iterations, a dataset that contains all updates of all users can be constructed to allow further optimizations to speed up the training.

7.1. Machine Learning Algorithms

Numerous learning algorithms seek to derive decision boundaries or estimates in low- and high-dimensional input spaces. While decision boundaries distinguish distinct classes for classification tasks, the output of regression problems can vary in dimension and range. The following sections address the most common modeling techniques implemented in insurance contracts and next-best-offer pipelines to prioritize and optimize workflows across several disciplines, including experimentation and risk assessment.

7.1.1. Decision Trees. Decision trees are nonparametric, supervised learning models with an extensive application background in classification and regression tasks. Each leaf node in a decision tree is associated with a class label or continuous value, while the root represents the binary decision for a specific data partition. Decision trees do not implicitly assume data distribution, can accommodate both classification and regression and use a utility function for splitting. Since decision trees can be not only stochastic but also extremely adaptive learners, data scientists use random forests to make them available for parallel training.

7.2. Real-time Data Analytics

The insurance operations generate large-scale data streams, composed of many varieties such as vision, text, structured documents like invoices, unstructured incoming or sent email content, and all varieties of customer-related data stored in databases, like web form responses. Most of

this data is structured or semi-structured. Large-scale data collection and understanding in real time can support not only insurance product service but also risk control. However, the machine learning development pipeline and production pipelines have different pipeline-level QoS requirements. The development pipeline should be executed many times and needs results in a reasonable time. The production pipelines are running almost continuously, with potentially huge volume application batch results produced for aggregate reporting and business intelligence.

We leverage data engineering technology to build a fast and scalable image ingestion pipeline and online data algorithm platform. The image preprocessing is as follows: the images of 3MB resolutions are resized to 30KB, 640x640. Then, the image data is transformed into an array format, converted to float32 type by the data preprocessing, and sent to the data retrieval pipeline. For retrieving the nearest k clusters, we tag the nearest nonzero row of weight matrices using the associated k cluster types. The predictions for each cluster type among the k predicted using the activation predict operation from the i-th cluster for vector addition, and the hash lookup is also performed on this weight table. The label of the cluster is selected according to activation predictors in the monitoring stage. The processing speed will not exceed 30ms for standard 640x640 images. The ensemble model uses post-processing, and the speed can meet the business requirements. The online data support is implemented in the online algorithm. These online models have demonstrated near real-time capability for image analysis online.

8. Case Studies

In this section, we illustrate our research in action with two case studies, demonstrating secure data engineering pipelines for federated insurance AI. It is to be noted that SEC-DEP is a research artifact with specific decision-making mechanisms. In a specific industry and business scenario, different technologies, including Statistical Data Validation methods and other secure or confidential computing technologies, can be adopted. We hope that the prototype system structures can provide a reference for the actual deployment of the project. Our first case study with a top-tier insurance corporation involves a federated AI architecture to recognize cracks, scratches, stains, ruptured edge tiles, and discoloration defects in images taken from the site's machinery. On the customer side, we realize specific tasks such as counting the number of defects in a photo and marking the position of the defects. Some tasks that are not sensitive can be handled directly on the customer's side, and fewer accurate results are required for these tasks; then the results are directly sent to the client for viewing. Our second collaborative research project with a top-tier insurance corporation focuses on personal damage detection via accident photo recognition. The industry has strict data privacy protection requirements corresponding to the commercial and risk processing of client's identity and damage.

8.1. Case Study 1: Health Insurance

In this section, we describe how to solve an important problem in health insurance using privacy-preserving machine learning. Note that this use case can easily be adapted to other types of insurance, such as disability, auto, and any other policy for employees and their families. An insurance premium is the price a customer pays to the insurance company for coverage. The most common type of premium pricing is set according to the claims history or current health status. Large health claims can have a substantial outcome on an insurance company's revenue and profit. The insurer charges the customer based on prior claim history in that employee pool. A wellness mileage reward app has a low participation rate, which causes the data to be imbalanced, resulting in the potential for unfairness when setting the premium price.

A new measure was implemented to encourage business owners and employees to adopt a healthier lifestyle. Employees of companies participating in a consortium of workplaces were

allowed to join a wellness council-funded Walking Program. The employees used an app supplied by the health and life insurer. The insurer used the wellness activity data to set wellness premium incentives for the companies involved in the program. Data was kept separate during the pilot project and only the wellness offer was given. The insurance company had to utilize private set operations to match the individual-provided reports and receipt numbers while employing differential privacy, a type of cryptographic privacy, so they could not see which companies or individuals were involved during the pilot phase. Small tests adjudicated by the insurance company demonstrated that reduced privacy had minimal effect on the value of the predictions.

8.2. Case Study 2: Auto Insurance

To put the approach into context, we evaluate the platform with a case study in the insurance domain. The goal of the case study is to accurately predict the insurance costs for customers based on customer demographics. We apply two popular learning models: a feed-forward neural network and a gradient-boosting decision tree. Since the data are very sensitive and insurance companies put a high value on the accuracy of estimated models, our task is to deliver accurate models as fast as possible. The privacy protection requirements are also critical for this case study since the usage of insurance data is highly regulated. This section presents a case study in the auto insurance domain. We show the performance of the infrastructure and the deep learning models. Since the time cost or monetary cost of federated learning is generally neglected in current works, we specifically explore the time consumption of building deep learning models with different numbers of clients in the federation as well as the performance of the collaborative learning model. The results are promising for industry applications that require the cooperation of different data owners in building deep learning models.

8.3. Case Study 3: Home Insurance

We are currently applying to the consortium for production-scale federated learning with a consortium of seven of the largest home insurers in Norway. In this pilot, the insurance companies use a secure federated learning platform to train models that predict the expected value in the case of home insurance for up to 1.5 million potential future customers. The models use property data from the property register that includes information about all homes in Norway, such as address, square meters, year of construction, and construction materials for both external and internal constructions, as well as estate and outdoor areas. Additionally, the model makes predictions related to critical in-house parameters tailored to the specific needs of the respective insurance company partners, allowing the insurance companies to calculate the risks associated with each home in the country. These models are trained on secure private data that reside with the data owner and do not need to be moved. Access control to the model is implemented at the tensor level, and the data are accessed only during training and are not visible to the data scientists who train the models.

The use of secure computation federated learning is a significant game-changer in the provision of personalized risk estimation to a broad market. The new technical approach provides high-value insurance-based predictions tailored to the individual needs of both existing and potential future policyholders. The predictions are dynamically updated to reflect current events and are based on a comprehensive set of house-specific and neighborhood-specific parameter risk proxies. Critically, the use of secure federated learning prevents insurance companies from becoming data factories, reducing privacy risks and the risk of distinguishing or targeting specific homes. We thus demonstrate how privacy, speed, and intelligence can be balanced while targeting efficiency and, at the same time, improving the quality, transparency, and fairness of the insurance industry.



Fig 5 : AI in Insurance

9. Challenges and Limitations

We conclude this paper by discussing important challenges in realizing our vision of secure data engineering pipelines for federated insurance AI. We discuss scientific challenges that need to be addressed as well as practical challenges in the deployment and operation of such secure pipelines. This section is the first to formally state the limitations of our solution and discuss realistic next steps.

There are important scientific challenges ahead. Our solutions for secure data and model management, perfect data and feature balancing, and distributed matrix completion are rather preliminary. There is a significant opportunity to align these solutions with insurance business problems to focus our efforts, develop metrics to quantify improvements and feedback loops for continuous learning, and research other missing ingredients, such as private feature computation, long-tail feature augmentation, multi-modal feature generation and validation, etc. In addition, while we have not yet seen any catastrophic large-scale attacks against federated machine learning and our pipeline, we also do not yet have a precise security guarantee. Sometimes, some of our features are only known in the case of a claim, and a counterfactual scenario can be generated given that certain features are different. To protect privacy, we may choose to hide or transform the feature such that it becomes meaningless. This new feature no longer carries semantics and can severely affect model accuracy, which can have significant implications for making favorable pricing decisions for our customers.

9.1. Scalability Issues

The scale of the federated data engineering problem in insurance is challenging. Large insurers frequently have tens of thousands of agents in hundreds of different roles. Every insurance company has many teams of data analysts, data scientists, software developers, GIS specialists, actuaries, underwriters, and model risk officers, as well as business support functions. Each of the thousands of models in use may have different data, derived data, and performance expectations. In addition, several stakeholders within the company are looking for aggregate indicators, reports, and visualizations. Ultimately responsible for building and deploying the models, the internal audit teams determine the breadth of information required to be considered and the techniques and technologies to be applied. Aggregating, anonymizing, extracting the length of the required inputs, and labeling thousands of data sets and models quickly, reliably, and continuously are required.

The overhaul of feature factory management principles is not an isolated technical concern. When confronted with a manageable number of model training instances, practitioners could tolerate a certain amount of manual feature engineering to reach benchmark signals that could be iteratively positioned and evaluated during long development cycles. However, at the pace of practical change in an operational environment in production with more than a small set of models, regression tasks with curated and readily available input sets constantly require maintenance after finding an environment located in a registry or repository of new features. The massive increase in training and re-tensioning leads to an exponentially increasing cost to balance categories with manual operations that would be impossible due to the relatively few

decisions. Therefore, intelligence concerning the first decisions, when production models are created, remains a significant barrier to the development of automated feature engineering. Because trees and reserves are built for many years and portfolios may have a longer life, observatory and transparency techniques in capability decisions must provide a longer perspective of the horizon.

9.2. Interoperability Concerns

Federated Insurance AI presents unique challenges in information sharing, ranging from non-federated models built on external data to distributed learning concerning artifacts that must be reconciled across geographies and with central data centers. These models need to increase data portability, protect model innovation, and facilitate sharing. Run atop industry-standard data engineering stacks, these techniques can help act as mapping translators. The architecture builds a secure enterprise data engineering platform to act as a trusted information intermediary. At the boundary between the regulated insurance business and the unregulated portions of the enterprise, these data engineering pipelines can help increase translation between the two, effectively providing a secure stage area that enables as close to zero trust information sharing as possible while still serving the needs of the enterprise as a whole.

Bridging the gap between traditional enterprise models of data warehouses, data marts, and data engineering with a truly federated learning solution will require new protocols and standards for structured and semi-structured data exchanges. Models and model data portability will help; however, they are part of the solution needed to support true data engineering collaboration. The enterprise platform represents the ideal unifying platform to help coordinate this data engineering 'pipework,' and whether existing unregulated enterprise data engineering tools can be used in a pseudo-federated manner or it becomes more efficient to stitch new real federated modules together, the industry needs open standards using universal entity definitions to facilitate the rapid innovation happening in the federated vertical.

10. Future Directions

Our SDS ecosystem is a proof of concept that provides flexibility and scalability. However, it has some limitations. In future work, we aim to improve our methodology and SDS ecosystem in terms of model optimization and privacy. As advanced insured analytics requires further optimization, our next priority is research on optimal model architecture and lighter-weight algorithms. We would like to formalize key metrics for model optimization and provide insights into optimal ML platform architecture with hyperparameter search. Our research area includes user-guided collaborative ML optimization, which seamlessly integrates collected industry knowledge into the pipeline. Conventional privacy-preserving distributed learning assumes honest participants. Insurance is a community and participants may not be honest, but the stakes are far too high to lose a customer.

Future work will address secure privacy-preserving technologies against semi-honest and Byzantine attacks by implementing game theoretic and automated data repairing techniques. Centralized ML favors architectural choices that improve the efficiency of gradient computation, but privacy-related considerations introduce choices that are completely different and may alter software and hardware design choices. Data obtained from AI and real-time flows, as close to real-time as possible, exist in only a few companies due to this significant barrier. These companies are among the global giants that are already harnessing the power of advanced analysis in insurance across various business lines. Consumer-led real-time data exchange needs to mature, if we are to ensure our products stay relevant and respond to emerging customer needs.

10.1. Advancements in AI Technology

In the early days of AI, the most prominent languages and environments for the development, exchange, and replication of software were LISP and Prolog. More recently, with the development of the cloud as a central platform for applications, the current context lowered the entry barriers, allowing the interaction of large data sets with the necessary high computing power. This allowed the design of a decentralized infrastructure, giving rise to the use of AI in various types of applications, from video games and self-driving cars to personal assistants and chatbots. In the approaches using GPU clusters, FPGA, and TPUs, we found deep learning development tools, libraries, and software resulting from AI modeling and infrastructure developments.

Concerning language, other modern languages promoted developments in scripting or dynamic programming environments, and Python has a solid foundation with machine learning libraries. On the other hand, languages such as Scala developed environments that favor high-performance computing using Spark with libraries for machine learning. The software that runs on hardware with high computing capacity and focuses on specific machine learning layers also has specific languages for developing, training, and using Convolutional Neural Networks for deep learning across a range of tasks and devices. The Keras library provides developers with an easy-to-understand format for rapid experimentation for all intents and purposes and has become one of the most used neural network software libraries.

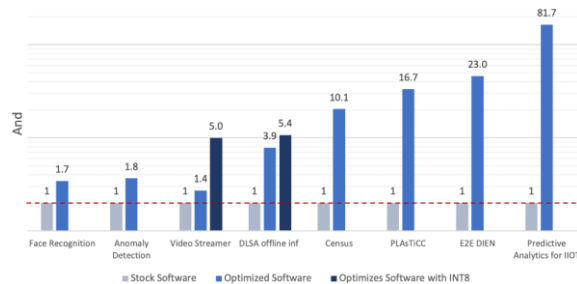


Fig 6 : AI Pipeline Optimization

10.2. Emerging Privacy Solutions

We reviewed the most interesting emerging privacy solutions for insurance companies: advanced cryptographic solutions supporting privacy via federated machine learning, distributed machine learning, and homomorphic encryption. Distributed machine learning is a promising way out of many of the training encryption problems and performance limitations inherent in allowing many other participants to see large amounts of encrypted data. However, despite being widely covered by the media, it is not mature and may potentially introduce security, and not privacy, concerns. In reality, there are fewer than 130 cybersecurity professionals worldwide working on any type of secure multi-party machine learning. Additionally, much of the development effort in multi-party machine learning research has not focused on ensuring a balance with performance for real-world problems. We cover the most promising emerging privacy solutions that could be of interest in the further research of federated data engineering in the context of insurance, particularly the following: K-federated learning, a generalization of federated learning to $K > s$ participants where, at each iteration, only s participants communicate with each other; ownerless machine learning, decentralized techniques shifting ownership of data and models so none of the potential adversaries can see them simultaneously.

Equation 3 : Latency-Optimized Secure Query Processing

where

T_q = Total query execution time,

T_d = Data retrieval time,

T_c = Computation time for encryption/decryption,

T_t = Transmission time in federated networks.

$$T_q = T_d + T_c + T_t$$

11. Conclusion

In conclusion, federated insurance AI leverages data generated across insurance customers in a privacy-preserving manner and allows machine learning systems to identify patterns and make decisions informed by these patterns. We found that a need for complete privacy in federated models leads to data sparsity, which raises some particular challenges for insurance companies. After all, we rely on the very same patterns generated by customers in the past to inform our predictions of the future. We presented meaningful contributions toward that goal in the form of matched data, generated by the open and privacy-preserving data exchange between data engineering and data science services data pipeline architecture.

Increased data sparsity and the federated learning paradigm together place additional constraints on the models allowed to be trained for intelligence at a given level of predictive accuracy. We therefore presented a novel approach leveraging a light outer model that can be trained without data but only the outcomes, and a deep inner model. The former uses publicly available externalized metadata, and the latter is trained on the more informative customer-generated models. Furthermore, data sources from wellness programs that allow us to resell data to data partners at a lower price than conventional data sources are characterized. With all this, we managed to adapt to challenges introduced by customer data sparsity and still deliver the necessary intelligence for insurance AI.

12. References

- [1] Ravi Kumar Vankayalapati, Venkata Krishna Azith Teja Ganti. (2022). AI-Driven Decision Support Systems: The Role Of High-Speed Storage And Cloud Integration In Business Insights. *Migration Letters*, 19(S8), 1871–1886. Retrieved from <https://migrationletters.com/index.php/ml/article/view/11596>
- [2] Avinash Pamisetty. (2022). Enhancing Cloudnative Applications WITH Ai AND MI: A Multicloud Strategy FOR Secure AND Scalable Business Operations. *Migration Letters*, 19(6), 1268–1284. Retrieved from <https://migrationletters.com/index.php/ml/article/view/11696>
- [3] Balaji Adusupalli. (2022). The Impact of Regulatory Technology (RegTech) on Corporate Compliance: A Study on Automation, AI, and Blockchain in Financial Reporting. *Mathematical Statistician and Engineering Applications*, 71(4), 16696–16710. Retrieved from <https://philstat.org/index.php/MSEA/article/view/2960>
- [4] Chakilam, C. (2022). Integrating Generative AI Models And Machine Learning Algorithms For Optimizing Clinical Trial Matching And Accessibility In Precision Medicine. *Migration Letters*, 19, 1918–1933.
- [5] Maguluri, K. K., Pandugula, C., Kalisetty, S., & Malleshm, G. (2022). Advancing Pain Medicine with AI and Neural Networks: Predictive Analytics and Personalized Treatment Plans for Chronic and Acute Pain Managements. *Journal of Artificial Intelligence and Big Data*, 2(1), 112-126.
- [6] Koppolu, H. K. R. 2022. Advancing Customer Experience Personalization with AI-Driven Data Engineering: Leveraging Deep Learning for Real-Time Customer Interaction. *Kurdish Studies*. Green Publication. <https://doi.org/10.53555/ks.v10i2.3736>.
- [7] Sriram, H. K. (2022). AI Neural Networks In Credit Risk Assessment: Redefining Consumer Credit Monitoring And Fraud Protection Through Generative AI Techniques. *Migration Letters*, 19(6), 1017-1032.
- [8] Chava, K. (2022). Redefining Pharmaceutical Distribution With AI-Infused Neural Networks: Generative AI Applications In Predictive Compliance And Operational Efficiency. *Migration Letters*, 19, 1905-1917.
- [9] Puli, V. O. R., & Maguluri, K. K. (2022). Deep Learning Applications In Materials Management For Pharmaceutical Supply Chains. *Migration Letters*, 19(6), 1144-1158.
- [10] Challa, K. (2022). Generative AI-Powered Solutions for Sustainable Financial Ecosystems: A Neural Network Approach to Driving Social and Environmental Impact. *Mathematical Statistician and Engineering*.

- [11] Sondinti, L. R. K., & Yasmeen, Z. (2022). Analyzing Behavioral Trends in Credit Card Fraud Patterns: Leveraging Federated Learning and Privacy-Preserving Artificial Intelligence Frameworks.
- [12] Malempati, M. (2022). Machine Learning and Generative Neural Networks in Adaptive Risk Management: Pioneering Secure Financial Frameworks. *Kurdish Studies*. Green Publication. <https://doi.org/10.53555/ks.v10i2>, 3718.
- [13] Pallav Kumar Kaulwar. (2022). The Role of Digital Transformation in Financial Audit and Assurance: Leveraging AI and Blockchain for Enhanced Transparency and Accuracy. *Mathematical Statistician and Engineering Applications*, 71(4), 16679–16695. Retrieved from <https://philstat.org/index.php/MSEA/article/view/2959>
- [14] Nuka, S. T. (2022). The Role of AI Driven Clinical Research in Medical Device Development: A Data Driven Approach to Regulatory Compliance and Quality Assurance. *Global Journal of Medical Case Reports*, 2(1), 1275.
- [15] Kannan, S. (2022). The Role Of AI And Machine Learning In Financial Services: A Neural Networkbased Framework For Predictive Analytics And Customercentric Innovations. *Migration Letters*, 19(6), 985-1000.
- [16] Maguluri, K. K., Pandugula, C., Kalisetty, S., & Mallesham, G. (2022). Advancing Pain Medicine with AI and Neural Networks: Predictive Analytics and Personalized Treatment Plans for Chronic and Acute Pain Managements. *Journal of Artificial Intelligence and Big Data*, 2(1), 112-126.
- [17] Vankayalapati, R. K. (2022). Harnessing Quantum Cloud Computing: Impacts on Cryptography, AI, and Pharmaceutical Innovation. *AI, and Pharmaceutical Innovation* (June 15, 2022).
- [18] Subhash Polineni, T. N., Pandugula, C., & Azith Teja Ganti, V. K. (2022). AI-Driven Automation in Monitoring Post-Operative Complications Across Health Systems. *Global Journal of Medical Case Reports*, 2(1), 1225.
- [19] Komaragiri, V. B. (2022). AI-Driven Maintenance Algorithms For Intelligent Network Systems: Leveraging Neural Networks To Predict And Optimize Performance In Dynamic Environments. *Migration Letters*, 19, 1949-1964.
- [20] Ganesan, P. (2020). PUBLIC CLOUD IN MULTI-CLOUD STRATEGIES INTEGRATION AND MANAGEMENT.
- [21] Annapareddy, V. N. (2022). Innovative AIdriven Strategies For Seamless Integration Of Electric Vehicle Charging With Residential Solar Systems. *Migration Letters*, 19(6), 1221-1236.
- [22] Vankayalapati, R. K. (2022). Composable Infrastructure: Towards Dynamic Resource Allocation in Multi-Cloud Environments. Available at SSRN 5121215.
- [23] Challa, S. R. (2022). Optimizing Retirement Planning Strategies: A Comparative Analysis of Traditional, Roth, and Rollover IRAs in LongTerm Wealth Management. *Universal Journal of Finance and Economics*, 2(1), 1276.
- [24] Chakilam, C. (2022). Generative AI-Driven Frameworks for Streamlining Patient Education and Treatment Logistics in Complex Healthcare Ecosystems. *Kurdish Studies*. Green Publication. <https://doi.org/10.53555/ks.v10i2>, 3719.
- [25] Subhash Polineni, T. N., Pandugula, C., & Azith Teja Ganti, V. K. (2022). AI-Driven Automation in Monitoring Post-Operative Complications Across Health Systems. *Global Journal of Medical Case Reports*, 2(1), 1225.
- [26] R. Daruvuri, "Harnessing vector databases: A comprehensive analysis of their role across industries," *International Journal of Science and Research Archive*, vol. 7, no. 2, pp. 703–705, Dec. 2022, doi: 10.30574/ijrsra.2022.7.2.0334.
- [27] Siramgari, D. (2022). Unlocking Access Language AI as a Catalyst for Digital Inclusion in India. Zenodo. <https://doi.org/10.5281/ZENODO.14279822>
- [28] Kalisetty, S., Vankayalapati, R. K., Reddy, L., Sondinti, K., & Valiki, S. (2022). AI-Native Cloud Platforms: Redefining Scalability and Flexibility in Artificial Intelligence Workflows. *Linguistic and Philosophical Investigations*, 21(1), 1-15.
- [29] Malempati, M. (2022). AI Neural Network Architectures For Personalized Payment Systems: Exploring Machine Learning's Role In Real-Time Consumer Insights. *Migration Letters*, 19(S8), 1934-1948.
- [30] Kalisetty, S., & Ganti, V. K. A. T. (2019). Transforming the Retail Landscape: Srinivas's Vision for Integrating Advanced Technologies in Supply Chain Efficiency and Customer Experience. *Online Journal of Materials Science*, 1, 1254.
- [30] Siramgari, D., & Korada, L. (2019). Privacy and Anonymity. Zenodo. <https://doi.org/10.5281/ZENODO.14567952>
- [31] Ganesan, P. (2020). DevOps Automation for Cloud Native Distributed Applications. *Journal of Scientific and Engineering Research*, 7(2), 342-347.
- [32] Komaragiri, V. B., & Edward, A. (2022). AI-Driven Vulnerability Management and Automated Threat Mitigation. *International Journal of Scientific Research and Management (IJSRM)*, 10(10), 981-998.
- [33] Ganti, V. K. A. T., & Valiki, S. (2022). Leveraging Neural Networks for Real-Time Blood Analysis in Critical Care Units. In *KURDISH*. Green Publication. <https://doi.org/10.53555/ks.v10i2.3642>

- [34] R. Daruvuri, "An improved AI framework for automating data analysis," *World Journal of Advanced Research and Reviews*, vol. 13, no. 1, pp. 863–866, Jan. 2022, doi: 10.30574/wjarr.2022.13.1.0749.
- [35] Ganesan, P. (2020). *Balancing Ethics in AI: Overcoming Bias, Enhancing Transparency, and Ensuring Accountability*. *North American Journal of Engineering Research*, 1(1).
- [36] Vankayalapati, R. K. (2022). *AI Clusters and Elastic Capacity Management: Designing Systems for Diverse Computational Demands*. Available at SSRN 5115889.
- [37] Siramgari, D. R. (2022). *Evolving Data Protection Techniques in Cloud Computing: Past, Present, and Future*. Zenodo. <https://doi.org/10.5281/ZENODO.14129065>
- [37] Vankayalapati, R. K., & Pandugula, C. (2022). *AI-Powered Self-Healing Cloud Infrastructures: A Paradigm For Autonomous Fault Recovery*. *Migration Letters*, 19(6), 1173-1187.
- [38] Maguluri, K. K., & Ganti, V. K. A. T. (2019). *Predictive Analytics in Biologics: Improving Production Outcomes Using Big Data*.
- [39] Sondinti, K., & Reddy, L. (2019). *Data-Driven Innovation in Finance: Crafting Intelligent Solutions for Customer-Centric Service Delivery and Competitive Advantage*. Available at SSRN 5111781.
- [40] Ganesan, P. (2021). *Cloud Migration Techniques for Enhancing Critical Public Services: Mobile Cloud-Based Big Healthcare Data Processing in Smart Cities*. *Journal of Scientific and Engineering Research*, 8(8), 236-244.
- [41] Polineni, T. N. S., & Ganti, V. K. A. T. (2019). *Revolutionizing Patient Care and Digital Infrastructure: Integrating Cloud Computing and Advanced Data Engineering for Industry Innovation*. *World*, 1, 1252.
- [42] Ganti, V. K. A. T. (2019). *Data Engineering Frameworks for Optimizing Community Health Surveillance Systems*. *Global Journal of Medical Case Reports*, 1, 1255.
- [43] Pandugula, C., & Yasmeen, Z. (2019). *A Comprehensive Study of Proactive Cybersecurity Models in Cloud-Driven Retail Technology Architectures*. *Universal Journal of Computer Sciences and Communications*, 1(1), 1253. Retrieved from <https://www.scipublications.com/journal/index.php/ujcsc/article/view/1253>
- [44] Burugulla, J. K. R. (2022). *The Role of Cloud Computing in Revolutionizing Business Banking Services: A Case Study on American Express's Digital Financial Ecosystem*. *Kurdish Studies*. Green Publication. <https://doi.org/10.53555/ks.v10i2.3720>.
- [45] Satyaveda Somepalli. (2022). *Beyond the Pill: How Customizable SaaS is Transforming Pharma*. *The Pharmaceutical and Chemical Journal*. <https://doi.org/10.5281/ZENODO.14785060>
- [46] Vankayalapati, R. K. (2020). *AI-Driven Decision Support Systems: The Role Of High-Speed Storage And Cloud Integration In Business Insights*. Available at SSRN 5103815.
- [47] Somepalli, S. (2021). *Dynamic Pricing and its Impact on the Utility Industry: Adoption and Benefits*. Zenodo. <https://doi.org/10.5281/ZENODO.14933981>
- [48] Yasmeen, Z. (2019). *The Role of Neural Networks in Advancing Wearable Healthcare Technology Analytics*.
- [49] Siramgari, D. (2022). *Enhancing Telecom Customer Experience Through AI Driven Personalization - A Comprehensive Framework*. Zenodo. <https://doi.org/10.5281/ZENODO.14533387>
- [50] Maguluri, K. K., Yasmeen, Z., & Nampalli, R. C. R. (2022). *Big Data Solutions For Mapping Genetic Markers Associated With Lifestyle Diseases*. *Migration Letters*, 19(6), 1188-1204.
- [51] Polineni, T. N. S., Maguluri, K. K., Yasmeen, Z., & Edward, A. (2022). *AI-Driven Insights Into End-Of-Life Decision-Making: Ethical, Legal, And Clinical Perspectives On Leveraging Machine Learning To Improve Patient Autonomy And Palliative Care Outcomes*. *Migration Letters*, 19(6), 1159-1172.
- [52] Ravi Kumar Vankayalapati, Venkata Krishna Azith Teja Ganti. (2022). *AI-Driven Decision Support Systems: The Role Of High-Speed Storage And Cloud Integration In Business Insights*. *Migration Letters*, 19(S8), 1871–1886. Retrieved from <https://migrationletters.com/index.php/ml/article/view/11596>