General Commerce Intelligence: Glocally Federated NLP-Based Engine for Privacy-Preserving and Sustainable Personalized Services of Multi-Merchants

Kyoung Jun Lee[1,2], Baek Jeong[1], Suhyeon Kim[1], Dam Kim[2], Dongju Park[2]
[1]Department of Big Data Analytics, Kyung Hee University, Seoul, Korea
[2]User Centric Artificial Intelligence Research Institute, Harex InfoTech, Seoul, Korea
{klee, ylbaek, kimsuuuu_99} @khu.ac.kr, {yedam.kim, dongju.park} @ubplatform.net

Abstract

One of the most crucial capabilities in the commercial sector is a personalized prediction of a customer's next purchase. We present a novel method of creating a commerce intelligence engine that caters to multiple merchants intended for the UB Platform, managed by e-payment company Harex InfoTech. To cultivate this intelligence, we utilized payment receipt data and created a Natural Language Processing (NLP)-based commerce model using a Transformer to accommodate multinational and merchant trade. Our model, called General Commerce Intelligence (GCI), provides a range of services for merchants, including product recommendations, product brainstorming, product bundling, event promotions, collaborative marketing, target marketing, and demand forecasting etc. To bolster user privacy and foster sustainable business collaboration, especially among micro-, small-, and medium-sized enterprises (MSMEs), the GCI model was trained through federated learning, especially with glocalization. This study delves into the structure, development, and assessment of GCI, showcasing its transformative capacity to implement User Centric AI and reshape the global commerce landscape to benefit MSMEs.

Introduction

AI offers promising prospects for enhancing business performance and enabling personalized customer service, contributing to increased sales. However, Artificial Intelligence (AI) implementation requires substantial data resources. While businesses have the option to engage with AI services offered by large technology platforms, most of these platforms require businesses to provide access to their data. This requirement has led to a growing discrepancy between businesses and Big Tech platforms. Therefore, the promotion of AI accessibility for smaller businesses or emerging markets, often referred to as the democratization of AI, is emerging as a pivotal direction in AI research and development (Björkegren 2023; Perrigo 2023; Seger et al. 2023).

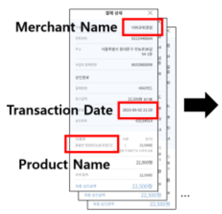
Because platforms leverage MSME's data to enhance AI, MSMEs may experience reduced autonomy in their data and operations. MSMEs lack the resources to independently develop AI systems and thus become heavily reliant on large platforms. Given these challenges, this study explores a new approach to developing an engine for multi-merchant platforms called General Commerce Intelligence (GCI).

GCI was conceived to create a GPT-like model tailored for commerce. GPT utilizes publicly available data (Brown et al. 2020), GCI relies on customer purchase data, which is not publicly accessible. Customers and companies involved in the transactions are reluctant to disclose their data. However, robust AI cannot be developed using only individual customer or company data. Federated learning (FL) has emerged as a crucial approach to address this challenge. Zhou et al. (2021) and Fernandez et al. (2023) emphasized FL's increment of competitiveness to data-scarce MSMEs.

Recommendation for single merchants may rely on product IDs. However, relying on individual-specific taxonomies becomes impractical when developing scalable AI for commerce encompassing multiple merchants. Therefore, a generalized commercial AI engine requires an NLP-based model that allows for the flexible interpretation of data. Sun et al. (2019), Yadav and Singh (2020), Lian and Li (2020), and Moreira et al. (2021) utilized a Transformer (Vaswani et al. 2017) to create recommendation systems; however, they were not NLP-based. Prior research has employed NLP methodology to construct product-level recommendation systems; however, there has been limited application of utilizing both merchants and product names within the context of natural language processing.

Architecture of GCI

A crucial aspect of commerce is the ability to forecast customers' subsequent purchases. GCI relies on receipt data as its primary source (Figure 1). We developed a commerce model that leverages NLP and transformers to cater to international and multivendor commerce. The performance based on language models improves as the foundational models are enhanced. This principle also applies to the GCI; as the GCI improves, the performance of its derivative services is also expected to increase.



Transaction Date	Merchant Name	Product Name
2023-05-02 21:20	ABC Chicken	Half & half boneless chicken
2023-05-20 18:24	DEF Gimbap	Cheese Gimbap
2023-05-27 18:27	GHI Korean Food	Galbi-tang
2023-06-04 15:27	JKL Chinese Food	Jajangmyeon
2023-06-21 11:22	MNO Burger	Bacon Cheese Burger
2023-07-03 18:34	PQR Korean Food	Jeyuk-bibimbap
2023-07-11 14:36	ABC Chicken	Half & half boneless chicken
2023-07-15 12:40	STU Cafe	Iced Vanilla Latte
...	...	...

Figure 1: Merchants and Product Names from Receipt Data

While developing the GCI, we encountered two primary challenges: securing training data and delivering services to end users sustainably. Users are often apprehensive about misusing their data. While they desire high-quality AI services, they are reluctant to risk the excessive leakage, movement, or analysis of their information. Similarly, MSMEs and sole proprietors are particularly sensitive to the data. They are generally open to collaboration but hesitant to share their valuable data assets. This concern is amplified when engaging with big tech platforms, as these businesses often feel powerless, fearing that the platform provider will entirely control their data. To address these concerns, we chose to train the GCI using FL. This approach allowed us to develop an AI platform that balances advancing the system's capabilities and safeguarding user privacy and business datasets. As depicted in Figure 2, the GCI is designed to offer merchants various services, including product recommendations, product brainstorming, product bundling, event promotions, collaborative marketing, target marketing, and demand forecasting.

The UB Platform, managed by Harex InfoTech, an e-payment company in Korea, maintains purchase receipt records through its payment service, UBpay. These records, essential for returns, refunds, and partial exchanges, detail the products and services sold by partner companies to their customers when and where these transactions occurred. The anonymized data served as the GCI's primary training and testing resources.

To safeguard user privacy, this process intentionally did not incorporate customer ID or other personally identifiable information from each partner into the training data. This practice aligns with the fact that OpenAI's GPT series does not use information about the author of a text to predict the next word. Although access to customer demographics or identity could potentially enhance the prediction accuracy for their next purchase, the GCI model was designed to avoid using such meta-information.

The absence of specific customer and product IDs sets the GCI apart from stand-alone recommendation systems, which may employ algorithms that utilize customer ID, demographic information, and product ID. With the GCI's focus on serving a multi-merchant commerce platform, it generally does not have access to use customer demographic information or product IDs. Owing to the pivotal importance of securing user data and a general hesitancy to disseminate information amongst competing stores, the platform is precluded from accessing or utilizing individual store data. Therefore, it relies on natural language data, such as product or service names and descriptions, along with any available purchase history specific to each merchant, to pragmatically and effectively forecast a customer's next purchase.

NLP-Based Recommendation Models for GCI

We developed NLP-based models by analyzing the purchased product names and associated merchant names (Hwangbo et al. 2021). For the inaugural GCI model, we

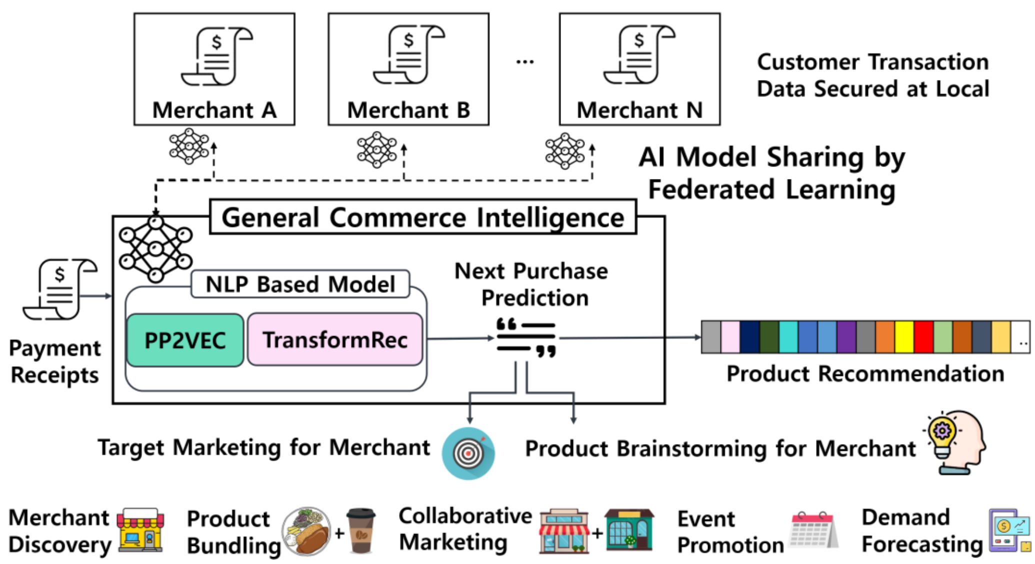


Figure 2: Architecture of GCI Engine and its Derived Services

introduced PP2Vec (Purchased Product to Vector), inspired by Word2Vec, which converts purchased product names into vector representations. We can identify products likely to be bought by users with similar purchasing patterns and items related to those they have previously purchased. Notably, this is achieved without delving into the user's details, thereby upholding user privacy while offering personalized recommendations.

TransformRec is a recommendation system that employs NLP techniques to analyze the relationships among words within the names of products purchased by users and the corresponding merchant names. This system represents an evolution of a previous extrapolative collaborative filtering recommendation model designed for a multi-merchant environment (Lee et al. 2021). Unlike earlier models, which were premised on sharing user data and purchase histories across various merchants, TransformRec focuses exclusively on the text data of products and merchant names (Lee et al. 2022a). This eliminates the need to identify similar users based on the shared data, thereby formulating recommendations that preserve user privacy.

We employ a space-based tokenization to identify similar products and merchants, wherein the product and merchant names are broken down into discrete tokenized units. This tokenization strategy facilitates the exploration of semantic similarities between product and merchant names, even when only partially identical words are present. Such a tokenization approach allows for a more comprehensive analysis of the relationships between the constitutive words of a product-merchant pairing. Figure 3 illustrates the enhanced connectivity enabled by the tokenization strategy. Without tokenization, the learning process is confined to establishing links between wholly identical merchant product names. For example, without tokenization, “Seafood Pizza” sold by “Pizzaya” and “Cheese Pizza” also sold by “Pizzaya” would be treated as separate, unrelated products during the learning process. Under this framework, only identical products sold by the same merchant, such as “Salmon Steak” sold by “Pocket,” would be linked.
However, tokenization establishes connections between product-merchant pairs even when only a subset of the tokens is identical. Consider the products "Seafood Pizza" and "Cheese Pizza," both sold by the merchant "Pizzaya”. Without tokenization, these two are treated as distinct products with no relationship; thus, they are not linked to understanding the system. Through tokenization, the common tokens "Pizzaya" and "Pizza" establish a connection between the two products. The system learns that the same merchant made purchases involving a pizza-based product. Similarly, tokenization enables a link between the "Seafood" token in "Seafood Pizza" sold by "Pizzaya" and the "Seafood" token in “Seafood Pasta” sold by "Lifestyle”. The system learns from the shared characteristics of seafood-containing dishes.

By tokenizing and analyzing the names of product-merchant pairs, we can thoroughly examine the relationships among transactions. This enables the proposed system to discern nuanced connections between products and merchants. To quantify the similarity between the tokenized pairs, we employed the Jaccard similarity metric. Based on this metric, our model generates recommendations by identifying product-merchant pairs with the highest similarity to the input query. This approach allows our model to make meaningful recommendations adaptively, even when new products and merchants are introduced into the system.

As the amount of training data increases, the same product-merchant pairs are repeated within the dataset, resulting in redundancy. To address this issue, we adopted a data pre-processing strategy inspired by Chen et al. (2019), as shown in Figure 4. We divided the transaction data into five units and assigned the last product-merchant pair within each unit as the label. To generate a comprehensive dataset, we generated an additional set of product-merchant pairs by shifting two pairs from each of the five transactions.
The proposed methods work well with limited data, suitable for MSMEs, unlike systems designed for big firms, potentially reducing AI gaps between SMEs and larger companies. A salient feature of this algorithm is its inherent language-agnostic design adaptable to various linguistic contexts. This versatility underscores the immense scalability of multi-merchant platforms underpinned by TransformRec, indicating its capacity to facilitate services in multilingual environments.

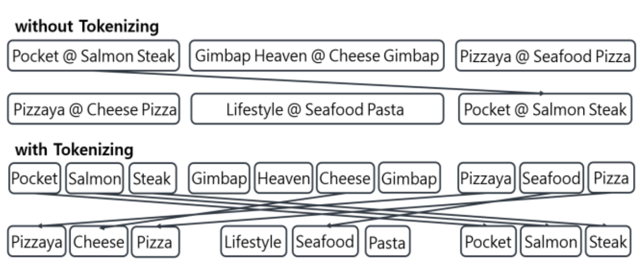


Figure 3: Tokenizing Product-Merchant Names

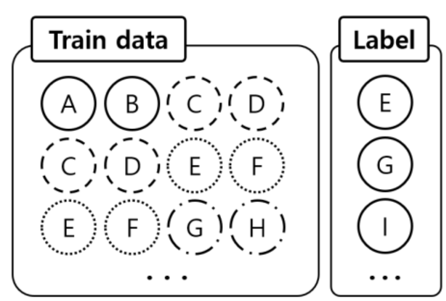


Figure 4: Preprocessing of Product-Merchant Pairs

Deriving Various Services from GCI

The product-merchant names obtained from GCI can be used to generate diverse services. Target marketing services are generated by swapping users and products in product-merchant lists. Forecasting can be performed by scrutinizing the product features in the recommendations. Synthesizing recommendation lists generates product bundling and collaborative marketing strategies. Furthermore, product brainstorming functions can be provided through analyzing non-existent product names in the recommendations. Event promotions can be generated by considering the context of the time and place of product-merchant recommendations.

Recommendation as Both a Service and a Source

The GCI generates personalized product-merchant pairs tailored to individual consumers. However, they are not static; they can be dynamically adjusted to accommodate various domain constraints and the unique circumstances of users or merchants. For instance, a customer's physical location is a crucial consideration, especially in food delivery services where the merchant's delivery radius may be limited. Thus, the lists of recommended pairs function dually; they serve as a service tailored to users and act as a source from which various added services can be derived.

Target Marketing as the Dual Recommendation

Recommending a product “A” for a merchant “B” to a customer “C” serves as a personalized suggestion. This process exhibits duality as it can become a targeted marketing strategy for merchants. In this reversed scenario, the merchant “B” can use the GCI to identify potential customers, like “C”, who are likely to be interested in the product “A”. This is accomplished by intersecting the list of users to whom the product “A” is recommended. Through this duality, the GCI system personalizes the shopping experience for customers and empowers merchants with actionable insights for precise and effective marketing.

Forecasting via Analysis of the AI-Generated Data

Each product suggested by the GCI can be interpreted as an item a user is highly likely to purchase. This predictive insight offers merchants a window into prospective buying behaviors. By systematically analyzing the frequency with which certain products appear in these recommendations, merchants can anticipate the demand for specific items. Furthermore, extracting pertinent keywords from AI-generated product-merchant pairings allows for identifying emerging trends in customer preferences. This is a notable example of how AI-generated data can be repurposed to provide actionable insights beyond their original intent.

Product Bundling and Collaborative Marketing

By synthesizing product-merchant pairs that frequently appear together in recommendations, the GCI engine provides a powerful strategy for product bundling. Product bundles often purchased together can stimulate customers' purchasing behavior, effectively encouraging them to buy more than one item at a time.

In addition to product bundling, we propose a unique approach that involves manual collaboration between merchants. This collaboration aims to create and market joint bundles of products from different merchants that customers will likely purchase together. This offers customers compelling package deals and fosters partnerships and collaborative marketing opportunities between merchants.

Brainstorming by Serendipitous Recommendation

Significant discoveries often emerged from unexpected results or perceived failures during experiments. GCI can yield product names that may not currently exist. While such outcomes might be perceived as anomalies, especially within a commercial context, these "anomalies" can be used as potential opportunities for product brainstorming and innovation. Essentially, these generated names can serve as creative stimuli, providing a springboard for conceptualizing and developing novel products that do not yet exist.

For example, the GCI engine generated food names such as "Rose Sauce Fried Chicken Set", “Stir-fried Spicy Toast”, “Small Octopus Cream Spaghetti”, and "Red Pepper Egg Tarts”, none of which were found to exist at the time of generation. Nonetheless, subsequent online searches revealed the existence of recipes for "Stir-fried Spicy Toast" and "Small Octopus Cream Spaghetti”, demonstrating the real-world feasibility of these dishes. Notably, in a compelling illustration of the potential impact of the GCI, Genesis BBQ, a partner in this research, launched a new product called the “Rose Sauce Fried Chicken Set” in Apr. 2022, a name originally generated by the GCI.

Promotion Tailored to Contextual Information

Transaction data include product names and purchase timestamps, offering contextual information. An Event Promotion service leverages this context to inform and enhance the recommendations it generates, aiming to make these suggestions more relevant and timelier for the end user.

These context-aware recommendations can be configured to reflect different time granularities, such as specific months, dates, or days of the week. This feature extends to recognizing and incorporating recurring patterns at regular intervals, such as specific anniversaries that may occur annually (e.g., a birthday) or monthly (e.g., payday). This level of detail fosters the development of a personalized and effective promotional strategy. Figure 5 shows a sample event promotion that incorporated these dates. In this example, the

Figure 5: Promotion Incorporating Date

recommendations generated are sensitive to the contextual significance of specific dates, as evidenced by the system suggesting a Christmas cake on Dec. 24, acknowledging the cultural and temporal significance of Christmas Eve.

Validation of the Approach

To validate the approach, we acquired transaction data from two collaborative partners: Ulsan Pedal, a multivendor food delivery application operated by the city of Ulsan, and Genesis BBQ, a leading chicken franchise in South Korea. The dataset from Ulsan Pedal comprised approximately 200,000 entries spanning from Mar. 2021 to May 2023, whereas the dataset from Genesis BBQ contained approximately 350,000 entries recorded from Feb. 2023 to May 2023. Additionally, data from BBQ's online platform were collected, encompassing 9,663 events involving 3,678 users from Apr. 2022 to Nov. 2022. Our research uses real store names and product names in the data. The Ulsan pedal data is multi-merchant, and there are various merchant names and their product and merchant names, such as Kimbap Heaven, Spicy Pork Bulgogi over Rice, and Mr. Pizza. Ulsan Pedal has 2,655 merchants, and BBQ has 2,002 merchants.

To accommodate the Federated Learning (FL) setting, the data from both groups were divided into two datasets excluding the BBQ ordering app data, which was small in size, resulting in four distinct datasets or clients. The Ulsan Pedal data were partitioned based on the merchant's name, reflecting its multi-merchant nature, whereas the BBQ data were segmented according to unique customer numbers. After preprocessing, the final data counts were recorded: BBQ1 contained 20,138 entries; BBQ2 19,869 entries; Ulsan Pedal 1 had 34,203 entries; and Ulsan Pedal 2 33,894 entries.

Data pre-processing involves tokenizing each client's data space. A tokenizer was constructed based on the tokenized data. After the tokenization process, the natural language data were numerically encoded using this tokenizer, a process like Label Encoding. After preprocessing, the datasets were allocated to distinct training, validation, and test sets. The data were partitioned in a 9:1 ratio and designated as training and validation sets. Table 1 illustrates the performance of TransformRec when applied to the Ulsan Pedal. To evaluate the effects of tokenization, a comparative analysis was conducted between TransformRec with and without tokenization, using Hit-Rate (HR) as a metric. TransformRec incorporated with tokenization exhibited superior performance.

	HR@3	HR@5	HR@10
TransformRec	5.604%	6.295%	7.329%
TransformRec without Tokenizing	0.843%	1.687%	2.182%
Random	0.010%	0.024%	0.044%

Table 1: Performance with/without Tokenization

HR@10	Ulsan Pedal	BBQ Order App	BBQ Mall
PP2Vec	4.1%	7.8%	8.1%
TransformRec	6.0%	28.0%	14.9%

Table 2: Comparison of PP2Vec and TransformRec

Table 2 presents a comparison of PP2Vec and TransformRec. Through tokenizing, TransformRec better understands the relationships between tokens, leading to a significant enhancement in its recommendation capabilities.

Federated Learning (FL) with Glocalization

The efficacy of FL has been previously substantiated in the medical field for convolutional neural network models, as corroborated by existing studies (Kaissis et al. 2021; Liu et al. 2022; Pati et al. 2022). Several studies have applied FL to Transformer (Park et al. 2021; Thwal et al. 2023). Lin et al. (2021) was the first to apply FL specifically to Transformer models for NLP tasks.

We implemented FL rather than sharing the entirety of the Transformer model weights using a glocalization technique. Glocal is a portmanteau term for globalization with a local twist. This technique involves sharing only the selected layers, as shown in Figure 6. Specific layers are selectively included or excluded from the FL process. This selective inclusion was based on the properties and roles of the layers. The embedding layers of the Encoder, the Decoder, and the output layer was excluded from the FL. One reason for this exclusion is that these layers are highly influenced by the vocabulary size, which leads to fluctuating parameter counts that may not be consistent across different clients or nodes in the FL setup. In contrast to the excluded layers, the internal layers within both the Encoder and the Decoder are actively engaged in FL. These layers are designed to share consistent parameters across clients or nodes.

This was done with the FedAvg (McMahan et al. 2017). The specific counts of the parameters across the different

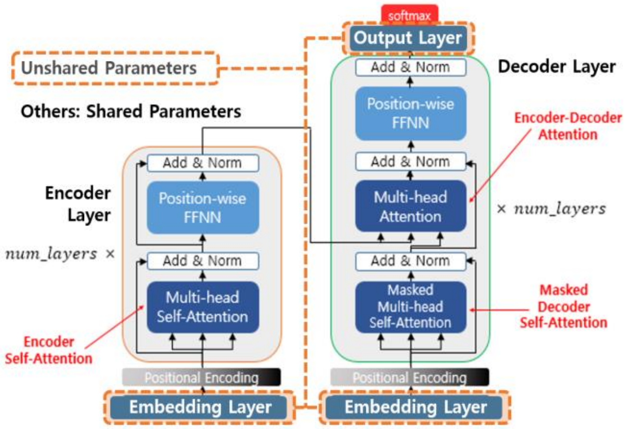


Figure 6: FL of Transformer with Glocalization

	BBQ 1	BBQ 2	Ulsan Pedal 1	Ulsan Pedal 2
Encoder Layer	132,480			
Decoder Layer	198,784			
Shared Total	331,264			
Encoder Embedding Layer	47,104	47,616	2,446,464	2,379,520
Decoder Embedding Layer	47,104	47,616	2,446,464	2,379,520
Output Layer	47,472	47,988	2,465,577	2,398,110
Unshared Total	141,680	143,220	7,358,505	7,157,150
Total	472,944	474,484	7,689,769	7,488,414

Table 3: Number of (Un)Shared Parameters of Each Model

layers of the transformer architecture under this FL setup are summarized in Table 3.

The unshared parameters differed substantially, with vocabulary sizes for BBQ1, BBQ2, Ulsan Pedal 1, and Ulsan Pedal 2 of 368, 372, 19,113, and 18,590, respectively. The hyperparameters set for both the dataset and model included NUM_LAYERS = 1, D_MODEL = 128, NUM_HEADS = 4, UNITS = 256, DROPOUT = 0.2, EPOCHS = 1, BATCH_SIZE = 128, BUFFER_SIZE = 20,000, and ROUNDS = 500, with an early stopping mechanism based on the validation loss and patience of 20 rounds. The training was concluded in the 175th round because of early stopping, and it took approximately 7,000 seconds, which was approximately 117 minutes.

Evaluation and Results

For each entry in the test set, the model is tasked with producing ten predictions corresponding to the potential products the user is likely to purchase. In the generated predictions, certain merchant or menu names may not correspond to actual genuine names in the dataset. In such instances, an adjustment was made to substitute the non-genuine names with the closest genuine match. The Jaccard similarity measure was employed to quantify the closeness of matches. Performance was evaluated using the HR@10 and Precision@10 metric, calculated as the proportion of cases in which the item appears within the top 10 recommended items. The evaluation results are listed in Table 4.

The local models (a) and data-sharing models (c) exhibit signs of overfitting (Figure 7). This is indicated by the lack of a further decrease in the validation loss as the training progressed. FL yielded superior results for all clients compared to Local and Data-Shared training approaches. Interestingly, training using shared data resulted in poorer outcomes than training using locally trained models. This finding highlights the significant implication that FL offers more effective solutions than mere data aggregation when data vary extensively across different contexts.

		BBQ 1	BBQ 2	Ulsan Pedal 1	Ulsan Pedal 2
HR @10	Local	0.443	0.455	0.110	0.117
	Glocal FL	0.478	0.522	0.147	0.140
	Data-Shared	0.425	0.404	0.097	0.099
Precision @10	Local	0.044	0.046	0.011	0.012
	Glocal FL	0.048	0.052	0.015	0.014
	Data-Shared	0.041	0.041	0.006	0.006

Table 4: FL with Glocalization is the Best

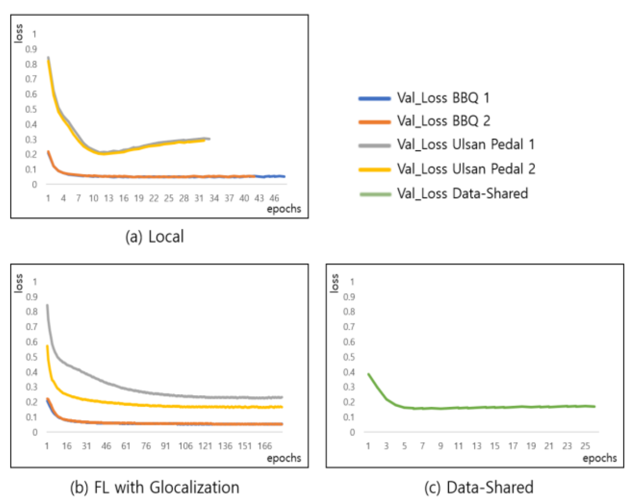


Figure 7: Validation Loss Trends

Glocalized FL in NLP presents a solution uniquely tailored to clients with varying vocabulary sizes who are hesitant to disclose their tokens. This strategy enhances the client and data security. Moreover, it fosters an increased willingness among clients to collaborate, as they are not mandated to disclose their entire model.

With glocalization, clients must contribute only a selected portion of their model - specifically, one trained on their proprietary data - rather than sharing the entirety of their model. This structure encourages partners to be more active and willing to participate in FL initiatives. Additionally, it benefits central platforms by ensuring they are not required to share the entire initial model investment with all participants. This effectively circumvents the potential issues associated with free riders or entities that might seek to benefit without contributing equally to the collaborative effort.

Glocalized FL facilitates seamless vocabulary expansion on various platforms. If a client's dataset is insufficient for an effective FL, they can selectively share specific tokens with the main platform, augmenting the model's vocabulary without compromising data security. Previous studies have opted for selective parameter federations (Mishchenko et al. 2023; Liang et al. 2020; Sun et al. 2021; Kalra et al. 2023; Qi et al. 2023). However, to our knowledge, none of these studies has addressed the specific issue of token information exchange or standardization within this context.

Deployment

As depicted in Figure 8, the backend architecture equipped with GCI utilizes a method where multiple AI servers (clients during federated learning) independently manage the integrated storage of model weights.

Leveraging the Amazon Simple Storage Service (S3) object storage system, it plays a pivotal role in mediating the upload and download of model weights generated by each local AI server and the global server. The local AI servers and the global server are configured with Amazon Elastic Compute Cloud (EC2). The global server aggregates weights uploaded to S3 from local AI servers and generates new synthetic weights using a federated learning algorithm. After completing a learning cycle (a set number of epochs), local AI servers upload their weights to S3 and download the aggregated and updated weights from the global server to continuously improve their models. This architecture is scalable and can consolidate the overall knowledge necessary to enhance the performance of the entire model without centralizing local data.

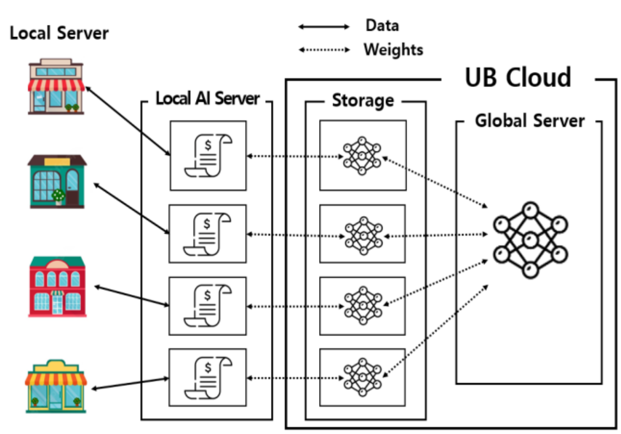


Figure 8: Architectural Design for Back-end Implementation of Federated Learning

Figure 9 shows an operational example of a service provided through GCI: (a) shows the app screen of the Ulsan Pedal and (b) shows the order delivery app of a chicken franchise. An AI recommendation service user leaves a review of the service as a post on social media, as shown in (c).

From July 22, 2022, to June 26, 2023, Ulsan Pedal recorded 29,933 purchases. Of these transactions, 1,891 orders were generated through AI recommendations, yielding a hit-rate of 6.32%. Amazon uses its AI recommendation system effectively, predicting with about 5% accuracy what customers want to buy from the millions of products it offers (Agrawal et al. 2017). The Ulsan Pedal is currently in its initial phase, indicating that the dataset it contributes to the FL process is in a nascent stage and is anticipated to expand. As the FL network attracts additional partners, we project enhancements to the hit-rate. Based on the current pace of collaboration and data integration, there is potential for an upward trajectory in the hit-rate of the algorithm. Our projections suggest that this key performance indicator may achieve double-digit figures in future assessments as the federated network evolves.

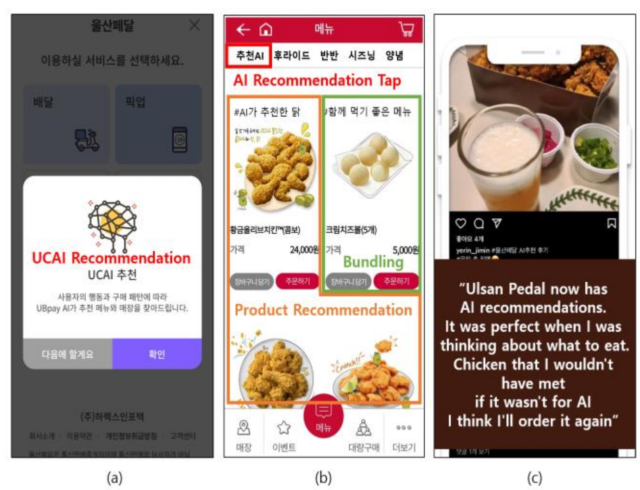


Figure 9: GCI Service Screens and Service Testimonials

Overview of the Development Process

This project, representing significant industry-academy collaboration, extended over three years. Its overarching objective was cultivating a thriving global commerce ecosystem that caters to MSMEs by deploying User Centric Artificial Intelligence (UCAI) technologies and business models. A central component of this initiative involved developing and refining the GCI, a novel framework designed to enhance the efficacy of commerce-related algorithms. Concurrently, this project marked a concerted foray into experimental trials with FL to establish robust, scalable, and privacy-preserving mechanisms for the collaborative training of AI models across multiple merchants. The chronological journeys and key milestones are listed in Table 5.

In the first year, our research efforts centered on exploring User Centric Artificial Intelligence (UCAI) and commerce models driven by NLP. This foundational phase culminated in the developing of a recommendation model based on Transformer. The following year, we expanded our scope to encompass deploying various AI services, including recommendations, bundling, and target marketing. The Ulsan Pedal became the first platform to integrate AI services during this period. The integration of GCI with FL was a milestone in the third phase. The third year witnessed the initiation of research on generative AI. This line of inquiry manifests itself in developing conversational commerce services and persuasive personalized messaging engines.

	1st Year	2nd Year	3rd Year
GCI	·PP2Vec ·TransformRec	·GCI Engine with Derived Services	· GCI with Generative AI
Service Development	·Product Recommendation	·Derived Service Pilot (e.g. Product Brainstorming)	·Conversational Service Pilot
Service Deployment	N.A.	· Deployed to Ulsan Pedal with Product Recommendation and Target Marketing Services	·Deployed to BBQ with Recommendation, Target Marketing, and Event Promotion
Performance Test	·Model Test with Open Data	·Model Test with Real Data (e.g., Ulsan Pedal Data)	·A/B Test on Real Data (e.g., Ulsan Pedal Data)
FL	·FL as a UCAI tool	·FL with Glocalization	·FL in real setting

Table 5: Project Timelines

Summary and Conclusion

The project was started initially to realize the vision of the UCAI. Over the course of the project, we have come to define the UCAI and realized that it is a research project that applies the UCAI to real-world online and offline commerce. To date, the team has defined UCAI as follows:1) Pursue sustainability of AI providers in achieving user-centered goals as much as possible. 2) Data and services are handled with the convenience and benefits of the user in mind and not from the perspective of the AI provider. 3) Protecting the privacy of individuals and business users as much as possible.

The UCAI's vision is not only applicable to commerce but also to managing personal lives. We developed the AMPER (Aim-Measure-Predict-Evaluate-Recommend) structure (Lee et al. 2022b) for Digital Me services, which can be defined as an AI-based product and service combination system that enables real-time management and improvement of a person's health, financial, and psychological states. AMPER measures the user's current state, predicts the user's future state, evaluates the user's future state, and recommends desirable actions to reach the target state and maximize the user's state improvement. To verify the proposed AMPER structure, we used data from EdNet (Choi et al. 2020) and MIMIC-III (Johnson et al. 2016). In the case of healthcare, it is undesirable to combine or share user data because they contain sensitive information. Therefore, Personalized FL was applied to the prediction algorithm of AMPER, and discovered the PFL Model improves its performance by 31.37% compared with the Local Model.

The UB Platform is designed as a User Centric Artificial Intelligence (AI) sharing platform to serve individual consumers and MSMEs. This design is grounded in the methodologies delineated in this study and seeks to actualize the vision of the UCAI. Unlike traditional models in which service-providing platforms acquire and control user data, the UB Platform adopts a contrasting approach. The platform furnishes AI models in this structure, whereas individual users and MSMEs retain ownership and control of their respective datasets (Piao et al. 2023).

This emphasis on privacy preservation and user data ownership will likely encourage wider and more enthusiastic participation in AI service platforms. This has the potential to enhance the long-term sustainability of platforms. Moreover, mitigating the concentration of data control may reduce the risk of societal polarization attributable to AI technologies. Ultimately, this framework contributes to the broader democratization of AI, promoting equitable access to and benefits from these advanced technologies.

References

Agrawal, A.; Gans, J.; and Goldfarb, A. 2017. How AI will Change Strategy: A Thought Experiment. *Harvard Business Review*, 2-6.

Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. 2020. Language Models are Few-shot Learners. *Advances in Neural Information Processing Systems*, 33, 1877-1901.

Björkegren, D. 2023. Artificial Intelligence for the Poor - How to Harness the Power of AI in the Developing World. *Foreign Affairs*.

Chen, Q.; Zhao, H.; Li, W.; Huang, P.; and Ou, W. 2019. Behavior Sequence Transformer for e-commerce Recommendation in Alibaba. *In Proceedings of the 1st International Workshop on Deep Learning Practice for High-Dimensional Sparse Data*, 1-4.

Choi, Y.; Lee, Y.; Shin, D.; Cho, J.; Park, S.; Lee, S.; Baek, J.; Bae, C.; Kim, B.; and Heo, J. 2020. EdNet: A Large-scale Hierarchical Dataset in Eduation. *In Artificial Intelligence in Education: 21st International Conference*, Ifrane, Morocco, II (21), 69-73.

Fernandez, J. D.; Brennecke, M.; Barbereau, T.; Rieger, A.; and Fridgen, G. 2023. Federated Learning: Organizational Opportunities, Challenges, and Adoption Strategies. arXiv:2308.02219.

Hwangbo, Y.; Lee, K.; Jeong, B.; and Park, K. 2021. Recommendation System with Minimized Transaction data. *Data Science and Management*, 4, 40-45.

Johnson, A.; Pollard, T.; and Mark, R. 2016. MIMIC-III Clinical Database (version 1.4). PhysioNet. https://doi.org/10.13026/C2XW26.

Kaissis, G.; Ziller, A.; Passerat-Palmbach, J.; Ryffel, T.; Usynin, D.; Trask, A.; Jr, I. L.; Mancuso, J.; Jungmann, F.; Steinborn, M. M.; et al. 2021. End-to-end Privacy Preserving Deep Learning on Multi-institutional Medical Imaging. *Nature Machine Intelligence*, 3(6), 473-484.

Kalra, S.; Wen, J.; Cresswell, J. C.; Volkovs, M.; and Tizhoosh, H. R. 2023. Decentralized Federated Learning through Proxy Model Sharing. *Nature Communications*, 14(1), 2899

Lee, K.; Hwangbo, Y.; Jeong, B.; Yoo, J.; and Park, K. 2021. Extrapolative Collaborative Filtering Recommendation System with Word2Vec for Purchased Product for SMEs. *Sustainability*, 13, 7156.

Lee, K.; Hwangbo, Y.; Jeong, H.; Jeong, B.; and Park, J. 2022a. TransformRec: A User-Centric Recommender System for e-commerce using Transformers. *In Proceedings of the 23rd International Conference on Electronic Commerce*. Daegu, Korea.

Lee, K.; Jeong, B.; Hwangbo, Y.; Kim, Y.; Bae, S.; and Baek, T. 2022b. AMPER(Aim-Measure-Predict-Evaluate-Recommend):. The Paradigm of Digital Me. *In Proceedings of the 23rd International Conference on Electronic Commerce*. Daegu, Korea.

Lian, M.; and Li, J. 2020. Financial Product Recommendation System based on Transformer. *In 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference*, 1, 2547-2551.

Liang, P. P.; Liu, T.; Ziyin, L.; Allen, N. B.; Auerbach, R. P.; Brent, D.; Salakhutdinov, R.; and Morency, L. P. 2020. Think Locally, Act Globally: Federated Learning with Local and Global Representations. arXiv:2001.01523.

Lin, B. Y.; He, C.; Zeng, Z.; Wang, H.; Huang, Y.; Dupuy, C.; Gupta, R.; Soltanolkotabi, M.; Ren, X.; and Avestimehr, S. 2021. FedNLP: Benchmarking Federated Learning Methods for Natural Language Processing Tasks. arXiv:2104.08815.

Liu, Z.; Chen, Y.; Zhao, Y.; Yu, H.; Liu, Y.; Bao, R.; Jiang, J.; Nie, Z.; Xu, Q.; Yang, Q. 2022. Contribution-aware federated learning for smart healthcare. *In Proceedings of the AAAI*, 36:11, 12396-12404.

McMahan, B.; Moore, E.; Ramage, D.; Hampson, S.; and Arcas, B. A. 2017. Communication-efficient Learning of Deep Networks from Decentralized Data. *In Artificial Intelligence and Statistic*s, 1273-1282.

Mishchenko, K.; Islamov, R.; Gorbunov, E.; and Horváth, S. 2023. Partially Personalized Federated Learning: Breaking the Curse of Data Heterogeneity. arXiv:2305.18285.

Moreira, G.; Rabhi, S.; Lee, J. M.; Ak, R.; and Oldridge, E. 2021. Transformers4Rec: Bridging the Gap between NLP and Sequential/Session-based Recommendation. *In Proceedings of the 15th ACM Conference on Recommender Systems*, 143-153.

Park, S.; Kim, G.; Kim, J.; Kim, B.; and Ye, J. C. 2021. Federated Split Vision Transformer for COVID-19 CXR Diagnosis using Task-agnostic Training. arXiv:2111.01338.

Pati, S.; Baid, U.; Edwards, B.; Sheller, M.; Wang, S. H.; Reina, G. A.; Foley, P.; Gruzdev, A.; Karkada, D.; Davatzikos, C.; et al. 2022. Federated Learning Enables Big Data for Rare Cancer Boundary Detection. *Nature Communications*, 13:1, 7346.

Perrigo, B. 2023. The Indian Startup Making AI Fairer - While Helping the Poor. *Time Magazine*.

Piao, Y.; Qiu, Y.; Jeong, B.; and Lee, K. J. 2023. An Explorative Study for Federated Learning-based AI Sharing Platform Development Methodology. *In Proceedings of the 24th International Conference on Electronic Commerce*. Xi'an, China.

Qi, T.; Wu, F.; Wu, C.; He, L.; Huang, Y.; and Xie, X. 2023. Differentially Private Knowledge Transfer for Federated Learning. *Nature Communications*, 14(1), 3785.

Seger, E.; Ovadya, A.; Siddarth, D.; Garfinkel, B.; and Dafoe, A. 2023. Democratising AI: Multiple Meanings, Goals, and Methods. arXiv:2303.12642.

Sun, B.; Huo, H.; Yang, Y.; and Bai, B. 2021. PartialFed: Cross-domain Personalized Federated Learning via Partial Initialization. *In Advances in Neural Information Processing Systems*, 34, 23309-23320.

Sun, F.; Liu, J.; Wu, J.; Pei, C.; Lin, X.; Ou, W.; and Jiang, P. 2019. BERT4Rec: Sequential Recommendation with Bidirectional Encoder Representations from Transformer. *In Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, 1441-1450.

Thwal, C. M.; Tun, Y. L.; Kim, K.; Park, S. B.; and Hong, C. S. 2023. Transformers with Attentive Federated Aggregation for Time Series Stock Forecasting. *International Conference on Information Networking*, 499-504.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; and Polosukhin, I. 2017. Attention is All You Need. *Advances in Neural Information Processing Systems*, 30.

Yadav, N.; and Singh, A. K. 2020. Bi-directional Encoder Representation of Transformer Model for Sequential Music Recommender System. *In Proceedings of the 12th Annual Meeting of the Forum for Information Retrieval Evaluation*, 49-53.

Zhou, J.; Zhang, S.; Lu, Q.; Dai, W.; Chen, M.; Liu, X.; Pirttikangas, S.; Shi, Y.; Zhang, W.; and Herrera-Viedma, E. 2021. A Survey on Federated Learning and its Applications for Accelerating Industrial Internet of Things. arXiv:2104.10501.