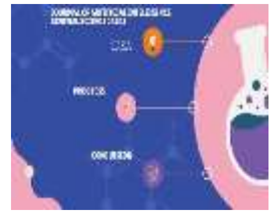
Vol.4, Issue 01, May, 2024
Journal of Artificial Intelligence General Science JAIGS

Home page http://jaigs.org



Towards Improved Privacy in Digital Marketing: A Unified Approach to User Modeling with Deep Learning on a Data Monetization Platform

Bhuvi Chopra[1],Vinayak Raja[2]

[1]Product manager, Google

[2]Software engineer, Meta

ARTICLEINFO
Article History:
Received:
01.04.2024
Accepted:
15.04.2024
Online: 13.05.2024
Keyword; Privacy Enhancement, Digital Marketing, Data Privacy, Machine Learning, Privacy-Preserving Techniques

ABSTRACT

This paper introduces an innovative method for safeguarding user privacy in digital marketing campaigns through the application of deep learning techniques on a data monetization platform. This framework empowers users to maintain authority over their personal data while enabling marketers to pinpoint suitable target audiences. The system consists of several key stages Data representation learning in hyperbolic space captures latent user interests across various data sources with hierarchical structures. Subsequently, Generative Adversarial Networks generate synthetic user interests from these embedding. To preserve user privacy, Federated Learning is utilized for decentralized user monetization, Data privacy, modeling training, ensuring data remains undisclosed to marketers. Lastly, a hyperbolic embedding, Federated learning targeting strategy, rooted in recommendation systems, utilizes learned user interests to identify optimal target audiences for digital marketing campaigns. In sum, this approach offers a comprehensive solution for privacy-preserving user modeling in digital marketing.

Introduction

In today's digital landscape, the utilization of personal data stands as a pivotal element in marketing strategies for both businesses and consumers. For businesses, it facilitates the creation of targeted campaigns, enhancing the likelihood of conversions by delivering tailored messages to specific audiences. Additionally, personal data offers valuable insights into consumer behavior, preferences, and trends, which inform strategic business decisions and enhance products and services. Conversely, for consumers, personalized and relevant advertising fosters engagement and aids in decision-making processes, ultimately enhancing the overall customer experience. However, amidst these benefits, consumers are increasingly privacy-conscious and concerned about the control over their personal information, particularly as tech giants leverage data access for competitive advantage. Moreover, consumers are beginning to recognize the financial value of their data, prompting demands for tangible returns.

In response to these concerns, privacy regulations such as the European Union's General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA) mandate stringent measures to protect consumer privacy in digital marketing practices. This regulatory landscape significantly impacts marketing strategies, influencing the effectiveness of targeted marketing and necessitating explorations of alternative avenues for data collection and analysis in compliance with regulations.

A novel solution arises in the form of data monetization platforms, facilitating the exchange of granular personal information between consumers and marketers for mutual benefit. These platforms operate within a two-sided market framework, enabling commercial transactions of data between consumers and marketers. Consequently, these platforms not only introduce new data-based business models but also pave the way for tangible value co- creation. Examples abound where consumers receive cash or discounts in exchange for demographic and behavioral data, highlighting the potential of these platforms in fostering fair and transparent data transactions.

This study, conducted in collaboration with a European data monetization platform, aims to enhance user and marketer experiences within the platform while ensuring compliance with data privacy regulations, particularly GDPR. The proposed approach offers a comprehensive solution for privacy-preserving user modeling in digital marketing campaigns. Leveraging hierarchical user interests represented in hyperbolic space through hyperbolic embeddings, this approach captures consumer behavior patterns across diverse data sources. To safeguard consumer privacy, synthesized user representations are generated using Generative Adversarial Networks (GAN), preserving user interests while maintaining indistinguishability from the original data. Training occurs through Federated Learning (FL), a distributed learning method ensuring data privacy and communication efficiency. This solution empowers the platform to identify target audiences matching specific campaigns while adhering to data privacy regulations and granting users control over their data. Marketers can specify consumer characteristics, such as interests, demographics, or online behavior, to create targeted user lists, optimizing campaign reach and effectiveness.

The paramount importance of data privacy in digital marketing underscores the significance of the proposed approach. By safeguarding user personal data while facilitating effective targeting for marketers, this study contributes to the field of e-commerce marketing by striking a balance between personalization and data privacy. Advanced deep learning techniques, including representation learning in hyperbolic space, GAN, and FL, ensure accurate targeting and personalization without compromising user privacy or violating regulations.

The paper proceeds as follows: Section 2 reviews related work covering representation learning, privacy-preserving machine learning approaches, and user-campaign matching mechanisms. Section 3 provides a detailed explanation of the system design. Section 4 outlines the data utilized in this study from the collaborating data monetization platform and presents results from each step of the proposed approach. Section 5 discusses the business implications and concludes the study.

Related Work:

Representation Learning for Hierarchical User Data:

Representation learning aims to extract meaningful features from raw data, presenting them in a more efficient and informative manner. In the realm of digital marketing campaigns, users often exhibit hierarchical structures in their interests and behaviors, posing a challenge in representing such hierarchical user data in a low-dimensional space while preserving their hierarchical relationships and similarities. Platforms like Facebook, Instagram, and TikTok employ multi-level categorization to aid users in navigating businesses, necessitating the extraction of knowledge from hierarchical data structures.

Hyperbolic embeddings have emerged as a potent solution to tackle this challenge. Unlike traditional Euclidean geometry, hyperbolic geometry proves more adept at modeling complex networks with hierarchical data structures. This is attributed to the constant negative curvature of hyperbolic space, enabling it to capture high-quality hierarchy information in a lower-dimensional space with minimal distortion. The utility of hyperbolic embeddings, particularly the Poincaré ball model, is evident in preserving distances between categories and hierarchies in data, as demonstrated in Equation 1 below.

Machine Learning Approaches for Privacy-Preserving with GAN:

Privacy preservation is paramount when handling sensitive user data, especially in domains susceptible to data leakage threats. Conventional privacy protection techniques, such as deidentification techniques and removal of personal identification values, have shown limitations, including potential vulnerabilities to attackers possessing background knowledge and negative impacts on data utility.

Generative Adversarial Networks (GANs) have emerged as a promising solution to address these challenges. GANs excel in synthesizing high-quality artificial samples closely resembling the distribution of original training data, rendering it challenging to discern between artificial and genuine data. Variants such as Deep Convolutional GANs (DCGANs) have proven effective in generating realistic images, while adaptations like table-GAN and CTGAN aim to synthesize relational tables and tabular data, respectively. Despite challenges, GANs have been widely embraced as a privacy-preserving mechanism across various domains subject to privacy regulations, including healthcare and finance.

In healthcare, GANs like medGAN and medBGAN generate artificial patient electronic health records, preserving statistical properties while safeguarding sensitive patient information. Likewise, the finance industry leverages GANs to generate synthetic customer datasets for portfolio analysis, fraud detection, and risk assessment, without compromising real customer data. GANs thus offer an effective solution for preserving privacy while enabling the use of realistic and statistically similar datasets for diverse purposes.

Machine Learning Approaches for Privacy-Preserving with Federated Learning:

In recent years, the drawbacks of Centralized Learning have become increasingly apparent due to the escalating complexity and volume of data. Centralized Learning places a significant strain on the network responsible for exchanging extensive data and challenges the server's capacity to process large data aggregates while ensuring data protection. These challenges have prompted the exploration of alternative systems that distribute the machine learning workload across multiple devices or mobile devices, exemplified by Federated Learning.

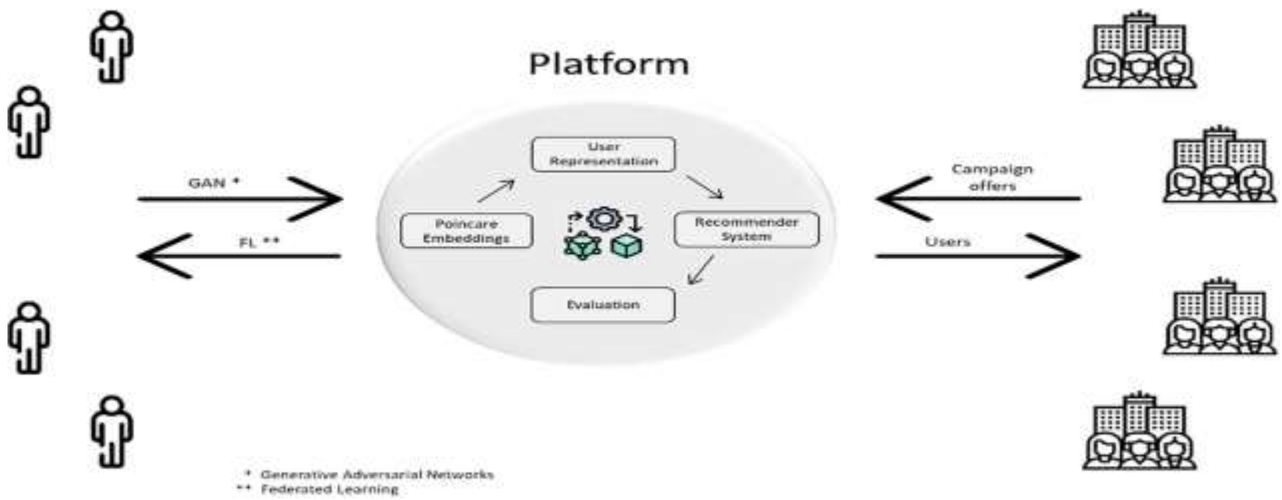
Federated Learning (FL) facilitates training data across numerous devices, coordinated by one or more central servers. This approach addresses the tension between data privacy and sharing for decentralized devices, as data remains undisclosed to a central server. Consequently, FL emerges as an apt solution for applications dealing with privacy-sensitive data. Across various industries, FL has found application in domains like banking for credit card fraud detection, image detection and representation, and healthcare for disease prediction and biomedical imaging analysis using health records. These applications underscore FL's potential to safeguard privacy while retaining the utility of shared data.

Of particular note is FL's promising role in privacy-preserving training for mobile devices, commonly referred to as on-device FL. In an era dominated by ubiquitous mobile devices generating vast amounts of personal data, ensuring user privacy and confidentiality becomes paramount. FL's decentralized framework enables the efficient utilization of data from mobile devices without compromising user privacy. For instance, on-device FL holds significant potential in enhancing privacy-preserving mobile health applications. These applications necessitate the collection and processing of sensitive health data from wearable devices like heart rate monitors and fitness trackers. Leveraging on-device FL, mobile health applications can develop personalized models for users while upholding their privacy.

Methods:

Platform Design Overview:

Drawing upon the principles and methodologies expounded in Section 2, we delineate the design of a privacy-preserving platform tailored for targeted marketing campaigns. The platform harnesses hyperbolic embeddings, Generative Adversarial Networks (GANs), and Federated Learning (FL) to ensure privacy, efficacy, and precision in targeting users based on their interests and preferences. Figure 1 offers a visual depiction of the platform design, showcasing several pivotal components. At its core, the platform focuses on modeling users' representations utilizing Poincaré Embeddings, as elucidated in Section 2.1. These embeddings adeptly capture the hierarchical nuances of users' interests and preferences, subsequently facilitating their alignment with campaigns initiated by marketers via a recommender system.



Privacy Concerns:

As delineated in Sections 2.2 and 2.3, the platform integrates Generative Adversarial Networks (GANs) and Federated Learning (FL) techniques, depicted on the left side of Figure 1. GANs are deployed to generate synthetic user data, safeguarding users' data privacy, while FL facilitates decentralized training processes on users' own devices, aiming to optimize communication efficiency and defense mechanisms to uphold data privacy.

On the right side of Figure 1, the utilization of user representations to identify the target audience aligning with the appropriate campaign is illustrated. When a marketer seeks to disseminate an offer via the platform's mobile app, they must delineate the target audience by specifying user characteristics such as interests, demographics, or online behavior. Specifically, for each campaign conducted on the platform, a user list is curated, enabling marketers to target users meeting campaign criteria and those more predisposed to accepting the offer.

Nevertheless, neither marketers nor the platform possess knowledge regarding which users may receive a specific campaign offer. Only the learned latent user embedding generated from synthetic user data is utilized to curate the user list aligning with the campaign offer. This process is facilitated through a series of deep learning techniques elaborated upon in the subsequent sections.

User Modeling Using Poincaré Embeddings:

We employ Poincaré embeddings to proficiently model users within a hyperbolic space, facilitating the efficient representation of hierarchical data derived from user characteristics and interests. The Poincaré model yields a collection of vectors or coordinates for each category on the Poincaré ball, furnishing valuable insights into the positioning of each interest or characteristic within the space.

To elaborate, each user is mapped into the hyperbolic space by computing an average of their interests and characteristics, generating a unique n-dimensional vector as a Poincaré embedding. Subsequently, these Poincaré embeddings of users are utilized to form target user groups with similar characteristics using PoincareKMeans. Notably, the implementation of this clustering method diverges from the original version of K-Means utilizing Euclidean distance as a similarity measure. The unique context of the hyperbolic space necessitates the consideration of non-linear spaces to accurately gauge comparable features without diminishing the original graph distances. Hence, users are partitioned into K groups based on hyperbolic distances between their embeddings on the Poincaré ball, enabling the capture of hierarchical relationships by clustering users belonging to the same category in proximity to each other. Refer to Section 1 in the Appendix for further details on the training process and selection of hyperparameters for the Poincaré model.

With the establishment of target groups, we formulate avatars representing the shared characteristics of users within each group. This process offers several advantages: firstly, aggregating user information to a higher level mitigates the risk of adversaries re-identifying individual users based on their interests and characteristics. Secondly, despite anonymization, the use of avatars ensures the data remains valuable for marketing purposes, enabling marketers to identify and target specific audience segments effectively. Thirdly, hierarchical user characteristics naturally align with the Poincaré embedding representation, rendering the creation of target groups more suitable than those generated from similarity metrics in Euclidean space.

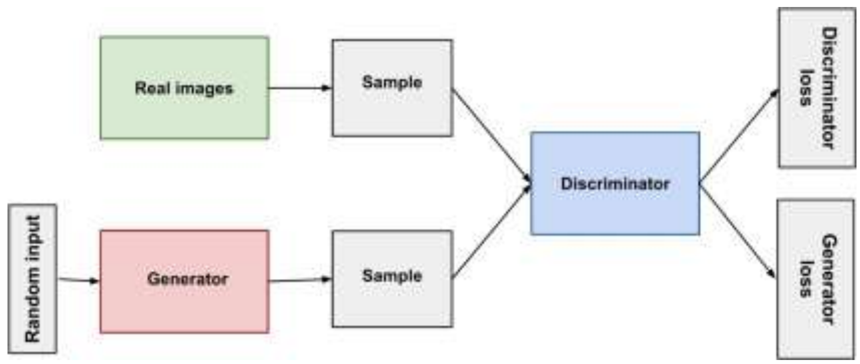
User Representation Synthesized Using GAN:

To synthesize user representations while safeguarding their privacy, we implement a customized GAN architecture. As depicted in Figure 2, the fundamental concept of GAN remains consistent: training two neural networks simultaneously, a Generator (G) and a Discriminator (D), in a manner where G generates synthetic data resembling the original data distribution and D discerns fake data from authentic data to optimize the min-max loss function, as illustrated in Equation 2:

$$\min_{G} \max_{D} V(G, D) = \mathbb{E}_{x \sim p_{data}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_{z}(z)} [\log (1 - D(G(z)))]$$

Where $p_{data}(x)$ denotes the original data distribution, $p_{z}(z)$ represents the simple noise distribution (typically a normal distribution), $x \sim p_{data}(x)$ denotes the expected value across all instances of original data, and $z \sim p_{z}(z)$ denotes the expected value across all random inputs to the Generator. Essentially, the objective of the generator G is to minimize the value function $V(G,D)$, while the discriminator D endeavors to maximize it.

Specifically, before applying the embedding methods, GAN generates synthetic data indistinguishable from the original data to protect user privacy. The training process commences by feeding random noise generated by G and actual data from the original dataset into D. Subsequently, D classifies synthetic data as real or fake. In case of misclassification, D incurs penalties through the discriminator loss. Finally, hyperparameters are updated via backpropagation.



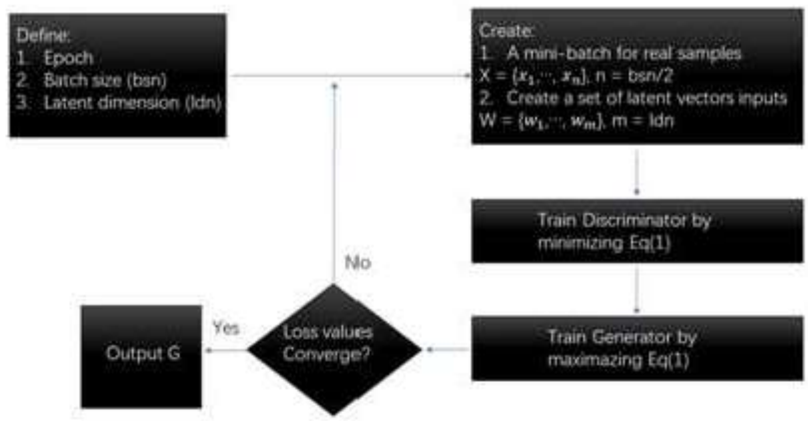


Fig. 3 Step-by-step illustration of GAN training algorithm

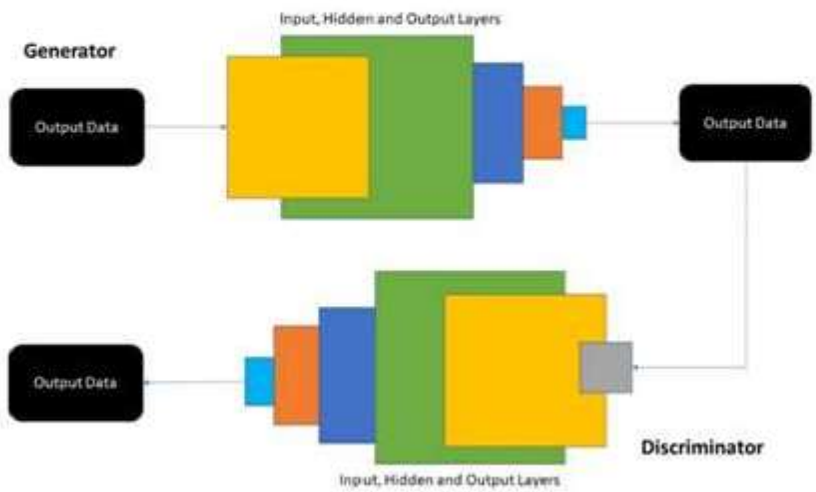


Fig. 4 Proposed GAN architecture for generator and discriminator

This iterative process continues until the discriminator loss converges, thereby minimizing the probability of errors. The efficacy of the generator (G) is intricately linked to the performance of the discriminator (D). G cannot be trained independently; an evaluation from D is imperative to update G's hyperparameters. G incurs penalties if D classifies its output as artificial, thus prompting adjustments through backpropagation, starting from D's output and propagating back into G. Figure 3 outlines the GAN training algorithm.

Additionally, we have tailored the GAN architecture to suit our specific use case. Figure 4 illustrates the custom GAN architecture, incorporating the following modifications. Firstly, our customized generator comprises one input layer, three fully connected hidden layers, and one output layer. LeakyReLU activation functions are applied to the hidden layers, while the output layer utilizes the tanh function to capture potential correlations between variables in user data. Secondly, the discriminator in our architecture features one input layer, four fully connected hidden layers, and one output layer. LeakyReLU activation functions are utilized for all hidden layers, the Sigmoid function for the output layer, and Dropout to mitigate overfitting. Furthermore, we employ the Adam version of stochastic gradient descent with a learning rate of 0.0002 and a momentum of 0.5. Refer to Section 2 in the Appendix for detailed insights into the training process of the custom GAN architecture.

The implementation of GAN plays a pivotal role in this study, facilitating the generation of synthetic data from original user information while upholding user privacy. This measure prevents user identification and ensures the confidentiality of their information. In conjunction with user modeling using hyperbolic embeddings, the synthetic data generated by GAN preserves the underlying structure and patterns of the original user data, bolstering privacy protection while maintaining data utility for digital marketing endeavors.

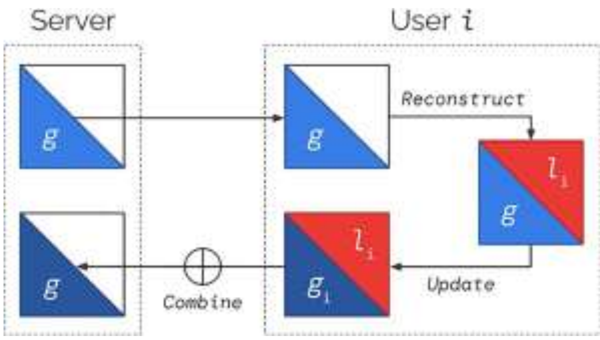
Distributed Training Using Federated Learning:

In this study, the distributed training of the model is achieved through a specific approach to Federated Learning (FL) known as Partially Local Federated Learning. This approach is employed to handle scenarios where the model consists of user-specific parameters, as observed in matrix factorization tasks. Sending updates of user embeddings to the server during the training of a global federated model is undesirable in such cases, as it may expose potentially sensitive individual preferences. To mitigate this risk, the model is divided into global and local parameters.

Specifically, the matrix containing users' preferences is factorized into a user matrix and an item matrix, generating a k-dimensional user-specific embedding for each user. This methodology ensures that certain parameters are not transmitted to the server. However, it necessitates clients to maintain their user embeddings across multiple rounds, which may be impractical. In large-scale cross-device settings, users are unlikely to be sampled more than once during the training process, leading to performance degradation.

To address this challenge, a federated reconstruction framework is employed in this study. This framework eliminates the requirement for users to retain their local parameters across rounds by reconstructing them as needed. A reconstruction algorithm is utilized to restore the local parameters. The process is depicted in Figure 5: in each round, the server retains and disseminates the item matrix (global parameters) to the selected users. Subsequently, each user employs one or more stages of Stochastic Gradient Descent (SGD) to freeze the item matrix and train their user embedding (local variables). Following this, each user freezes their user embedding and employs one or more steps of SGD to train the item matrix. Finally, updates to the item matrix are aggregated across users, and the server's copy of the item matrix is updated for utilization in the subsequent round.

It is noteworthy that the training process in this study diverges from the conventional Federated Learning process, which typically employs federated averaging for aggregation. Instead, a reconstruction optimizer is utilized to reconstruct the parameters that remain local, such as user embeddings. Both the server and user optimizer utilize the same SGD optimizer, albeit with differing learning rates. Refer to Section 3 in the Appendix for a detailed illustration of the steps involved in the FL training framework.



Matching Users with Marketing Campaigns Using HyperML:

The final phase of the process entails harnessing the synthesized user embeddings and integrating them into a recommender system model to accurately pinpoint the list of most appropriate users for a given campaign offer. Since user characteristics and interests are encapsulated by Poincaré embeddings, we adopt a specialized approach known as Hyperbolic Metric Learning (HyperML), tailored for recommender systems operating in hyperbolic space, such as the manifold Poincaré Ball, which computes similarities. Particularly, the training procedure with HyperML occurs on the Poincaré Ball manifold, facilitating distance computations in hyperbolic space. This enables the learning of Poincaré embeddings of insights, master categories, and users, with the ultimate aim of predicting new users who may exhibit interest in a specific insight or category.

The selection of appropriate users is pivotal for campaign success. Hence, it is imperative to identify a user base that not only ensures marketers target their offers towards those most inclined to accept the campaign but also ensures that clients receive offers aligning with their needs and desires. Consequently, the primary objective is to recommend users for a particular campaign based on their interests and personal details, utilizing this uniquely tailored approach.

The advantages of employing this specialized methodology for matching users with marketing campaigns are manifold. By leveraging HyperML and the Poincaré Ball manifold, the recommender system can more effectively identify suitable users for specific campaigns from their Poincaré embeddings, leading to enhanced targeting and more efficient marketing strategies. Consequently, this can result in increased conversion rates and improved return on investment (ROI) for marketers. Additionally, this approach facilitates enhanced personalization of marketing campaigns, ensuring users receive offers that are pertinent and tailored to their preferences and needs. This enhances user satisfaction and cultivates loyalty to both the brand and the platform.

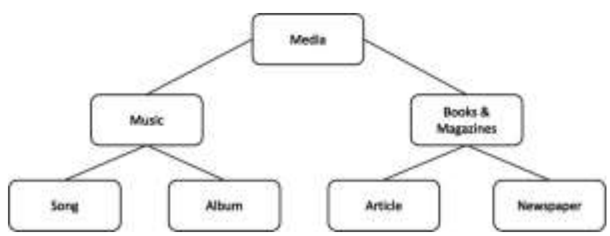
Results:

Data Description:

The dataset utilized in this study comprises synthetic user data modeled on profiles resembling those of 10,000 real-world Facebook users. This synthetic dataset is crafted to retain the structure and distribution of the original data while safeguarding user privacy. It delineates users' interests based on their likes on Facebook pages. Importantly, it should be emphasized that no actual Facebook profile data was accessed or utilized directly in this research; the data employed is entirely synthetic. Access to the synthetic dataset can be arranged through the corresponding author, subject to approval by the collaborating data monetization platform. Additionally, supplementary materials, including code and models, are available upon reasonable request and with appropriate permissions from the authors.

Each Facebook page in the synthetic dataset is associated with an intrinsic category, organized hierarchically with a depth of three levels. Consequently, user interest hierarchies play a significant role in user modeling. As illustrated in Figure 6, for instance, the category "Media" encompasses child categories such as "Music" and "Books & Magazines," with "Music" further subdivided into "Song" and "Album." Thus, user interests are delineated as a hierarchical dataset comprising various levels of categories. In total, there are 1563 categories.

To effectively match users with campaigns, it is imperative to identify users with specific interests and duly consider their preferences.



- For categories where the user possesses a genuine interest, the value assigned is 1.

- For categories where the user lacks interest (explicitly), the value is -1.

Figure 6: Example of Hierarchical Facebook Category

For categories not assigned values of 1 or -1, they are regarded as unknown interests (value = 0), signifying uncertainty regarding whether the user has an interest or not.

Results from Poincaré Embeddings:

In an ideal scenario, users with similar interests within the same parent category should be represented closely. However, achieving this using traditional one-hot encoding in Euclidean space poses challenges. To tackle this

hurdle, we adopt a Poincaré embedding model to learn user representations. This involves transforming user interests into hyperbolic space and calculating similarity using non-Euclidean distance metrics.

Poincaré embeddings are obtained for all users, resulting in a set of n-dimensional vectors or coordinates for each category in hyperbolic space, where n ranges between 25 to 200. To assess the quality of Poincaré embeddings, we employ the same methodologies as described by [37]: Reconstruction error relative to the embedding dimension, which evaluates the representation capacity through hierarchy reconstruction, and Link prediction, accomplished by dividing the data into train, validation, and test sets to gauge generalization performance. We compare these tasks using Poincaré distance against traditional Euclidean distance, represented by d(x, y)=‖ x − y ‖2.

These evaluations are quantified by two metrics:

- Mean Rank: the average rank of all observations within each sample.

- MAP (Mean Average Precision): a metric capturing the preservation quality of each vertex's neighborhoods.

The evaluation outcomes of Poincaré embeddings are depicted in Table 1. Analysis of the reconstruction task reveals consistently superior representation quality with Poincaré embeddings compared to Euclidean embeddings. Furthermore, quality substantially improves with increasing dimensionality, although the enhancement becomes marginal for 200 dimensions. Similarly, for the Link Prediction task, Poincaré embeddings exhibit superior representation quality.

Table 1 Evaluation of embedding quality obtained from reconstruction and link prediction tasks in hyperbolic space and Euclidean space

		Reconstruction				Link prediction			
		25D	50D	100D	200D	25D	50D	100D	200D
Euclidean	Mean Rank↓	4.56	4.28	4.19	4.12	3.95	3.85	3.73	3.51
	MAP↑	0.439	0.445	0.448	0.451	0.392	0.408	0.411	0.453
Poincaré	Mean Rank↓	2.59	2.55	**2.50**	2.53	3.19	2.98	3.07	**2.76**
	MAP↑	0.534	0.534	0.536	**0.54**	0.413	0.418	0.416	**0.538**

The bold indicates the optimal performance obtained from the model tasks and parameters. For Mean Rank, the lower the better. For MAP, the larger the better

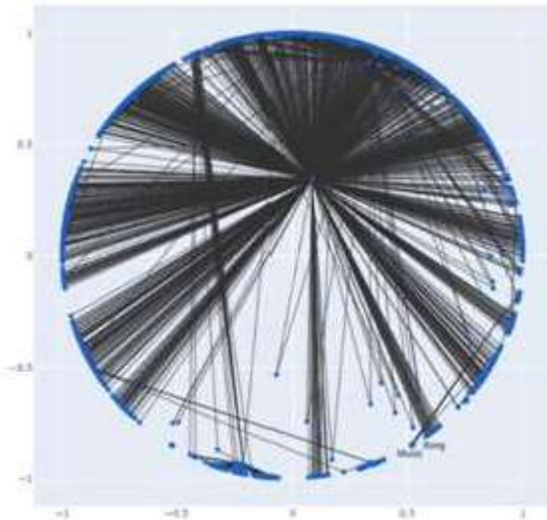


Fig. 7 Representation of the Poincaré embeddings in 2D

Additionally, there was a notable decrease in Mean Rank from 3.2 to 2.76, accompanied by an increase in the MAP value from 0.41 to 0.538, signifying a substantial enhancement with increased dimensionality. Following fine-tuning of our model, the optimal model exhibited the following specifications: 20 negative samples, 0 burn-in initialization, no regularization, and 200 epochs in a 200-dimensional space. These findings underscore the superior representation of the dataset in hyperbolic space, as evidenced by the distances between interests and the preservation of the hierarchy across the three levels of categories.

Figure 7 showcases the obtained embeddings of categories on the Poincaré Model, depicted as blue dots. The straight black lines delineate the relationships between the hierarchy levels of the categories. Notably, the figure illustrates that more similar categories are positioned closer together, exemplified by the proximity of "Music" and "Song."

Moreover, to further illustrate the representation quality of the Poincaré embeddings, we segmented all users into multiple groups with similar characteristics and interests using the PoincaréKMeans model. As PoincaréKMeans operates as an unsupervised learning algorithm, and given that evaluation metrics for clustering analysis are still evolving, the determination of the number of clusters primarily relies on domain knowledge and intuition. Figure 8 presents the depiction of user partitions into 6 target groups using hyperbolic embeddings. Appendix Section A elaborates on the group characteristics derived from the clustering analysis.

Such representations enhance our understanding of the characteristics of each group, potentially increasing the accuracy of representing categories for a user, as users within the same group are more likely to share similar preferences.

Discussions and conclusions

This study introduced an innovative approach to privacy-preserving user modeling for digital marketing campaigns through the utilization of deep learning techniques on a data monetization platform. The core objective was to devise a solution that facilitates marketers in accurately identifying suitable target audiences for their campaigns while empowering users to retain control over their personal data and ensuring adherence to data privacy regulations. The proposed methodology integrated representation learning in hyperbolic space, Generative Adversarial Networks (GAN), and Federated Learning (FL) to capture latent user interests, generate synthetic representations, and conduct decentralized training, all without divulging user data to marketers. By implementing a targeting strategy based on recommendation systems, the acquired user interests were leveraged to pinpoint the optimal target audience for digital marketing campaigns.

This study's contributions encompass the development of a privacy-preserving user modeling mechanism that strikes a balance between personalization and data privacy, the practical application of deep learning techniques for user modeling in digital marketing, and the exploration of new business models fostering value co-creation between consumers and marketers on data monetization platforms.

However, the research faces several limitations. Firstly, the Poincaré Ball model may not capture the relations between categories and users accurately, potentially leading to inaccurate user representations. Additionally, users' interests may evolve over time, impacting the model's effectiveness. Secondly, while GAN techniques are valuable for generating synthetic user profiles, they may struggle with generating categorical data and addressing dataset imbalances. Future investigations could explore alternative GAN variations or architectures for enhanced results. Thirdly, while FL techniques offer advantages such as reducing communication constraints and bolstering privacy protection, they are not foolproof, and limitations exist in terms of the level of protection they afford. Future research avenues could explore applications of differential privacy or secure aggregation to further fortify users' information against attacks. Lastly, the performance evaluation of the proposed approach was conducted using a restricted set of metrics, potentially overlooking pertinent aspects of user modeling and privacy preservation. Moreover, the study was conducted in collaboration with a single data monetization platform, potentially limiting the generalizability of the findings to other platforms or contexts.

This study marks a significant stride towards a more transparent, ethical, and sustainable digital marketing ecosystem benefiting both consumers and marketers. Future research trajectories could involve exploring the adaptability of the proposed approach to different online platforms and E-commerce marketing contexts, refining the proposed architectures with advancements in deep learning techniques, and broadening the evaluation framework to encompass a wider array of metrics such as user satisfaction, user privacy risk, and potential impacts on marketing outcomes. By addressing these limitations and building upon the contributions of this study, future research endeavors can deepen our understanding of privacy-preserving user modeling and its implications for the digital marketing realm.

References List

[1]. Shuford, J. (2023). Contribution of Artificial Intelligence in Improving Accessibility for Individuals with Disabilities. Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online), 2(2), 421-433.

DOI: https://doi.org/10.60087/jklst.vol2.n2.p433

[2]. Chentha, A. K., Sreeja, T. M., Hanno, R., Purushotham, S. M. A., &Gandrapu, B. B. (2013). A Review of the Association between Obesity and Depression. Int J Biol Med Res, 4(3), 3520-3522.

[3]. Gadde, S. S., &Kalli, V. D. R. (2020). Descriptive analysis of machine learning and its application in healthcare. *Int J Comp Sci Trends Technol*, *8*(2), 189-196.

[4]. Atacho, C. N. P. (2023). A Community-Based Approach to Flood Vulnerability Assessment: The Case of El Cardón Sector. *Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online)*, *2*(2), 434-482. DOI:https://doi.org/10.60087/jklst.vol2.n2.p482

[5]. jimmy, fnu. (2023). Understanding Ransomware Attacks: Trends and Prevention Strategies. Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online), 2(1), 180-210. https://doi.org/10.60087/jklst.vol2.n1.p214

[6]. Bayani, S. V., Prakash, S., &Malaiyappan, J. N. A. (2023). Unifying Assurance A Framework for Ensuring Cloud Compliance in AIML Deployment. *Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online)*, *2*(3), 457-472.DOI: https://doi.org/10.60087/jklst.vol2.n3.p472

[7]. Bayani, S. V., Prakash, S., &Shanmugam, L. (2023). Data Guardianship: Safeguarding Compliance in AI/ML Cloud Ecosystems. *Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online)*, *2*(3), 436-456.

DOI: https://doi.org/10.60087/jklst.vol2.n3.p456

[8]. Karamthulla, M. J., Malaiyappan, J. N. A., & Prakash, S. (2023). AI-powered Self-healing Systems for Fault Tolerant Platform Engineering: Case Studies and Challenges. *Journal of*

Knowledge Learning and Science Technology ISSN: 2959-6386 (online), *2*(2), 327-338. DOI: https://doi.org/10.60087/jklst.vol2.n2.p338

[9]. Prakash, S., Venkatasubbu, S., &Konidena, B. K. (2023). Unlocking Insights: AI/ML Applications in Regulatory Reporting for US Banks. *Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online)*, *1*(1), 177-184.DOI: https://doi.org/10.60087/jklst.vol1.n1.p184

[10]. Prakash, S., Venkatasubbu, S., &Konidena, B. K. (2023). From Burden to Advantage: Leveraging AI/ML for Regulatory Reporting in US Banking. *Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online)*, *1*(1), 167-176. DOI: https://doi.org/10.60087/jklst.vol1.n1.p176

[11]. Prakash, S., Venkatasubbu, S., &Konidena, B. K. (2022). Streamlining Regulatory Reporting in US Banking: A Deep Dive into AI/ML Solutions. *Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online)*, *1*(1), 148-166.DOI: https://doi.org/10.60087/jklst.vol1.n1.p166

[12]. Tomar, M., &Jeyaraman, J. (2023). Reference Data Management: A Cornerstone of Financial Data Integrity. Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online), 2(1), 137-144.DOI: https://doi.org/10.60087/jklst.vol2.n1.p144

[13]. Tomar, M., &Periyasamy, V. (2023). The Role of Reference Data in Financial Data Analysis: Challenges and Opportunities. Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online), 1(1), 90-99.

DOI: https://doi.org/10.60087/jklst.vol1.n1.p99

[14]. Tomar, M., &Periyasamy, V. (2023). Leveraging Advanced Analytics for Reference Data Analysis in Finance. *Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online)*, *2*(1), 128-136.

DOI: https://doi.org/10.60087/jklst.vol2.n1.p136

[15]. Sharma, K. K., Tomar, M., &Tadimarri, A. (2023). Unlocking Sales Potential: How AI Revolutionizes Marketing Strategies. *Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online)*, *2*(2), 231-250.

DOI: https://doi.org/10.60087/jklst.vol2.n2.p250

[16]. Sharma, K. K., Tomar, M., &Tadimarri, A. (2023). Optimizing Sales Funnel Efficiency: Deep Learning Techniques for Lead Scoring. *Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online)*, *2*(2), 261-274.

DOI: https://doi.org/10.60087/jklst.vol2.n2.p274

[17]. Shanmugam, L., Tillu, R., &Tomar, M. (2023). Federated Learning Architecture: Design, Implementation, and Challenges in Distributed AI Systems. *Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online)*, *2*(2), 371-384.

DOI: https://doi.org/10.60087/jklst.vol2.n2.p384

[18]. Sharma, K. K., Tomar, M., &Tadimarri, A. (2023). AI-driven Marketing: Transforming Sales Processes for Success in the Digital Age. *Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online)*, *2*(2), 250-260.

DOI: https://doi.org/10.60087/jklst.vol2.n2.p260

[19]. Gadde, S. S., &Kalli, V. D. (2021). The Resemblance of Library and Information Science with Medical Science. International Journal for Research in Applied Science & Engineering Technology, 11(9), 323-327.

[20]. Gadde, S. S., &Kalli, V. D. R. (2020). Technology Engineering for Medical Devices-A Lean Manufacturing Plant Viewpoint. Technology, 9(4).

[21]. Gadde, S. S., &Kalli, V. D. R. (2020). Medical Device Qualification Use. International Journal of Advanced Research in Computer and Communication Engineering, 9(4), 50-55.

[22]. Gadde, S. S., &Kalli, V. D. R. (2020). Artificial Intelligence To Detect Heart Rate Variability. *International Journal of Engineering Trends and Applications*, *7*(3), 6-10.

[23]. Chentha, A. K., Sreeja, T. M., Hanno, R., Purushotham, S. M. A., &Gandrapu, B. B. (2013). A Review of the Association between Obesity and Depression. *Int J Biol Med Res*, *4*(3), 3520-3522.

[24]. Tao, Y. (2022). Algorithm-architecture co-design for domain-specific accelerators in communication and artificial intelligence (Doctoral dissertation).

https://deepblue.lib.umich.edu/handle/2027.42/172593

[25]. Tao, Y., Cho, S. G., & Zhang, Z. (2020). A configurable successive-cancellation list polar decoder using split-tree architecture. IEEE Journal of Solid-State Circuits, 56(2), 612-623.

DOI: https://doi.org/10.1109/JSSC.2020.3005763

[26]. Tao, Y., & Choi, C. (2022, May). High-Throughput Split-Tree Architecture for Nonbinary SCL Polar Decoder. In 2022 IEEE International Symposium on Circuits and Systems (ISCAS) (pp. 2057-2061). IEEE.

DOI: https://doi.org/10.1109/ISCAS48785.2022.9937445

[27]. Tao, Y. (2022). Algorithm-architecture co-design for domain-specific accelerators in communication and artificial intelligence (Doctoral dissertation).

https://deepblue.lib.umich.edu/handle/2027.42/172593

[28]. Mahalingam, H., VelupillaiMeikandan, P., Thenmozhi, K., Moria, K. M., Lakshmi, C., Chidambaram, N., &Amirtharajan, R. (2023). Neural attractor-based adaptive key generator with DNA-coded security and privacy framework for multimedia data in cloud environments. Mathematics, 11(8), 1769.

https://doi.org/10.3390/math11081769

[29]. Padmapriya, V. M., Thenmozhi, K., Praveenkumar, P., &Amirtharajan, R. (2020). ECC joins first time with SC-FDMA for Mission "security". Multimedia Tools and Applications, 79(25), 17945-17967.

DOI https://doi.org/10.1007/s11042-020-08610-5

[30]. Padmapriya, V. M. (2018). Image transmission in 4g lte using dwt based sc-fdma system. Biomedical & Pharmacology Journal, 11(3), 1633.

DOI :https://dx.doi.org/10.13005/bpj/1531

[31]. Padmapriya, V. M., Priyanka, M., Shruthy, K. S., Shanmukh, S., Thenmozhi, K., &Amirtharajan, R. (2019, March). Chaos aided audio secure communication over SC-FDMA system. In 2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN) (pp. 1-5). IEEE.

https://doi.org/10.1109/ViTECoN.2019.8899413

[31]. Padmapriya, V. M., Thenmozhi, K., Praveenkumar, P., &Amirtharajan, R. (2022). Misconstrued voice on SC-FDMA for secured comprehension-a cooperative influence of DWT and ECC. Multimedia Tools and Applications, 81(5), 7201-7217.

DOI https://doi.org/10.1007/s11042-022-11996-z

[32]. Padmapriya, V. M., Sowmya, B., Sumanjali, M., &Jayapalan, A. (2019, March). Chaotic Encryption based secure Transmission. In 2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN) (pp. 1-5). IEEE.

DOI https://doi.org/10.1109/ViTECoN.2019.8899588

[33]. Sowmya, B., Padmapriya, V. M., Sivaraman, R., Rengarajan, A., Rajagopalan, S., &Upadhyay, H. N. (2021). Design and Implementation of Chao-Cryptic Architecture on FPGA for Secure Audio Communication. In Emerging Technologies in Data Mining and Information Security: Proceedings of IEMIS 2020, Volume 3 (pp. 135-144). Springer Singapore

https://link.springer.com/chapter/10.1007/978-981-15-9774-9_13

[34]. Padmapriya, V. M., Thenmozhi, K., Avila, J., Amirtharajan, R., &Praveenkumar, P. (2020). Real Time Authenticated Spectrum Access and Encrypted Image Transmission via Cloud Enabled Fusion centre. Wireless Personal Communications, 115, 2127-2148.

DOI https://doi.org/10.1007/s11277-020-07674-8

[35].Thakur, A., & Thakur, G. K. (2024). Developing GANs for Synthetic Medical Imaging Data: Enhancing Training and Research. Int. J. Adv. Multidiscip. Res, 11(1), 70-82.

DOI: http://dx.doi.org/10.22192/ijamr.2024.11.01.009

[36]. Shuford, J. (2023). Contribution of Artificial Intelligence in Improving Accessibility for Individuals with Disabilities. Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online), 2(2), 421-433.DOI: https://doi.org/10.60087/jklst.vol2.n2.p433

[37]. Schwartz, E. A., Bravo, J. P., Ahsan, M., Macias, L. A., McCafferty, C. L., Dangerfield, T. L., ...& Taylor, D. W. (2024). RNA targeting and cleavage by the type III-Dv CRISPR effector complex. Nature Communications, 15(1), 3324.

https://www.nature.com/articles/s41467-024-47506-y#Abs1

[38]. Saha, A., Ahsan, M., Arantes, P. R., Schmitz, M., Chanez, C., Jinek, M., & Palermo, G. (2024). An alpha-helical lid guides the target DNA toward catalysis in CRISPR-Cas12a. Nature Communications, 15(1), 1473.https://www.nature.com/articles/s41467-024-45762-6

[39]. Nierzwicki, Ł., Ahsan, M., & Palermo, G. (2023). The electronic structure of genome editors from the first principles. Electronic Structure, 5(1), 014003.DOI https://doi.org/10.1088/2516-1075/acb410

[40]. Bali, S. D., Ahsan, M., &Revanasiddappa, P. D. (2023). Structural Insights into the Antiparallel G-Quadruplex in the Presence of K+ and Mg2+ Ions. The Journal of Physical Chemistry B, 127(7), 1499-1512.https://doi.org/10.1021/acs.jpcb.2c05128

[41]. Ahsan, M., Pindi, C., &Senapati, S. (2022). Mechanism of darunavir binding to monomeric HIV-1 protease: A step forward in the rational design of dimerization inhibitors. Physical Chemistry Chemical Physics, 24(11), 7107-7120.https://doi.org/10.1039/D2CP00024E

[42]. Ahsan, M., Pindi, C., &Senapati, S. (2021). Hydrogen bonding catalysis by water in epoxide ring opening reaction. Journal of Molecular Graphics and Modelling, 105, 107894.https://doi.org/10.1016/j.jmgm.2021.107894

[43]. Ahsan, M., Pindi, C., &Senapati, S. (2020). Electrostatics plays a crucial role in HIV-1 protease substrate binding, drugs fail to take advantage. Biochemistry, 59(36), 3316-3331.

https://doi.org/10.1021/acs.biochem.0c00341

[44]. Pindi, C., Chirasani, V. R., Rahman, M. H., Ahsan, M., Revanasiddappa, P. D., &Senapati, S. (2020). Molecular basis of differential stability and temperature sensitivity of ZIKA versus dengue virus protein shells. Scientific Reports, 10(1), 8411.https://doi.org/10.1038/s41598-020-65288-3

[45]. Ahsan, M., &Senapati, S. (2019). Water plays a cocatalytic role in epoxide ring opening reaction in aspartate proteases: a QM/MM study. *The Journal of Physical Chemistry B*, *123*(38), 7955-7964.

https://doi.org/10.1021/acs.jpcb.9b04575

[46]. Dixit, S. M., Ahsan, M., &Senapati, S. (2019). Steering the lipid transfer to unravel the mechanism of cholesteryl ester transfer protein inhibition. Biochemistry, 58(36), 3789-3801.

https://doi.org/10.1021/acs.biochem.9b00301

[46]. Hasan, M. R., Gazi, M. S., & Gurung, N. (2024). Explainable AI in Credit Card Fraud Detection: Interpretable Models and Transparent Decision-making for Enhanced Trust and Compliance in the USA. *Journal of Computer Science and Technology Studies*, *6*(2), 01-12.

[47]. Gazi, M. S., Hasan, M. R., Gurung, N., & Mitra, A. (2024). Ethical Considerations in AI-driven Dynamic Pricing in the USA: Balancing Profit Maximization with Consumer Fairness and Transparency. *Journal of Economics, Finance and Accounting Studies*, *6*(2), 100-111.

[48]. Sarkar, M., Puja, A. R., & Chowdhury, F. R. (2024). Optimizing Marketing Strategies with RFM Method and K-Means Clustering-Based AI Customer Segmentation Analysis. *Journal of Business and Management Studies*, *6*(2), 54-60.

[49]. Jones, K., Spaeth, J., Rykowski, A., Manjunath, N., Vudutala, S. K., Malladi, R., & Mishra, A. (2020). U.S. Patent No. 10,659,295. Washington, DC: U.S. Patent and Trademark Office.

[50]. Malladi, R., Bukkapattanam, A., Wigley, C., Aggarwal, N., & Vudutala, S. K. (2021). U.S. Patent No. 11,087,020. Washington, DC: U.S. Patent and Trademark Office.

[51]. Jones, K., Pitchaimani, S., Viswanathan, S., Shah, M., Malladi, R., Allidina, A., ... & Brannon, J. B. (2023). U.S. Patent No. 11,797,528. Washington, DC: U.S. Patent and Trademark Office.