PFI-TT: Automated Recommender Systems for small businesses digital transformation to E-commerce.

Proposal submission on FastLane.nsf.gov [DEADLINE: July 14, 2021.]

- ☑ Project Summary [One (1) page max].
- ☑ FastLane documentation
  - ☐ Collaborators and other affiliations
  - ☐ Current and pending support
  - ☐ Bio sketch
  - ☐ Budget
  - ☐ Data management
  - ☐ Equipment and facilities
  - ☐ Postdoctoral management plan

Project Description. -Fifteen (15) pages max-

- ☑ Executive Summary (no more than one page)
- ☑ From NSF Basic Research to Addressing a Market Opportunity (suggested length: 4 -5 pages)
- ☑ Technical Challenges and Applied Research Plan (suggested length: 5-7 pages)
- ☑ Achieving Societal Impact through the Realization of Commercial Potential (2-4 pages)
- ☑ Project Team (suggested length: 1-2 pages)
- ☑ Partnerships (suggested length: 1-2 pages)
- ☑ Training Future Leaders in Innovation and Entrepreneurship (suggested length: 1-2 pages)
- ☑ Broadening Participation (suggested length: up to 1 page)

Others

- ☑ Letters of support (max. 3)

Resubmission Change Description

Previously declined proposal information: The reviewers made the following comments to the original PFI-TT proposal #2122812, *PFI-TT: Democratizing Recommender Systems for E-commerce with Automated Graph Neural Networks*.

C1: It is unclear what the commercial product/service will be as the output.

Response: This PFI-TT proposal aims to develop a recommendation system engine to help small and medium businesses (SMB) personalize the product offer for every individual customer. Utilize real-time personalization for a better user experience using collaborative data from multiple SMBs. The prototype we aim to make such technology affordable and easy to deploy by SMB managers.

C2: As an open-source software package, the team has not established how the IP will be protected or what path towards commercialization is for the outputs of this project.

Response: After receiving IP advice from the entrepreneurial office at Rice University, we have concluded that to protect our IP, our team will protect the outputs generated in this project first via trade secret, a standard in the industry. Such practices include Non-Disclosure Agreements and non-compete agreements, to name a few. In the same way, we plan to publish the advanced functions, such as automated machine learning, interpretable machine learning, and state-of-the-art recommendation modules, under the Berkeley Source Distribution (BSD) license, which forbidden users to develop their product based on these modules privately. Finally, our team has already approached an IP firm to explore patenting different systems that our team will implement in this PFI project that can later be licensed through Rice University.

C3: Seems unlikely that small-medium retailers would have even the basic ML expertise needed to utilize the tools developed.

Response: Our partnership with Amazon and Samsung was key for our NSF lineage results to understand the challenges of recommendation systems from a large corporation perspective. With the technical understanding of our previous fundamental research, in this PFI proposal, we have screened and selected a specific type of small-medium businesses with the data maturity and back-end team to help us develop a recommendation system engine tailored to a larger group of SMBs. For this resubmission, we have reached out to our partners Biarte and FreeFuse, which have the required technical support to utilize the prototype resulting in this PFI project.

C4: The broader impact of improving sales is not a very compelling argument. Additionally, there is minimal evidence to support that this argument is true or a quantification of how much that improvement is likely to be.

Response: We have updated our broader impacts to reflect the following societal benefits better:

Enhance the US competitiveness: as mentioned in the letter of support from our SMB partners Biarte and FreeFuse, both identify recommendation systems as the missing elements to fuel a powerful engine to drive their sales and improve engagement. A proper recommendation system has the chance to substantially drive their revenue as our team has already been seen with our lineage results with Amazon and Samsung. Furthermore, the strong support from companies such as Amazon and Samsung has been instrumental in our search to commercialize our lineage findings and ML technologies. Such industrial partnerships will be key when translating our technologies to our SMB partners Biarte and FreeFuse to democratize machine learning in the

US economy effectively. Biarte, one of our partners, has particularly expressed the following: 'Democratizing AI in search of supporting small businesses will boost small producers' social and economic development in the neighboring communities of Austin.'

Effectively enhance partnerships between academia and industry in the United States: recommendation systems are essential online channels: they shape the media we consume and the products we seek. By partnering with Biarte, FreeFuse, and the Rice University's entrepreneurial office, we aim to understand better the technological barriers that these SMB face and lower the complexity barrier for more small businesses to effectively transfer our NSF-funded fundamental recommendation engine research into the market, which is key to drive innovation and economic prosperity in the country.

C5: The target market of small to medium sized retailers is different than the partnerships the team has with very large retailers. The challenges and concerns of the target market may not be addressed well by using only these resources.

Response: More than 1.9 million US businesses are using Amazon as a marketplace, and as of 2019, 58 percent of Amazon's total sales were by SMBs. Recommendation systems are very complex and still costly technology that we want to develop with the help of partners such as Amazon and Samsung to solve particular problems of independent SMB owners. We still consider that the partnership with Amazon remains relevant, mainly because this technology has been implemented and matured by large firms with the workforce means and data to implement and capitalize on such advancements in recommendation systems. However, for this resubmission, to further enhance the impact of this PFI project, we have expanded our partnership to the other two partners. Biarte and FreeFuse are SMBs with operations online and have seen the importance of a strong recommendation engine amid the pandemic-accelerated changes in consumer purchase habits that demand better online opportunities.

On the other hand, another profile of companies in this project that may benefit directly is an early-stage startup. Although such startups may seem different from Biarte, they share similar characteristics and need when implementing recommendation engines to promote third-party products on their websites. The robust recommendation engine we propose aims to give startups like FreeFuse and Biarte a way to work together and fully use the capabilities of recommendation systems by sharing their data assets and compete in the big-data-driven e-commerce market.

C6: Minimal assessment of the leadership and entrepreneurship educational goals.

Response: Through our partnership with our on-campus accelerator and incubator, we will better address our leadership and entrepreneurial goals. Particularly by knowing what courses and workshops are available for our students and postdoctoral researcher. Also, after the submission of the original PFI proposal, our team successfully participated in the worlds most important student-led startup competition, the Rice Business Plan Competition. We strongly believe that participating in such events reflects a direct assessment of the business and technological potential of the proposed project here; particularly, such events have to go through a very rigorous selection process where startup ideas are evaluated by how well articulated are its commercialization potential. Additionally, the Rice Alliance for Technology and Entrepreneurship office has provided a letter of support outlining all the resources that will be available for our team during and after the completion of this PFI project, which includes summer programs, workshops, mentoring and networking, and various educational material available to us to validate many assumptions behind the technology commercialization of our proposed recommendation systems.

Project summary -1 page

1 Overview

This project aims to build upon the NSF award (#1750074), CAREER: Human-Centric Big Network Embedding grant, to further translate the academic research and technologies in Automated Graph Neural Networks into advanced recommendation systems to help small businesses adopt advanced machine learning (ML) technologies. The e-commerce industry is already heavy users of enterprise business analytics software. Many large enterprises can currently implement recommendation systems; however, existing approaches focus on heuristics to decide the optimal configuration of complex pipelines. The primary goal of this project is to develop a novel automated recommender system upon the NSF lineage's research fruits on automated and scalable graph neural networks. The product development will be based on the PI's leading automated machine learning project, i.e., AutoKeras. The PFI team will tailor the resulting automated recommender system to bridge the gap between machine learning expertise and software developments and reduce the requirements for software developers to launch ML-related services. This reduction in complexity and deployment cost can help small and medium businesses (SMB) leverage advanced ML technologies without the steep learning curve.

2 Intellectual merit

Modeling Complex User-Item Relationships with Efficient Deep Learning Solution: User-item relationship modeling has long been a critical factor for building a sound recommender system. Often, higher-order relationships between users and items are also the key information to discover the potential needs of customers. To tackle the problems above, one may leverage deep neural networks (DNN) and graph neural networks (GNN) to build a deep learning solution. However, the computational cost of DNN and GNNs is expensive, which may lead to an inefficient inference that damages the user experience. Furthermore, due to the over-expressive power of GNNs, current GNNs suffer from an over smoothing issue and, therefore, unable to model higher correlations between users and items effectively.

End-to-End Automated Recommender System: Typically, the development of an end-to-end recommender system pipeline requires intensive human efforts. The PI's previous efforts on AutoKeras provides neural architecture search which allows software engineers to conduct experiments on ML model without ML knowledge. Still, data processing is a heavy-duty job that requires domain experts to collaborate with data engineers to perform data processing tasks, which is expensive for small businesses. In addition, all of the existing AutoML solutions solely focus on ML model development, while data processing plays a key role in a successful recommender system. However, data processing automation is very challenging as the search space may be very large, which leads to a computational bottleneck. In addition, joint optimization of data processing and ML model is impossible without clearly identified application scenarios for shrinking search space.

3 Broader impacts

This PFI-TT project will expand the fundamental understanding and provide practical computational tools in dealing with an emerging and critical ML automation and interpretation problem for recommendation systems with significant applications in e-commerce. The proposed translational research's successful outcome will enable more SMB managers to build and deploy advanced recommendation system frameworks but at a fraction of current time and cost. Furthermore, the strong support from companies such as Amazon and Samsung has been instrumental in search to commercialize the PI's lineage findings and ML technologies. Such industrial partnerships will be key when translating NSF-funded technologies to the PFI partners Biarte and FreeFuse to effectively enhance the US e-commerce experience to simplify excessive amounts of data and offer more relevant products to online consumers. Recommenders are essential online channels: they shape the media people consume and the products they seek; therefore, it is critical to allow its access to broader SMBs in the US. By partnering with Biarte, FreeFuse, and the Rice University's entrepreneurial office, the PFI team effectively enhance partnerships between academia and industry in the US, which is key to drive innovation and economic prosperity in the country. Biarte, one of the partners, has expressed in their letter of support that "... Democratizing AI in search of supporting small businesses will boost small producers' social and economic development in the neighboring communities of Austin."

PFI-TT: Democratizing Deep Recommendation Systems for Small Businesses' Digital Transformation to E-commerce

1 Executive summary

1.1 The societal need and the customer

Small and medium businesses (SMB) worldwide are currently looking to improve their online operations to serve a market shifting to buy mainly on the web, which is often an overlooked collateral effect of the past stay-at-home restrictions worldwide. Recommendation systems are a valuable tool in e-commerce for SMB because they reduce data overload by providing meaningful results to customers that increase the possibility of converting a sale. While recommender systems have become an essential computational component in major giant e-commerce companies, their complicated nature prevents SMB from readily adopting and understanding powerful Machine Learning (ML) algorithms and recommendation systems in their daily online operations. Therefore, the competitiveness of SMB in the US can significantly benefit from easier-to-use and affordable recommendation systems.

1.2 The value proposition

This project aims to reduce the ML development speed and complexity to accelerate the adoption of advanced recommendation systems by SMB and help them thrive in the digital-first economy. Our team has developed a prototype to create recommendation systems [1, 2] at a fraction of current time and cost.

1.3 The innovation

This project builds upon *our highly praised open-source system, AutoKeras* [3], which has become one of the most used AutoML systems (with over 8,000 stars and 1,300 forks on Github), to provide automated recommender systems. Our proposed framework learns the graph representation for any modal data to reveal the intrinsic relatedness of data at the feature or instance level. This PFI project will allow SMB managers to automatically and efficiently develop their ML system.

1.4 The partnership

Our team has established a partnership with the following two SMB: Biarte offers various pantry products through sustainable end-to-end trade. This business is heavily transitioning its commercial operations to the web and will be at the center of this feasibility project as its transformation to an online presence is a sample of the challenges found in more SMB that our lineage project will aim to solve. On the other hand, FreeFuse is a video-streaming platform with the mission to enhance idea-sharing by making learning content more customized and engaging.

The resulting prototype will enable FreeFuse and Biarte to share their user and product data to mutually benefit from a better recommendation engine.

1.5 Training and leadership development in innovation and entrepreneurship

We are proposing an educational plan for our postdoctoral researcher that leverages the entrepreneurial resources available at Rice Alliance for Entrepreneurship. Mainly, our on-campus incubator offers an intensive program designed to provide our postdoc with the training, network, and co-working space to validate the business hypothesis of this project. The incubator administrators will assess these educational tools' results to prepare him to present on a go/no-go decision day, where our postdoc will receive feedback from mentors, entrepreneurs, and investors.

2 From NSF basic research to addressing a market opportunity

2.1 NSF lineage

This proposed project develops from the NSF lineage work entitled *CAREER: Human-Centric Big Network Embedding, Award Number 1750074*. Network embedding is currently employed to learn a low-dimensional representation in recommendation systems to facilitate network analytics applications, including node classification and network visualization. This NSF lineage project has developed advanced graph neural networks to improve network embedding learning in various real-world applications, including social network and modular biochemical analysis. The lineage project also investigates a novel direction to explore how human beings could better understand the results. This multidisciplinary research's successful progress is currently leading to advances in enabling domain experts to interactively and quickly analyze big network data with human knowledge, thus positively impacting various information systems' online activity. During our lineage project, Amazon has been a critical partner in this project by providing us with actual market demands and feedback on our results. We recognize the missing computational elements to bridge the gap between sophisticated recommendation systems and the technologies tailored to assist SMB managers in taking full advantage of advanced ML tools in e-commerce.

Besides, the ongoing project's current results have helped develop a human-centric framework for modeling and incorporating human knowledge in network embedding, tackling data challenges in ML, and enabling interpretation and interaction of network embedding results. Our team has investigated multiview learning and deep structured frameworks to integrate three human knowledge types from the node-, edge- and community-level into a unified framework. Given that real-world online activity could contain heterogeneous, large-scale, and dynamic human knowledge, our research group has developed corresponding solutions to handle the problems. Our team also developed global and local interpretation algorithms to explain network embedding and interactive learning algorithms to integrate user feedback to facilitate the human understanding of our research results.

We will automate the recommender systems by incorporating human expert knowledge in this project based on our previous successful outcomes. Due to the diverse data characteristics and modalities in recommendation systems in real-world applications, the underlying analyzing tools, such as Graph Neural Networks, must be laboriously and carefully designed. Our proposed framework plans to exploit the interpretation results and human knowledge to construct a robust and complete search space of analyzing tools, automatically optimizes the neural architectures, and improves network embedding learning by adapting to different scenarios.

2.2 Relevant NSF lineage results and broader impact

Our lineage project has already shown promising results reported in [4], where we presented Auto-GNN (AutoGNN), a Neural Architecture Search of Graph Neural Networks, to find the optimal neural architecture given a node classification task. Our team designed the search space, RCNAS controller, and constrained parameter sharing strategy explicitly for the message-passing-based GNN. Our experiment results show that the discovered neural architectures achieve competitive performance on both transductive and inductive learning tasks. Furthermore, the proposed RCNAS controller searches the well-performed architectures more efficiently, and the shared weight could be effective in the offspring network under constraints. Thus, this technique can be applied to facilitate the network analytics tasks in recommender systems.

One major application of the Auto-GNN project results is recommendation systems for e-commerce applications, as pictured in Fig. 1. Personalized recommendation systems are ubiquitous and the primary source of revenue for many online services such as E-commerce, advertising,

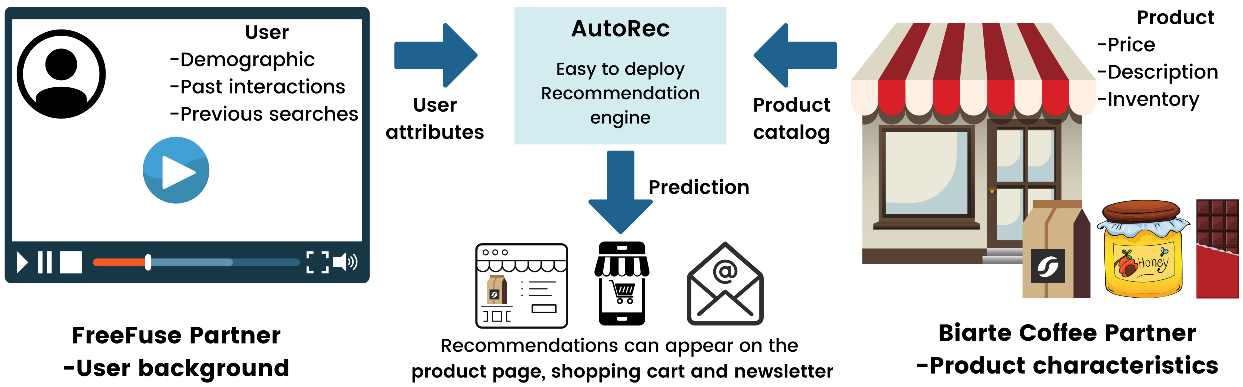

Figure 1: AutoRec application in e-commerce's recommendation systems.

and social media. At its core, a recommendation system estimating how likely a user will adopt an item based on historical interactions like purchases and clicks. As such, collaborative filtering (CF), which focuses on exploiting the past user-item interactions to achieve the prediction, remains to be a fundamental task towards effective personalized recommendation. CF's most common paradigm is to learn latent features (a.k.a. embedding) to represent a user and an item and perform prediction based on the embedding vectors. Matrix factorization is an early model that directly projects the user's single ID to her embedding. Later on, several researchers found that augmenting user-ID with their interaction history as the input can improve embedding quality. Because of the user-item interaction graph, researchers could see these improvements from using the user's subgraph structure — more specifically, her one-hop neighbors — to improve the embedding learning. Inspired by this, there is a surge of works recently that data scientists have proposed to model on such graph-structured data for practical user profiling. The essential idea behind them is to represent each user or item as a subgraph and then capture this subgraph's structure information by exploiting powerful graph neural networks GNNs.

Although GNNs based recommendation systems have shown promising performance in various industrial applications, such as social media, e-commerce, and advertising, it remains challenging to deploy GNNs-based systems in practice for two significant reasons. First, the user's interests often change over time, and most of the customer's subsequent behaviors are affected mainly by their recent actions, putting demands for model retraining frequently [5]. Second, the success of GNNs usually requires a lot of laborious works for architecture search. Thus, it is crucial to design an automated GNN framework to liberate people from such tedious works [6]. This project will provide a tailored neural architecture search for GNNs to tackle the challenges in recommender systems. By automating the GNN training, we accomplished an optimal GNN that captures users' actual interests on time.

In our previous research, Auto-GNN has achieved promising results in publicly available datasets. Now we propose to apply our techniques in real-world recommender systems. Specifically, we will leverage GNN to model the complex relationships among the users and items. Auto-GNN can be used to automatically identify the neural architectures and message passing mechanisms to capture the higher order relationships. In addition, we will also apply our technique in the feature engineering and feature interaction of recommenders systems to automatically identify the best solution in a data-driven manner.

2.3 Market analysis

Small and medium-sized businesses (SMB) worldwide are currently looking to improve their online operations to serve a market shifting to buy mainly on the web, which is often an overlooked collateral effect of the current lookdowns and stay-at-home restrictions worldwide. Recommendation systems are a valuable tool in e-commerce for SMBs because they reduce data overload by providing meaningful results to customers that increase the possibility of converting a sale. Allowing SMBs to build and deploy robust recommendation systems becomes more economically significant when knowing that only in the US, more than 47% of jobs come from SMBs.

Particularly, IBISWorld analysts indicate that the business analytics software industry will generate 86.7 billion US dollars in revenue in 2021 in the US; as shown in Fig. 2, the e-commerce market represents 17% of the market with a reported 11.3% annual growth over the past five years [7]. This industry heavily focuses on large enterprises with the data requirements and budget sizes necessary to take full advantage of recommendation systems, leaving an unmet need for smaller firms. **However, IBISWorld's report indicates that smaller businesses will become more critical to this industry over the next five years as they adopt enterprise technologies that will eventually become the standard**. We see an opportunity to serve this market by developing easy-to-use ML tools with intuitive user interfaces tailored to retail and e-commerce managing teams with a basic programming background.

Business analysts from IBISWorld also indicate that online businesses are already heavy users of enterprise software. Online companies are exceptionally prepared to adopt enterprise software because they already heavily use information technology as a core part of their business. The e-commerce market segment is relatively mature due to its early adoption of enterprise software and will continue growing as online businesses displace traditional competitors. Altogether, online companies generated more than 4.0% of industry revenue in 2020 of the business analytics market [7]. On the one hand, we believe that our lineage project can reach a more mature commercialization potential with translational research that can build on basic scientific research. To further lower the complexity of the production of less expensive production of tailor-made recommendation systems, we propose to apply AutoGNN to discover architectures for more applications such as graph classification and link prediction. By implementing more advanced graph convolution techniques in the search space to facilitate neural architecture search in different applications. On the other hand, we have created a team specifically designed to bridge the gap between our fundamental research and to find a product-market fit, specifically by understanding what are the needs in the market, defining our value proposition, understanding and defining customers, and finding a repeatable and scalable business model.

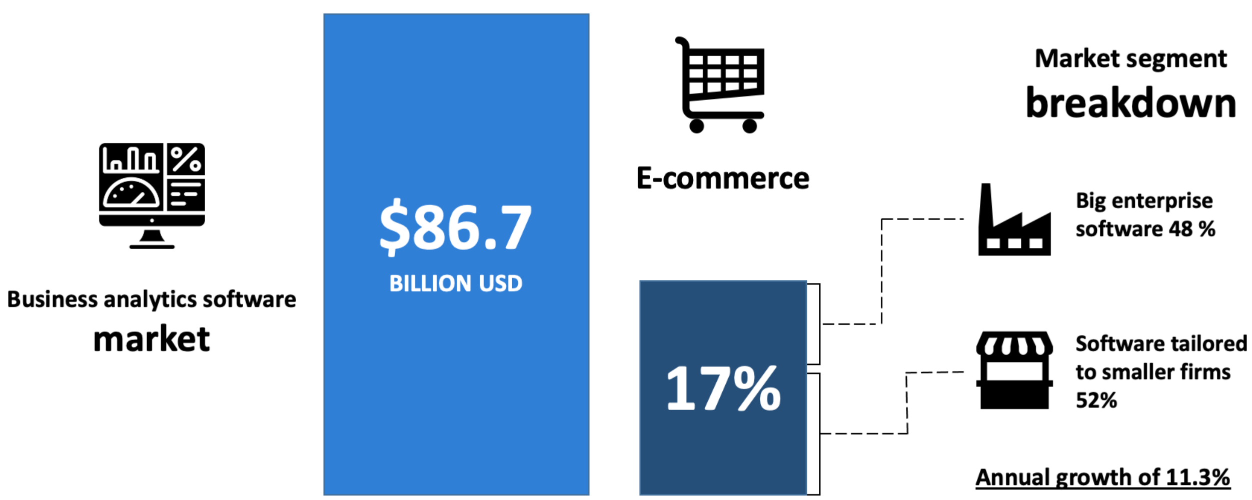

Figure 2: Business analytics software industry and e-commerce market segment concentration breakdown

2.4 Competitive technologies analysis

ML is a disrupting force across many industries; Statista projects ML-solutions' demand to grow more than 50% year to year over the next five years [8]. **Commercial recommendation technologies** such as Recombee [9], Crossing Minds [10], ExpertRec [11], and Strands [12], respectively, are increasingly investing in ML automation techniques. We are looking to complement these efforts by further expanding our human-centered explainability technologies that are already visible in our published research and our popular open-sourced projects. **We firmly believe that our competitive advantage is our strong focus on ML automation, which we inherently combine with interpretable artificial intelligence (AI) to maximize human knowledge with intelligent machine support**. Specifically, our work in Auto-GNN can be used in interpretation because of the inherent user-item connection of GNN by knowing how two items are correlated, i.e., when a user has bought an item or shown a strong interest in a product, human experts can use the visual representation of such correlations to improve their recommendation systems.

On the other hand, our indirect competitors also include the **Open Source** libraries Keras [13–15] and Tensor Flow [16–18]. While they offer elements to build recommendation systems, both libraries are not flexible for SMBs. Nevertheless, our goal is to push forward this effort by minimizing the business owners' learning curve and accelerating our recommendation system platform adoption. We do so by building proprietary software dedicated to ML automation on top of open-sourced neural architectures widely used by the AI community. Our vision is to use Automation and interpretation as key elements to make recommendation systems widely available for SMBs.

2.5 Intellectual property

Our team plans to protect the intellectual property generated in this project first via trade secret forms standard in enterprise solutions. Such practices include Non-Disclosure Agreements and non-compete clauses in the advent we have a spin-off startup, to name a few. In the same way, we plan to publish the advanced functions, such as automated machine learning, interpretable machine learning, and state-of-the-art recommendation modules, under the Berkeley Source Distribution (BSD) license, which forbidden users to develop their product based on these modules privately. Finally, our team has already approached an IP firm to explore patenting different systems that our team will implement in this PFI project.

3 Technical challenges and applied research plan

3.1 Innovation 1: deep learning for recommendation

3.1.1 Learning complex interactions between user and items

The core idea of developing a data-driven recommender system is collaborative filtering, which models the direct interactions between users and items. Traditional models perform collaborative filtering on user-item matrix [19–22] and employ matrix factorization techniques [23–28] to learn the latent factors of users and items for further predicting the preferences of users to items. Matrix factorization-based methods leverage the inner product to model the interactions which combine the multiplication of latent factors of users and items linearly. However, assuming users and items are linearly correlated is insufficient for complex real-world data distribution.

To address the problem, we propose to leverage the strong expressive power of deep neural network [29] to model the direct relations between users and items. Specifically, we will focus on implicit feedback, which indirectly reflects users' preferences through behaviors like watching videos, purchasing products, and clicking items. Compared to explicit feedback (i.e., ratings and reviews), implicit feedback can be tracked automatically and is thus much easier to be collected by content providers.

Figure 3 illustrates the neural collaborative filtering framework. The input of the framework is the sparse feature vector of a user/item. In this example, each entry representing whether the user/item has interacted with items/users. First, we will leverage a multi-layer perceptron to map the sparse feature vector into a dense user/item latent vector. Second, instead of leverage a simple inner product, we employ multiple deep neural network layers in Neural CF layers to model the non-linearity of the input user-item interaction. The neural architecture inside the Neural CF layers can be customized to discover certain structures of user-item interaction. Last but not least, the Neural CF layers pass the latent vector of interaction to the output layer for predicting the score y'_{ui} and compute the prediction error between the prediction y'_{ui} and the target y_{ui} to perform backpropagation and update the parameters of the framework. The framework's effectiveness has been validated [30] on various application scenarios, including movie recommendation and image recommendation.

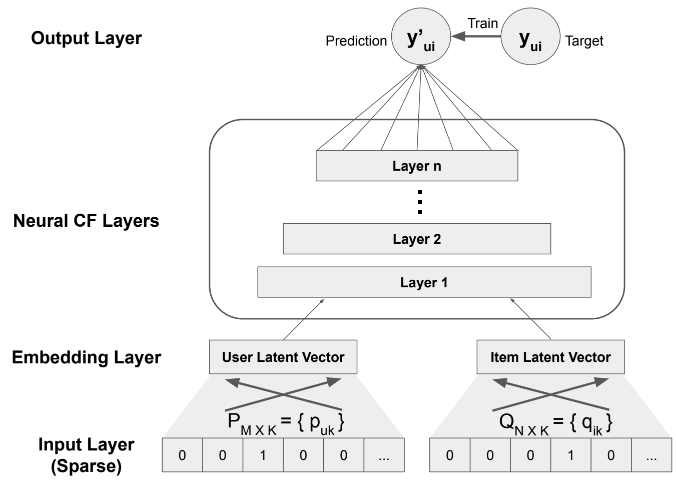

Figure 3: An overview of Neural Collaborative Filtering.

3.1.2 Modeling higher-order user-item relations

Following the recent success of deep learning on graph [31], graph neural networks (GNNs) have been widely adopted to recommendation problems, and various neural architectures [32–35] has been proposed to model deeper user-item relations to boost recommendation performances. By aggregating the neighborhood information for each node, GNNs learn the latent representation of nodes and perform downstream tasks such as node classification, link prediction, and recommendation. Furthermore, to model the higher-order correlation between users and items, GNNs extend the user-item matrix into a user-item bipartite adjacency matrix and form the bipartite graph where each node represents a user/item and each edge denoting the implicit/explicit feedback from the user to item. However, modern GNNs are suffered from several problems and over-smoothing problem [36] which lead to poor performance when capturing high order user-item relationships with deeper neural architecture. In addition, due to the computational complexity of the inference procedure, performing recommendations via classical GNNs will damage the efficiency of retrieving recommendation lists for users and, therefore, degrade the user experience.

To address the over-smoothing problem, we will develop our framework based on the previous research on group normalization technique [37]. Specifically, we relieved the over-smoothing issue from preserving group-structure characteristics and input features within node embeddings and proposed two over-smoothing metrics, i.e., group distance ratio and instance information gain. The group distance ratio measures the node embedding distances from different groups, which needs to be improved to separate the groups accompanied by various labels. The instance information gain measures the mutual information between input features and hidden node embeddings, which must be maintained to preserve the informative features for node classification. To optimize these two metrics, we developed a general module, called differentiable group normalization (DGN), applied between layers. It normalizes nodes within the same group independently and separates node distributions among different groups while keeping the input features during normalization. Experiments on real-world datasets demonstrate that DGN makes GNN models more robust to over-smoothing and achieves better performance with deeper GNNs.

Secondly, we address the retrieval efficiency based on our previous research, which learns a hash function for graph neural networks [38]. We investigate hashing with GNNs in this work and propose a simple yet effective discrete representation learning framework to learn continuous and discrete codes jointly. Our model consists of two components, a GNN encoder for extracting node representations and a hash layer for encoding representations to hash codes. A novel discrete optimization strategy based on a straight-through estimator (STE) with guidance is proposed to enable the hash layer differentiable and make our model is trained end-to-end. The principal idea is to avoid gradient magnification in the backpropagation of STE with continuous optimization guidance. Empirical results over several publicly available datasets demonstrated that our model can achieve comparable performance compared with its continuous counterpart and runs multiple times faster during inference.

3.2 Innovation 2: end-to-end automated deep recommender system

3.2.1 Automated CTR prediction

In terms of the application scenario, Click Through Rate (CTR) prediction is a crucial problem in many recommendation-related applications such as display advertising and search engine optimization. It drives the personalized experience for billions of users. Traditional solutions usually put effort into designing explicit feature interactions to capture the feature relationships and combine them with an MLP structure towards a two-tower model. However, the human-crafted architecture is often ad-hoc, and there lacks enough exploration on how to combine different types of explicit feature interactions with the implicit interactions learned from MLPs.

To address the problem, we will enable automated neural interaction discovery in our product based on our previous research on AutoCTR [39] to design a CTR prediction model automatically. It contains three key components to cope with three technical challenges. First, there are no dominant models in recommender systems such as the CNNs in CV tasks. We abstract and modularize simple yet representative operations in existing CTR prediction approaches to formulate a generalizable search space in AutoCTR. It also accommodates the heterogeneous and high-dimensional features in CTR prediction tasks. Second, since the designed search space is quite large, a good search algorithm is needed to provide an efficient exploration. We propose a hybrid search algorithm composed of an evolutionary algorithm and a learning-based algorithm based on a gradient-boosted tree. The algorithm provides a good balance between exploration and exploitation during the search process. It also utilizes a learning-to-rank loss, balancing the trade-off between different learning objectives and further enhancing search. Third, the CTR model is often learned on billions of data in practice. To further accelerate the search speed and reduce the space cost, AutoCTR involves a composited strategy of low-fidelity estimation, including data subsampling and hash size reduction. With abundant experiments on the three benchmark datasets, we empirically demonstrate the AutoCTR's effectiveness compared to human-crafted architectures and other classical NAS algorithms generalized from CV tasks. We also validate the generalizability and transferability of the discovered architecture across different datasets.

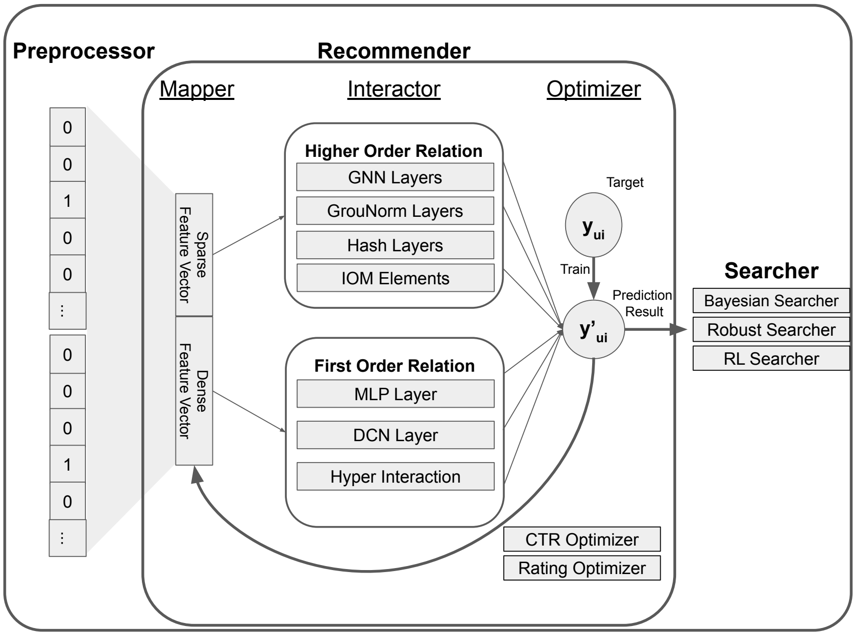

Figure 4: The structural diagram of the proposed automated recommender system.

3.2.2 Automated graph neural networks

As graph neural networks [40] have recently shown promise in various recommendation tasks [35, 41], we will automate the design of graph neural networks to enable better recommendation performance. Our system will focus on automation in two specific perspectives, including neural architecture design and aggregation optimization.

First, we will develop automated graph neural networks (AGNN) [42], which aims to find an optimal GNN architecture within a predefined search space. In AGGN, we have defined a tailored search space from general GNN architecture composed of layers of message-passing-based graph convolutions. To improve the search speed, we have designed a more efficient controller by considering a key property of GNN architecture–the variation of representation learning capacity with slight architecture modification. We have also explored the heterogeneous GNN architectures in the context of parameter sharing to train the architecture more stable with shared weight. We have shown that the GNNs discovered in this way consistently outperform the existing state-of-the-art handcrafted models. Our team can adapt this technique to the context of recommender systems to enable more advanced GNN models automatically.

In addition to the search of neural architectures, we will adapt the aggregation strategy for different users and items based on our study of PolicyGNN [43]. Our study shows that other nodes in a graph often need a different number of iterations of aggregation to achieve the best performance. For example, in the context of the user-item graph, different users may share the interests of other users in different hops away. Our team can thus personalize the recommendation strategy to meet the needs of different users. In our product, we will optimize this aggregation strategy with deep reinforcement learning in a data-driven manner to improve the overall performance.

3.2.3 Automated deep recommender system: AutoRec.

A fundamental challenge of a recommender system is the engineering cost to adapt to new recommendation scenarios. Specifically, realistic recommender systems are required to have the capacity to adapt to the constantly evolving data and tasks quickly or to explore different models systematically. One of the most prominent examples of this is that Netflix has never deployed the champion recommendation model of their 1M contest due to its engineering cost and the business' shifting from movie recommendation (rating prediction) to video streaming (click-through-rate prediction). In addition, although recommender systems start to capitalize on the power of deep learning, they have yet been able to convert model depth into raw performance because of the proneness to overfit, which leads to severe online-offline mismatch. Therefore, most of the industry's active recommendation models are shallow compared to their CV counterpart [44]. This calls for a new approach of recommendation model development that emphasizes both flexibility and the systematic exploration of existing and new neural architectures alike.

In the industry, Most recommender systems are highly specialized in handling specific data and tasks. For example, NCF [30] takes user-item implicit feedback data as inputs for the rating prediction task; DeepFM [45] leverages both numerical and categorical data for the CTR prediction task. However, a high degree of specialization comes at the expense of model adaptability and tuning complexity. As recommendation tasks evolve and additional data types are collected, the initially apt model can either become obsolete or require tremendous tuning efforts.

To bridge the gap, we developed an open-source automated machine learning (AutoML) platform extended from the TensorFlow [23] ecosystem named AutoRec, which focuses on neural architecture search (NAS) for deep recommendation models. AutoRec supports a highly flexible pipeline that accommodates both sparse and dense inputs, rating prediction and click-through rate (CTR) prediction tasks, and an array of recommendation models. We will incorporate AutoRec into our product to enable automatic pipeline construction and neural architecture search in industrial recommendation tasks.

3.3 Research and development plan

Our research and development will focus on an end-to-end, automated deep recommender system that our customers can readily adopt and deploy in real-world applications.

- □ **Developing deep recommender system.** We will build deep learning models to learn the complex interaction between users and items. We will also leverage the power of graph neural networks to model the higher-order user-item relations. Beyond the performance, we will develop interpretation functionalities to provide a better understanding of the models.
- □ **Automated deep recommender system** Based on our deep learning models, we will enable automated CTR prediction with neural architecture search, and we will also automate the design of neural graph networks to capture the higher-order relationships efficiently.
- □ **End-to-end robust deep recommender system.** We will wrap all of our models under the same framework with unified interfaces. We will also enhance the robustness of the system with robust loss functions.

3.4 Risk and mitigation plan

Our research and development will center on a recommendation system engine that our customers can readily adopt. High technology adoption barriers for FreeFuse and Biarte: there is a challenge in having the correct data structure from our partners to be promptly used in our recommendation engine. Another potential risk is that their current online infrastructure is not prepared to scale the implementation of our prototype quickly. To reduce such risks, our team will first adopt a gradual but iterative service development that will contemplate the compatibility of our products. Second, incremental iteration will help us find any further incompatibilities and challenges and propose a joint solution for our partners. Finally, we will provide guidelines on how Biarte and FreeFuse should collect their data to become gradually easy for our partners to prepare their data assets to use in the recommendation engine prototyped here. A success metric for our implementation with our partners Biarte and FreeFuse will be directly related to allowing them to find a way to share their information in a way that privacy is taking care of while helping them provide more meaningful recommendations to their end-users. In the same way, we plan to offer an easy way to integrate the recommendation engine into their operations, which, if this project is successful, will decrease in complexity in the successive development iterations of our service.

The sections above discuss the technical challenges and proposed solutions. Still, two technical risks might be raised by our customers. First, data quality might be a potential risk as target customers are small business which may not have sufficient or quality data to train an ML-based recommender system. In this case, we will first build a rule-based recommender system based on customers' domain expertise and collect the data. Then, based on the amount and quality of the collected data, we will build a hybrid recommender system that generate recommendation result based on domain expertise and ML models and gradually increase the portion of autoML model with increasing amount and quality of the data. Secondly, as the portion of autoML model increasing, customer may not trust the model as the whole operation is a black-box to our customer. To address their concern, we will provide a friendly interface to the customers to learn about the operation process of the model and visualize the data and recommendation outcome to allow customers to surveillance the whole process. In addition, interpretable machine learning techniques will be adopted to provide model interpretation in the interface to explain the complex interaction within the neural model to increase the credibility of the autoML model.

3.5 Timeline and milestones

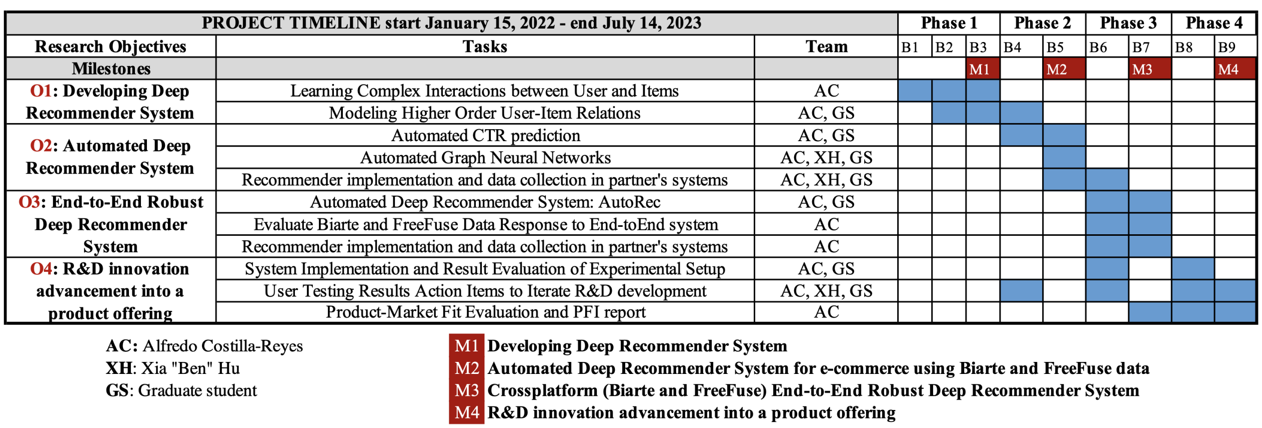

Figure 5: Recommendation system's PFI feasibility project timeline.

Overall, research and development will take 18 months. We will first start with implementing the system and then focus on enabling AutoML in our system. After that, we will spend half a year to polish and finalize our product based on the community's feedback. The milestone and timeline are summarized as follows.

- □ Developing deep recommender systems (6 months): We will evolve our product for commercial purposes. We will build upon our previous research lineage to develop an easy-to-use deep recommender system for a user-item recommendation.

- □ Developing automated deep recommender system (4 months): We will enable automated machine learning in a deep recommender system. The team will focus on neural architecture search for the automatic design of fully connected networks and graph neural networks.

After that, we will apply the robust loss to improve the robustness of the system. Finally, we will test our AutoML product internally and run evaluations on cloud services.

- □ End-to-end robust deep recommender system (4 months): We will closely support end-users to try our AutoML system. We will collect some feedbacks and improve our product. In this period, our objective is to ensure we provide a human-friendly interface for end-users.
- □ Finalizing prototype (4 months): Based on the customer's feedback, we will complete our work with improved interfaces. We will present our product in two options based on the input. First, we will tentatively wrap the product as software installed on personal computers or servers. Alternatively, we could deploy our development as a cloud service with a web-based user interface. We will choose one of the above options (or both) based on the user feedback from the open-source version.

4 Achieving societal impact through the realization of commercial potential

4.1 Commercialization strategy

We want to highlight that we have formed a team to push forward our technology's research development and commercialization efforts. Our current partnerships with Biarte and FreeFuse result from the product-market fit efforts and our first attempt to find a customer for our recommendation engine. With this being said, our team is entertaining the possibility of spin-off this technology from Rice University and forming an independent company. We are aware that a way to fund further development is by actively seeking non-dilutive funding support from federal agencies such as the Small Business Innovation Research grant. We believe that our successful results in the CAREER: Human-Centric Big Network Embedding grant prepared us well to propose this PFI grant to seek to commercialize our research.

We believe that the first step in our go-to-market strategy is to create a community when starting a business. The open-source community has also been an essential portion of our commercialization efforts. Notably, their demand, measure through downloads and reviews, has historically correlated to useful package features to the community. Releasing open-source packages serve two goals: first, it helps us gauge and understand download activity and user-feedback reviews to measure the success of our packages, and second, it is also a great way to build a community of programmers, researchers, and more importantly, other entrepreneurs that want to commercialize technologies in recommendation systems.

In addition, note that the technology commercialization office at Rice University will also be a valuable asset in our pursuit of technology commercialization. Mainly in two areas, their mentorship and entrepreneurial curriculum will be invaluable for our team, and secondly, they already possess a strong network of investors and domain experts. Finally, note that Rice University's incubator and accelerator, along with the customer discovery training provided by I-Corps, will be necessary for our commercialization strategy and the development of a spin-off company.

The NSF I-corps program will also help us gauge the success of our business partnerships. At this point, one of the business hypotheses we need to validate or invalidate is that our business model will follow conventional practices in the business intelligence industry, divided into non-recurring and recurring revenue streams. **Non-recurring revenue streams** involve setup infrastructure costs such as AI & ML model building, platform installation, operator training, deployment expenses, and data preprocessing. Our **recurring revenue streams** may follow standard models of cloud container companies [46] such as AWS®. They include ML deployment operation, license fees, data volume analysis, and extra premium features related to security and data encryption. We also believe that our software as a service (SaaS) expected margin is 80 to 85%, and our projected monthly burn rate is expected to be $40,000 for the first year. The I-corps program and their business model canvas methodology can undoubtedly help us gauge the validity of our value proposition and product-market fit in our quest to translate academic research and technologies into commercial use rapidly.

4.2 Assessment plan to gauge the success of the research partnerships and third-party collaboration

We have had multiple meetings with our partners, and we jointly defined the success metric for our partners and us as shown below:

"The primary metric our team can track and compare at FreeFuse.com to address the success of this partnership is directly related to the time our user engages with the recommendations provided by the proof-of-concept recommendation system provided by Dr. Xia (Ben) Hu's team." (Letter of support 1. Dr. Mike Liu, CEO, and founder. **FreeFuse**.)

"For this project, we expect to see an improvement in the following internal metrics that will measure this partnership work's success: website visits statistics and feedback from our clients and employees, customer engagement, click-through rate, and conversion rate in our website. In addition, we understand the PI can use this information to perfect the recommender technology and understand different market interrogatives, such as calculating the pricing of the technology that resulted in this feasibility project." (Letter of support 2. Carlos Torrebiarte, CEO and founder. **Biarte Coffee**.)

5 Project team

This team comprises expertise from data mining and machine learning systems (Hu), entrepreneurship and, data science (Costilla Reyes), including faculty and graduate students with strong experience developing AutoRec from at the Data Lab at Rice University.

PI Dr. Xia "Ben" Hu is an Associate Professor at Rice University in the Department of Computer Science. He has published more than 100 papers in major data mining venues. His articles have received seven Best Paper Award (candidate), and he is the recipient of the JP Morgan AI Faculty Award, the Adobe Data Science Award, and the NSF CAREER Award. An open-source package developed by his group, namely AutoKeras, has become the most used automated deep learning system on Github (with over 8,000 stars and 1,200 forks). Dr. Hu's work on deep collaborative filtering, anomaly detection, and knowledge graphs is part of the TensorFlow package, Apple production system, and Bing production system. Hu's work has been cited more than 10,000 times, with an h-index of 41. He was the conference General Co-Chair for WSDM 2020.

Dr. Alfredo Costilla-Reyes, will be brought as a postdoctoral researcher at Rice University in the Department of Computer Sciences. He graduated from Entrepreneurship and Technology Commercialization program at Mays Business School and the doctorate program in Electrical Engineering, both from Texas A&M University. Dr. Alfredo has been a recipient of the NSF I-Corps Site, 2017-2018 Kirchner, Silicon Labs and the McFerrin-Entrepreneurship Fellowships, and the prestigious Mexico National Youth Award, presented by the president of Mexico for his contributions in science, technology, and entrepreneurship. Specifically, Dr. Alfredo has led projects regarding embedded software and systems for future agriculture, battery-less wearable consumer electronics, application-specific integrated circuits, and wireless systems for IoT applications. His research and entrepreneurial endeavors have participated in YCombinator's YC120 event, Silicon Valley Bank Trek, McFerrin Center for Entrepreneurship business incubator, and Rice University's OwlSpark accelerator.

The team has established long-term collaboration, co-authored publications, and worked together on active DARPA and NSF projects to conduct fundamental research in developing interpretable and automated machine learning systems. In addition, our past collaborations have provided a solid ground for the technology commercialization of our NSF lineage technologies.

6 Partnerships

Outcomes of the proposed project will facilitate an automated recommender system in e-commerce applications by providing a more effective ranked list of items/products automatically. To better understand user needs and commercialization opportunities, as well as evaluate the effectiveness and efficiency of the developed systems on real-world datasets, PI Hu and the team will actively collaborate with corresponding domain experts from leading companies, including Adobe, Apple, Amazon, and Samsung, through funded joint research and development. The support of Amazon and Samsung are critical in pursuing our current technologies' commercialization (see letters of support). Notably, as the leading e-commerce, Amazon is very well positioned to push forward our efforts, and their experience with real-world problems will help us deliver better solutions tailored to small businesses around the globe. Moreover, we have collaborated extensively with Samsung's Mobile Ads AI team, which has provided vital specific problem needs, collaborative data curation, evaluation platforms, performance metrics, and feedback. Both partnerships are key for our commercialization strategy moving forward.

FreeFuse is an early-stage startup with the mission to enhance the exchange of ideas by making learning content more customized and engaging. Multiple recommendation systems are a mandatory requirement for startups like FreeFuse to show our relevant products to our visitors, improve customer engagement and incentivize sharing among our user's networks.

"Through this letter, we acknowledge the specific roles and responsibilities we will fulfill in this partnership. In the event the funding is granted, we would expect our part in this project to include:

- □ *Forming a true partnership and grant access to FreeFuse's user data and recommendation system's performance metrics.*
- □ *Help gather and share input from our users and customers to improve the recommendation prototype developed during and beyond this PFI-TT project timeline.*
- □ *Collaborate closely with other entrepreneurs who are part of this proposal to work jointly to provide better recommendations to the end-users.*
- □ *Collaborate in creative and innovative ideas on how to allow the resulting recommendation system for other early-stage startups and small businesses."*

(Letter of support 1. Dr. Mike Liu, CEO and founder. **FreeFuse**.)

Biarte Coffee is a Texas company that offers end-to-end coffee trade from the farmer to end-consumers that enjoy the finalized roasted coffee. We skip intermediaries and source our products directly from the producers. For my team, the digital transformation of our business has been a challenge, and during the COVID-19 lockdowns, it became clear that such digital transformation has accelerated and will become the standard in the next few years, and Biarte should be prompt to adopt advanced AI technologies. As a business, we need advanced machine learning tools that are affordable and easy to use to be indeed able to embrace a cost-effective solution in our operations, a need that I know other of our fellow retailers also require.

"As part of this PFI collaboration, we will be glad to share data regarding our products and the customers we serve. For example, a few categories of the products we offer online include coffee, all-natural honey, Biarte craft chocolate, premium quality cacao beans, accessories, and many product subcategories.

Such data will be available to the PI's team for this PFI project. In addition, my team will provide technical cooperation regarding how our current digital operation work so that the PI's team can get a sense of what type of data we would have for the recommendation engine." (Letter of support 2. Carlos Torrebiarte, CEO and founder. **Biarte Coffee**.)

The team has collaborated with the SMBs mentioned above owners and established methods for data-sharing through direct access with their companies. The purpose of such a broad range of collaborators and partnerships is to demonstrate the applicability and effectiveness of the proposed algorithms on real-world datasets and solve problems that matter to business owners.

7 Training future leaders in innovation and entrepreneurship

This PFI project will also help our team learn and implement a lean methodology into our daily tasks. Specifically, we want to benefit from the I-Corps training to validate and invalidate our hypothesis more systematically. At the same time, this PFI opportunity is already encouraging to look for partnerships beyond academia into the industry and other startup-tailored ecosystems. An additional leadership development goal is to help our graduate students and postdoctoral researchers properly articulate a very technical idea into solutions that can find the support of investors.

The Rice Alliance for Technology and Entrepreneurship office at Rice University has offered a letter of support outlining the different entrepreneurial and leadership tools available at Rice University that we will use in this PFI proposal to enable our postdoctoral researcher and graduate students to develop the skills needed for the successful commercialization of our NSF-funded research in engineering. The following programs were offered by the Rice Alliance for Technology and Entrepreneurship office at Rice University. First, the OwlSpark Accelerator is an intense, immersive 12-week summer program that will provide our team members with firsthand experience launching a tech startup. OwlSpark delivers entrepreneurship education throughout the summer, teaches relevant business fundamentals, pairs teams with knowledgeable mentors and industry experts, features office hours with successful entrepreneurs, and yields key connections. **The assessment plan of this accelerator includes evaluating the outcomes and impact of this training program with the Bayou Startup Showcase, a pitch presentation to the Houston startup community, and the university accelerator mentors.** An excellent opportunity for our team to showcase the PFI outputs of this proposal. OwlSpark has the resources and collaboration to provide us with exposure and direct access to the local entrepreneurial ecosystem, increasing the probability of continued success beyond the proposed PFI project.

Second, Rice University is one of the managers of the i-Corps Southwest Node, which has already walked us through the customer-discovery process of this NSF program. Assuredly, we want to mention that Rice Alliance for Entrepreneurship officials have already advised our team has already to participate in the I-Corps Teams program with Dr. Alfredo Costilla as the Entrepreneurial Lead (EL), Dr. Hu as the Technical Lead (TL), and our partners at Biarte and FreeFuse s as the Industry Mentor (IM).

Third, the Liu Idea Lab for Innovation and Entrepreneurship (Lilie) is the home of experiential learning and co-curricular activities in entrepreneurship and innovation at Rice University. We are currently considering different graduate courses and co-curricular offerings at the curricular level that fosters a growing university-wide entrepreneurship community available to graduate students and postdoctoral researchers participating in the project. Such courses include fundamentals in entrepreneurship to management and accounting classes.

It is worth highlighting that the Rice Alliance for Technology and Entrepreneurship is home to the Rice Business Plan Competition, the world's largest and richest intercollegiate student startup competition, and OwlSpark is Rice University's internationally recognized initiative devoted to the support of technology commercialization and the launch of technology companies through strong entrepreneurial-fundamentals education. Since 2013, 79 founders have successfully launched 33 startups through The Rice Alliance, making Rice University's entrepreneurial education offerings a catalyst for building successful ventures through education. As a result, our group will benefit significantly from the support we will receive in creating technology-based companies and the commercialization of new technologies in Houston.

Finally, we would like to take this proposal's opportunity to empower our team, including faculty members, postdoctoral researchers, and Ph.D. students, by harnessing the opportunity to learn and grow our entrepreneurial spirit. This proposal's I-corps component will help our graduate students and postdoctoral researcher serve as the Entrepreneurial Lead and find this project's commercialization. Such activities heavily stress the importance of reaching out to industry experts and getting their opinion about their current needs. As it is true in science, defining the correct business problem is more important than starting a solution based solely on a groundless hypothesis. We also plan to train the team with Disciplined Entrepreneurship: 24 Steps to a Successful Startup [47] developed at MIT, consisting of a series of lessons and methodology to formulate and test business assumptions. Another implementation we will put in place in this project is The Startup Owner's Manual [48], an Step-By-Step Guide for Building a Great Company developed by Steve Blank, which is a method highly recognized for its replicable product-market fit testing process.

8 Broadening participation

This proposal's inclusion plans extend beyond the research team and commercial partners to use our developments to serve SMBs such as Biarte and FreeFuse that are the foundation to increase the United States' domestic and international economic competitiveness. As we explained in this document, the current stay-at-home restrictions worldwide have made the importance of e-commerce in the global economy visible. Another critical issue is that while large retailers quickly adopt ML-based recommendation systems, these advanced tools are still costly and complicated, limiting smaller business owners' ability to benefit from their data. This project's mission is to close the currently growing digital divide between small businesses and large enterprises by providing access to high-quality and affordable ML tools for small businesses and be part of the vision of a more inclusive digital world for small business owners.

We want to provide more accessible ML and help women, minorities, and persons with disabilities grow their online businesses and thrive in the digital era in our quest to build an equal world by expanding the participation of women in the PI's research group and individuals from underrepresented groups in STEM, such as the founders of both partners Biarte and FreeFuse. Not to mention that by partnering with Biarte, FreeFuse, and the Rice University's entrepreneurial office, we effectively enhance partnerships between academia and industry in the United States, which is key to drive innovation and economic prosperity in the country. Therefore, the feedback of our partners will be critical to know what is important for the SMBs in this proposal, which will also help assess the reach of the broader impacts of this PFI proposal. As an example of the importance of such conversations, Biarte has perfectly summarized the PFI project's broader impact that matter the most to them in their letter of support:

"Having lived and grown up at a coffee farm, I firmly believe that this PFI project will empower small business owners to create a positive impact. I have witnessed how my family farm was a sustainable economic motor for the region and a leader in sustainable practices and fair trade. Democratizing AI in search of supporting small businesses will boost small producers' social and economic development in the neighboring communities of Austin. Furthermore, Biarte has a close connection to other producers in Texas who can significantly benefit from the outputs of this PFI-TT project." (Letter of support 2. Carlos Torrebiarte, CEO and founder. **Biarte Coffee**.)

References

[1] Ting-Hsiang Wang, Xia Hu, Haifeng Jin, Qingquan Song, Xiaotian Han, and Zirui Liu. Autorec: An automated recommender system. In *Fourteenth ACM Conference on Recommender Systems*, pages 582–584, 2020.

[2] DATA (Data Analytics at Texas AM) Lab. *AutoRec, an open-source automated machine learning (AutoML) platform extended from the TensorFlow ecosystem*, 2020 (Accessed February 9, 2021). https://autorec.ai/.

[3] Haifeng Jin, Qingquan Song, and Xia Hu. Auto-keras: An efficient neural architecture search system. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1946–1956, 2019.

[4] Kaixiong Zhou, Qingquan Song, Xiao Huang, and Xia Hu. Auto-gnn: Neural architecture search of graph neural networks. *CoRR*, abs/1909.03184, 2019.

[5] Yufei Feng, Fuyu Lv, Weichen Shen, Menghan Wang, Fei Sun, Yu Zhu, and Keping Yang. Deep session interest network for click-through rate prediction. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, pages 2301–2307, 2019.

[6] Thomas Elsken, Jan Hendrik Metzen, Frank Hutter, et al. Neural architecture search: A survey. *J. Mach. Learn. Res.*, 20(55):1–21, 2019.

[7] Dan Cook. *IBISWorld US Industry (NAICS) Report 51121C. Business Analytics Enterprise Software Publishing in the US*, July 2020. Retrieved from IBISWorld database.

[8] Tractica. *Revenues from the artificial intelligence software market worldwide from 2018 to 2025, by region (in billion U.S. dollars) [Graph]*, April 22, 2019 (Accessed October 6, 2020). https://www.statista.com/statistics/721747/worldwide-artificial-intelligence-market-by-region/.

[9] PitchBook Data. *Recombee | Private Company Profile*, February 2021. Retrieved from PitchBook database.

[10] PitchBook Data. *Crossing Minds | Private Company Profile*, February 2021. Retrieved from PitchBook database.

[11] PitchBook Data. *ExpertRec | Private Company Profile*, February 2021. Retrieved from PitchBook database.

[12] PitchBook Data. *Strands | Private Company Profile*, February 2021. Retrieved from PitchBook database.

[13] Nikhil Ketkar. Introduction to keras. In *Deep learning with Python*, pages 97–111. Springer, 2017.

[14] Antonio Gulli and Sujit Pal. *Deep learning with Keras*. Packt Publishing Ltd, 2017.

[15] Jojo Moolayil, Jojo Moolayil, and Suresh John. *Learn Keras for Deep Neural Networks*. Springer, 2019.

[16] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. Tensorflow: A system for large-scale machine learning. In *12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16)*, pages 265–283, 2016.

[17] Nishant Shukla. *Machine learning with TensorFlow*. Manning Publications Co., 2018.

[18] Joshua V Dillon, Ian Langmore, Dustin Tran, Eugene Brevdo, Srinivas Vasudevan, Dave Moore, Brian Patton, Alex Alemi, Matt Hoffman, and Rif A Saurous. Tensorflow distributions. *arXiv preprint arXiv:1711.10604*, 2017.

[19] Jonathan L Herlocker, Joseph A Konstan, Loren G Terveen, and John T Riedl. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS)*, 22(1):5–53, 2004.

[20] J Ben Schafer, Dan Frankowski, Jon Herlocker, and Shilad Sen. Collaborative filtering recommender systems. In *The adaptive web*, pages 291–324. Springer, 2007.

[21] Robin Burke. Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction*, 12(4):331–370, 2002.

[22] Gediminas Adomavicius and Alexander Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE transactions on knowledge and data engineering*, 17(6):734–749, 2005.

[23] Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, 2009.

[24] Andriy Mnih and Russ R Salakhutdinov. Probabilistic matrix factorization. *Advances in neural information processing systems*, 20:1257–1264, 2007.

[25] Ruslan Salakhutdinov and Andriy Mnih. Bayesian probabilistic matrix factorization using markov chain monte carlo. In *Proceedings of the 25th international conference on Machine learning*, pages 880–887, 2008.

[26] Xiangnan He, Hanwang Zhang, Min-Yen Kan, and Tat-Seng Chua. Fast matrix factorization for online recommendation with implicit feedback. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, pages 549–558, 2016.

[27] Steffen Rendle. Factorization machines. In *2010 IEEE International Conference on Data Mining*, pages 995–1000. IEEE, 2010.

[28] Yehuda Koren. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 426–434, 2008.

[29] Maithra Raghu, Ben Poole, Jon Kleinberg, Surya Ganguli, and Jascha Sohl-Dickstein. On the expressive power of deep neural networks. In *international conference on machine learning*, pages 2847–2854. PMLR, 2017.

[30] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. Neural collaborative filtering. In *Proceedings of the 26th international conference on world wide web*, pages 173–182, 2017.

[31] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and S Yu Philip. A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 2020.

[32] Rianne van den Berg, Thomas N Kipf, and Max Welling. Graph convolutional matrix completion. *arXiv preprint arXiv:1706.02263*, 2017.

[33] Jiani Zhang, Xingjian Shi, Shenglin Zhao, and Irwin King. Star-gcn: Stacked and reconstructed graph convolutional networks for recommender systems. *arXiv preprint arXiv:1905.13129*, 2019.

[34] Xiang Wang, Xiangnan He, Meng Wang, Fuli Feng, and Tat-Seng Chua. Neural graph collaborative filtering. In *Proceedings of the 42nd international ACM SIGIR conference on Research and development in Information Retrieval*, pages 165–174, 2019.

[35] Rex Ying, Ruining He, Kaifeng Chen, Pong Eksombatchai, William L Hamilton, and Jure Leskovec. Graph convolutional neural networks for web-scale recommender systems. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 974–983, 2018.

[36] Deli Chen, Yankai Lin, Wei Li, Peng Li, Jie Zhou, and Xu Sun. Measuring and relieving the over-smoothing problem for graph neural networks from the topological view.

[37] Kaixiong Zhou, Xiao Huang, Yuening Li, Daochen Zha, Rui Chen, and Xia Hu. Towards deeper graph neural networks with differentiable group normalization. *Advances in Neural Information Processing Systems*, 33, 2020.

[38] Qiaoyu Tan, Ninghao Liu, Xing Zhao, Hongxia Yang, Jingren Zhou, and Xia Hu. Learning to hash with graph neural networks for recommender systems. In *Proceedings of The Web Conference 2020*, pages 1988–1998, 2020.

[39] Qingquan Song, Dehua Cheng, Hanning Zhou, Jiyan Yang, Yuandong Tian, and Xia Hu. Towards automated neural interaction discovery for click-through rate prediction. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 945–955, 2020.

[40] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

[41] Wenqi Fan, Yao Ma, Qing Li, Yuan He, Eric Zhao, Jiliang Tang, and Dawei Yin. Graph neural networks for social recommendation. In *The World Wide Web Conference*, pages 417–426, 2019.

[42] Kaixiong Zhou, Qingquan Song, Xiao Huang, and Xia Hu. Auto-gnn: Neural architecture search of graph neural networks. *arXiv preprint arXiv:1909.03184*, 2019.

[43] Kwei-Herng Lai, Daochen Zha, Kaixiong Zhou, and Xia Hu. Policy-gnn: Aggregation optimization for graph neural networks. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 461–471, 2020.

[44] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[45] Huifeng Guo, Ruiming Tang, Yunming Ye, Zhenguo Li, and Xiuqiang He. Deepfm: a factorization-machine based neural network for ctr prediction. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, pages 1725–1731, 2017.

[46] Donna L Hoffman and Thomas P Novak. A conceptual framework for considering web-based business models and potential revenue streams. *International Journal of Marketing Education*, 1(1):7–34, 2005.

[47] Bill Aulet. *Disciplined entrepreneurship: 24 steps to a successful startup*. John Wiley & Sons, 2013.

[48] Steve Blank and Bob Dorf. *The startup owner's manual: The step-by-step guide for building a great company*. John Wiley & Sons, 2020.