# CHAPTER TWO

# LITERATURE REVIEW

## 2. 0 INTRODUCTION

The purpose of this literature review is to critically evaluate the existing research on deep learning models for personalized book recommendations. The aim is to identify the strengths and limitations of current approaches, highlight the research gaps and challenges and provide a basis for the development of a novel deep learning model for personalized book recommendation.

## 2.1 BOOK RECOMMENDATION SYSTEMS

Book recommendation systems are software applications that suggest books to users based on their interests and preferences. The goal of these systems is to help users discover books that they might be interested in reading but might not otherwise have discovered on their own.

There are two types of recommendation systems i.e., personalized and non-personalized recommendation system. Personalized recommendation systems recommend product or items to users based on preference criteria of individual (Sana, James & Nasseh, 2019). This system makes use of the individual past records of user’s ratings in order to find more suitable items that are yet to be rated by the user as recommendations (Khalid & Jamshed, 2020). On the other hand, non-personalized recommendation systems recommend base on the current popular trends that are relevant to all users. These trends can be the top 10 most popular items among users, information on a certain update or the current updates of a particular environment which are suitable for every new user (Khatwani & Chandak, 2016).

In recent years, deep learning models have become increasingly popular for book recommendations (Dhanashri et al., 2020). These models use artificial neural networks and other machine learning techniques to process large amounts of data and make recommendations based on patterns and trends in the data (Balaji, Pranshu & Deepali, 2020). Deep learning models can incorporate a wide range of data, including user behavior, book metadata, and social network information, to make highly personalized recommendations (Alexandros & Balázs, 2017).

However, there are several challenges associated with developing effective book recommendation systems. These include the cold start problem, which occurs when there is not enough user data to make accurate recommendations for new users, and the problem of data sparsity, which occurs when there is not enough data on certain books or users to make accurate recommendations. Additionally, privacy concerns related to user data collection and usage are an important issue that must be addressed when developing book recommendation systems.

This brief overview of book recommendation systems sets the stage for the subsequent sections of the literature review, which will dive deeper into the different types of recommendation systems and the challenges associated with developing effective deep learning models for personalized book recommendations.

## 2.2 Personalized Recommendation system

The emergence of personalized recommendation system can be traced back to the extensive research in cognitive science, approximation theory, information retrieval, forecasting theories, and also to management science related and to consumer choice modeling in marketing i.e., E-commerce. The recommender systems emerged as an autonomous research domain in the mid-1990s. This is the period where researchers started to focus on the recommendation problems that rely solely on ratings structure. The recommendation problem is reduced to the problem of recommender system to be estimating ratings for the items that have not been seen or rated by the user. Intuitively, this prediction or estimation is usually based on the user’s previous rated items. In this sense, if the estimate ratings for the yet unrated items can be obtain, then the item(s) with the highest estimated rating(s) can be predicted to the user (Gediminas, Alexander et al., 2005).

Personalized recommendation system is a computer-based system that offers individualized recommendations to users based on their previous behavior or history (Khatwani & Chandak, 2016). According to Dina F. et al. (2018), a recommender system in the context of education provides intelligent recommendations to students about study materials, courses, and related information based on their prior decisions. To provide such recommendations, the system employs a statistical model that is trained using a learning algorithm based on input from previous learning participants.

In a personalized recommendation system, users and items are the two main objects that play a crucial role. Users provide their interests about items, and this information is accumulated as input data in a utility matrix. The utility matrix represents the order of preference for particular items by the user in the combination of customer-item value. There are two classifications of recommender systems: user-based and item-based. In a user-based system, the user's interests, dislikes, and ratings are used to recommend unrated items. In contrast, an item-based system utilizes the relationship among items to generate recommendations for users (Geetha & Karthika, 2019).

Personalized recommendation systems play a vital role in both academia and industries. Many companies use personalized RS in their sales promotion, utilizing various platforms. For instance, the majority of the most-watched movies on YouTube and other online video databases come from the RS (Aminu & Naomie, 2019). The main purpose of building a recommender system is to provide maximum information required for the sole aim of personalizing learning and interests depending on the interactive patterns of users (Bouihi & Bahaj, 2019; Q. Zhang & D. Yang, P, 2020).

The algorithm for estimating the not-yet-rated items in the personalized recommender system is classified into three, namely: collaborative filtering, content base and hybrid approach (Kunal, Akshaykumar et al., 2017).

All authors agree that a personalized recommendation system is a system created with the major aim of recommending useful individualized information or items to users based on past learned historical records relating to the individual’s tastes or preferences.

### 2.2.1 Type of recommendation systems

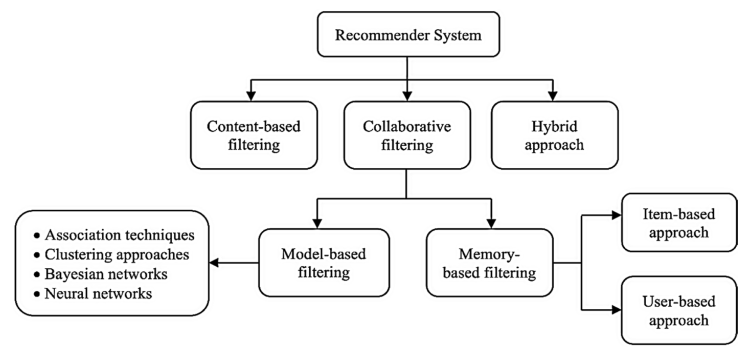
There are three main categories of recommendation systems: content-based recommender systems, collaborative recommender systems, and hybrid recommender systems (Roy & Dutta, 2022). Figure 1 provides a visual depiction of these different types of recommender systems.

### 2.2.1.1 Content-based filtering (CBF)

Content-based recommender systems are widely used in information or book retrieval. The process involves manually assigning terms and selecting appropriate techniques to compare these terms with the information in the user's profile. A learning algorithm is then applied to perform the chosen techniques and provide relevant recommendations to the user (Javed & Shaukat, 2021).

In content-based recommender systems, all the data items are collected into different item profiles based on their description or features. For example, in the case of a book, the features will be author, publisher, etc. When a user gives a positive rating to an item, then the other

items present in that item profile are aggregated together to build a user profile. This user profile combines all the item profiles, whose items are rated positively by the user. Items present in this user profile are then recommended to the user (Roy & Dutta, 2022).



*Fig. 1 Types of recommendation systems*

Term frequency (TF) and inverse document frequency (IDF) are important concepts used in information retrieval and content-based filtering systems, such as content-based recommenders. These concepts determine the relative importance of a document, article, news item, film, etc. Content-based filtering is a common approach in recommender system design, relying on item descriptions and user profiles. In a content-based recommender system, keywords are used to describe items, and a personal profile is constructed to suggest items that align with the user's preferences. In other words, these algorithms aim to recommend items similar to those that the user has liked or currently shows interest in. To achieve this, multiple candidate items are compared with items previously rated by the user, and the best matching items are recommended. Yannick et al. (2010) employed a content-based algorithm to filter and recommend the best results to users. The advantages of content-based recommender systems include transparency, independence, and recommendations for unclassified entities. However, they also have drawbacks such as limited serendipity, partial content analysis, and overspecialization (Adomavicius, Tuzhilin, et al., 2005).

By leveraging a content-based recommender system, recommendations can be generated based on the user's provided data or implicit interactions with the system. As the user provides more data or accepts suggested activities, the system becomes increasingly accurate in generating personalized recommendations. Compared to collaborative filtering systems, content-based approaches offer easier implementation and have been used in many projects due to their ability to generate highly relevant and transparent results. Content-based filtering strategies often employ techniques for content extraction, characterization, and building user and item profiles. However, these techniques have limitations, such as mismatches between user profile items and profile terms, leading to lower performance. Notable examples of content-based recommendations include Fab, which suggests web pages, and ELFI, which recommends financial information from a database (Lops, Gemmis & Semeraro, 2011).

Content-based filtering systems adapt their behavior to individual users by learning their preferences from previously relevant documents. The process involves comparing representations of item content with representations of user interests to find the most relevant items for each user. This task requires finding the best representation for both items (item profile) and users (user profile). A user profile represents a mapping of the user's real-world interests to a simplified domain model, while an item's profile describes its content using keywords or other means of representation. By enabling matching between user and item profiles, a common representation is established (Javed & Shaukat, 2021).

Content-based recommendation systems operate by analyzing a collection of documents and/or descriptions of items that have been previously rated by a user. These systems construct a model or profile of the user's interests based on the features of the rated objects (Lops, Gemmis & Semeraro, 2011). As described by Mladenic (2018), the profile serves as a structured representation of user interests and is utilized to recommend new and interesting items. The recommendation process primarily involves comparing the attributes of the user profile with the attributes of a content object (Khatwani & Chandak, 2016). This comparison results in a relevance judgment that indicates the user's level of interest in the object. A well-constructed profile that accurately reflects user preferences greatly enhances the effectiveness of an information retrieval process (Geetha & Renuka, 2019).

An example of the profile's utility is its application in filtering search results. By leveraging the user profile, it becomes possible to determine whether a user would be interested in a specific web page. If not, the system can prevent the display of that particular page (Lops, Gemmis & Semeraro, 2011).

#### High-Level Architecture of Content-based Systems

Content-based Information Filtering (IF) systems require effective techniques for representing items and generating user profiles, as well as strategies for comparing the user profile with item representations (Adomavicius, Tuzhilin, et al., 2005). In their study, Lops et al. (2011) outlined a three-step recommendation process, with each step handled by a distinct component:

1. **Content Analyzer**: The Content Analyzer is responsible for processing information without a predefined structure, such as text, and extracting relevant structured information. Its main task is to represent the content of items (e.g., documents, web pages, news, product descriptions) from various sources in a format suitable for further processing. Through feature extraction techniques, the component analyzes data items and transforms their representation from the original information space to the target space, such as representing web pages as keyword vectors. The output of the Content Analyzer serves as input for the Profile Learner and Filtering Component.
2. **Profile Learner**: The Profile Learner module collects representative data on user preferences and generalizes this data to construct the user profile. Typically, machine learning techniques are employed to infer a model of user interests based on items previously liked or disliked. For example, in a web page recommender system, the Profile Learner can utilize a relevance feedback method (Rocchio, J., 1971), where vectors of positive and negative examples are combined to create a prototype vector representing the user profile. Training examples consist of web pages that received positive or negative feedback from the user.
3. **Filtering Component**: The Filtering Component leverages the user profile to suggest relevant items by comparing the profile representation with the representations of items to be recommended. This comparison results in a binary or continuous relevance judgment, often computed using similarity metrics. In the case of continuous relevance, a ranked list of potentially interesting items is generated. In the previous example, the matching process involves computing the cosine similarity between the prototype vector and the item vectors.

The recommendation process begins with the Content Analyzer, which applies techniques from Information Retrieval systems to process item descriptions from information sources. The Content Analyzer extracts features such as keywords, n-grams, and concepts from unstructured text to generate a structured representation of items stored in the Represented Items repository.

To construct and update the user profile for the active user (Ua) who requires recommendations, the system collects the user's reactions to items and records them in the Feedback repository. These reactions, known as annotations or feedback, along with the corresponding item descriptions, are used during the process of learning a model that predicts the relevance of newly presented items. Users can either provide explicit feedback, where they actively evaluate items, or implicit feedback, which is derived from monitoring and analyzing user activities (Lops, Gemmis & Semeraro, 2011).

Explicit feedback can be obtained through approaches such as like/dislike ratings, numerical ratings, or text comments. Implicit feedback methods assign a relevance score based on user actions, such as saving, discarding, printing, sounds, manipulation of physical objects, or bookmarking of items. While explicit feedback is straightforward, the adoption of numeric or symbolic scales can increase cognitive load. Implicit feedback methods do not require direct user involvement but may be subject to biases, such as interruptions during reading (Meddeb, Maraoui & Zrigui, 2021).

According to Nilashi et al. (2013), user feedback and ratings can take different forms, which include:

* **Numerical ratings**: These ratings are represented by numbers on either discrete or continuous scales. Discrete rating scales, such as the popular five-star system or Likert response scales used in questionnaires, provide a limited range for users to rate items. Continuous rating scales often involve sliders that users can adjust to indicate a specific value.
* **Binary rating scale**: This type of rating allows users to categorize items into two distinct classes, typically denoted as "like" or "dislike." For example, platforms like YouTube offer users the option to give movies a thumbs-up or thumbs-down rating.
* **Ordinal ratings**: Users are presented with a set of options such as "strongly agree," "neutral," "disagree," or "strongly disagree." They are then asked to choose the option that best represents their opinion about a particular item, typically through the use of questionnaires.
* **Unary rating**: Users can assign items to a single positive class using unary ratings. A prominent example is Facebook's "Like" button, which allows users to express their positive opinion about a post. Implicit unary ratings can also be inferred from user actions like purchasing products on a web shop or clicking links on a webpage. The absence of a rating indicates a lack of information connecting the user to the item, potentially because the purchase was made elsewhere.

These different types of ratings provide users with diverse ways to express their preferences and opinions about items.

The Profile Learner component utilizes a training set (TRa) specific to the active user (Ua), consisting of pairs ⟨Ik, Rk⟩, where Rk represents the rating provided by Ua for the item representation Ik. By employing supervised learning algorithms on the labeled item representations, the Profile Learner generates a predictive model known as the user profile. This model is stored in a profile repository and utilized by the Filtering Component for future recommendations. When presented with a new item representation, the Filtering Component compares its features to those in the user profile, predicting the user's interest. Typically, the Filtering Component incorporates strategies to rank potentially interesting items based on their relevance to the user profile. The top-ranked items form a recommendation list (La) that is presented to the user (Ua). As user preferences change over time, the user profile is continuously updated by incorporating up-to-date information. User feedback on the recommendations helps refine the learning process, creating a feedback-learning cycle that adapts to the dynamic nature of user preferences (Lops, Gemmis, & Semeraro, 2011).

In summary, Content-based Information Filtering systems involve the Content Analyzer for item representation, the Profile Learner for constructing user profiles, and the Filtering Component for suggesting relevant items based on user profiles. These components interact in a recommendation process that incorporates user feedback to adapt to evolving user preferences.

### 2.2.1.2 Collaborative Filtering (CF)

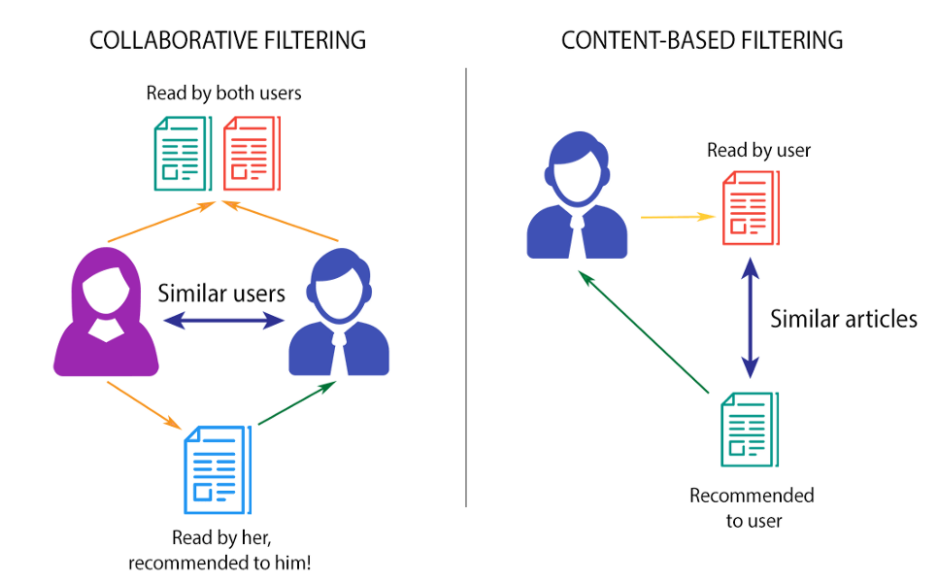
The term collaborative filtering was coined in 1992 by Goldberg et al., who proposed that information filtering becomes more effective when human group relational preferences or involvements are consider while recommending items to users (Kunal, Akshaykumar, et al., 2017). This method of recommendation tries to predict the utility of items for a particular user based on the items previously rated by other similar users (Gediminas, Alexanda, et al., 2005). According to Kunal et al. (2017), collaborative filtering is a method of recommendation system in which recommendations are generated to various users by actively comparing the preferences of one active user with another who have qualified products similarly with the user in the past. Conversely, Sana et al. (2019) describes collaborative filtering as a technique that help make a recommendation by finding users similar interests to predict relatively. CF are based on the idea that people who agree with the evaluation of items in the past are likely to agree again in future.

Collaborative filtering method are grouped into two general classes: neighborhood based and model-based (Gediminas, Alexanda, et al., 2005; Kunal, Akshaykumar, et al., 2017).

*A. Neighborhood Based Method*

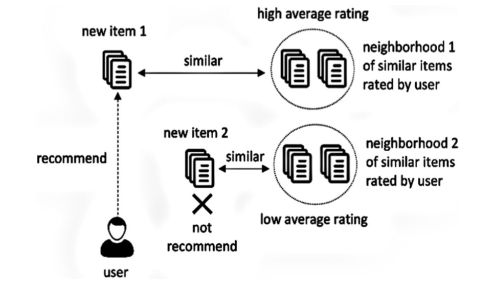
The neighborhood-based method of collaborative filtering, also known as memory-based or heuristic-based CF, leverages the user-item ratings stored in memory to make predictions about the preferences for new items. This can be accomplished through two approaches: user-based and item-based recommendation (Qian, Jie, et al., 2020).

In the user-based approach, the interest of a user (referred to as "u") in an item (referred to as "i") is determined by considering the ratings given to that item by other users who exhibit similar rating patterns. These users are known as "neighbors". By examining the ratings of these neighbors, the system can estimate the likelihood of user u being interested in item i (Dhanashri, Nandani, et al., 2020). This approach as explained by Roy & Dutta (2022), the user rating of a new item is calculated by finding other users from the user neighborhoods who has previously rated that same item. If a new item receives positive ratings from the user neighborhood, the new item is recommended to the user. Figure 2 depicts the user-based filtering approach.



*Fig. 2 User-based collaborative filtering*

On the other hand, the item-based approach involves predicting the rating that user (u) would assign to item (i) based on the ratings that user u has given to items similar to i. In this approach, the similarity between two items is calculated by examining the ratings provided by other users of the system. By analyzing these ratings, the system can identify items that are similar in terms of user preference and use this information to make predictions about user u's rating for item i (Bushra, Imran, et al., 2019). Also, Roy & Dutta (2022) addresses the fact that, in the item-based approach, an item-neighborhood is built consisting of all similar items which the user has rated previously. Then that user’s rating for a different new item is predicted by calculating the weighted average of all ratings present in a similar item-neighborhood as show in the figure 3.



*Fig. 3 Item-based collaborative filtering*

As stated by Nilashi et al. (2013), all collaborative filtering methods share a capability to utilize the past ratings of users in order to predict or recommend new content that an individual user will like. The real assumption is highly based in the idea of likeness between users or between products, with the similarity being expressed as a function of agreement between past ratings or preferences.

Overall, the neighborhood-based method of collaborative filtering utilizes the existing user-item ratings to determine recommendations for new items. The user-based approach focuses on finding users with similar rating patterns, while the item-based approach concentrates on identifying similar items based on user ratings. By employing these methods, the system can provide personalized recommendations that align with user preferences.

Kunal et al. (2017) classified the main advantages of neighborhood-based methods as follow:

* **Simplicity**: Neighborhood-based methods are relatively straightforward to implement, making them accessible and easy to work with.
* **Efficiency**: This method does not require expensive training phases that need to be regularly performed in large commercial applications. This saves computational resources and time.
* **Justifiability**: Neighborhood-based methods provide concise yet comprehensive explanations for their predictions. This makes it easier for users to understand why certain recommendations are made.
* **Stability**: Despite the continuous addition of items, users, and ratings typically encountered in major e-commerce applications, neighborhood-based methods remain relatively unaffected. They maintain stability in their recommendations even as the system evolves over time.

*B. Model-Based Recommendation Methods*

The model-based recommendation method differs from neighborhood-based systems in that it doesn't directly use stored ratings for prediction. Instead, it utilizes these ratings to acquire knowledge and build a predictive model. In model-based approaches, the aim is to replicate and model the interactions between users and items, incorporating factors that represent the hidden characteristics of both users and items within the system. These factors can include user preferences and item categories, among others.

To implement a model-based system, the available data (usually in the form of a training dataset) is used to train the predictive model. The model is trained to understand the patterns and relationships between users, items, and their hidden characteristics. Once the model is trained, it can be applied to predict ratings for new items based on user behavior and item attributes.

Numerous model-based approaches exist for the task of item recommendation, each employing specific techniques. These techniques include Bayesian Clustering, which utilizes probabilistic clustering algorithms; Latent Semantic Analysis, which applies matrix factorization techniques to uncover latent factors in user-item interactions; Support Vector Machines, which employ machine learning algorithms for classification and regression tasks; and Singular Value Decomposition, which decomposes the user-item rating matrix into lower-dimensional representations (Nilashi, Bagherifard, et al., 2012; Geetha & Renuka, 2019)

These model-based approaches offer a more sophisticated and comprehensive way to make recommendations by capturing underlying patterns and relationships within the data. They leverage machine learning and statistical techniques to create predictive models that can effectively estimate user ratings for new items, enhancing the recommendation process.

### 2.2.1.3 Hybrid approach

Hybrid recommender systems have gained popularity as an effective approach to enhancing the accuracy of predictions in recommender systems. These systems aim to achieve better recommendation accuracy by combining collaborative filtering, content-based filtering, and other techniques. The hybrid approach can be implemented in several ways. Empirical evaluations consistently demonstrate that hybrid methods generate more accurate recommendations than independent approaches, such as pure collaborative and content-based methods (Da'u & Salim, 2019).

According to Roy and Dutta (2022), hybrid techniques involve the combination of two or more recommendation techniques to overcome limitations and enhance the accuracy and effectiveness of individual techniques. These combinations generally lead to improved performance and accuracy in recommender applications. Different approaches can be used to incorporate different techniques, such as combining the outcomes of separate techniques or employing content-based filtering within a collaborative method. Common approaches to hybridization include meta-level, feature-augmentation, feature-combination, mixed hybridization, cascade hybridization, switching hybridization, and weighted hybridization.

Geetha, Safa, Fancy, and Saranya (2018) also state that a hybrid recommender system combines multiple recommendation techniques to generate recommendations. Compared to collaborative or content-based systems, hybrid systems typically achieve higher recommendation accuracy. By combining both approaches, the collective knowledge is increased, leading to more informed recommendations. This knowledge enhancement presents promising opportunities to explore novel ways of enriching collaborative filtering algorithms with content data and content-based algorithms with user behavior data.

Netflix serves as an exemplary case of the successful implementation of hybrid recommender systems. The recommendations provided by Netflix are generated by analyzing the viewing and searching history of similar users (collaborative filtering) and suggesting movies that share similar characteristics with highly rated movies from a user's past preferences (content-based filtering) (Geetha & Renuka, 2019).

In summary, the authors agree that hybrid recommender systems, which combine collaborative filtering, content-based filtering, and other techniques, are effective in improving the accuracy of recommendations. They emphasize that hybridization enhances the performance and accuracy of individual techniques by overcoming limitations. Various approaches can be used to combine different techniques, leading to improved recommendation outcomes. The authors also highlight that hybrid systems typically achieve higher recommendation accuracy compared to purely collaborative or content-based systems. By combining approaches, the collective knowledge is increased, enabling more informed recommendations. The successful implementation of hybrid recommender systems by Netflix serves as a notable example.

## 2.3 Deep Learning in Recommendation Systems

### 2.3.1 Introduction to deep learning

Deep learning, also known as deep structured learning or hierarchical learning, is a branch of machine learning that encompasses a range of techniques designed to create high-level abstract models from data. The architecture of deep learning draws inspiration from artificial intelligence and seeks to simulate the deep learning process and sensory layers of the human brain. By learning from past experiences, deep learning models can make decisions to solve specific problems encountered by machines (Bochie, Gilbert, Gantert et al., 2021).

One of the primary objectives of deep learning techniques is to automatically extract relevant features and abstractions from data. This process involves identifying the most critical information required for a given task (Mijwil & Abttan, 2021). These techniques are particularly effective when dealing with large volumes of unsupervised data, enabling the models to learn representations that capture the inherent structure and patterns present in the data. Deep learning models are capable of achieving high accuracy in tasks such as data classification and decision-making (Al-Zubaidi, Mijwil & Alsaadi, 2019).

The field of image recognition has witnessed the remarkable impact of deep learning techniques. In the 2016 ImageNet image classification competition, deep learning achieved unprecedented accuracy, surpassing 97% (Haiming et al., 2021). This breakthrough showcases the power of deep learning in surpassing traditional methods and pushing the boundaries of image recognition capabilities.

Deep learning has also revolutionized machine translation, particularly with the development of the Google Neural Machine Translation System (GNMT). Wu, Schuster, Chen, et al. (2016) demonstrated that GNMT achieved performance levels close to human translation in both English-Spanish and English-French translations. This advancement not only improves communication across language barriers but also showcases the potential of deep learning to tackle complex language tasks.

Furthermore, deep learning has made significant contributions to the field of speech recognition. Leading players in this domain, including Baidu, Xunfei, and Sogou, have utilized deep learning techniques to achieve Chinese speech recognition accuracies exceeding 97% (Haiming et al., 2021). Such breakthroughs in speech recognition have paved the way for applications in voice assistants, transcription services, and other speech-related tasks.

In the realm of online advertising, deep learning has emerged as a powerful tool for predicting click-through rates. Industry giants like Google and Microsoft have leveraged deep learning to improve the accuracy of their advertising platforms (Cheng, Koc, Harmsen, et al., 2016; Zhu, Shan, Mao, 2017). This application of deep learning has revolutionized online advertising by enabling targeted and personalized ad delivery, optimizing marketing campaigns, and enhancing user experience.

Deep learning has experienced remarkable growth in the realms of big data and artificial intelligence (Silver, Huang, Maddison, et al., 2016). By combining low-level features, deep learning creates richer, high-level abstractions that automatically uncover distributed data features. This transformative approach eliminates the need for manual feature design in traditional machine learning, leading to breakthroughs in various domains, including image recognition, machine translation, speech recognition, online advertising, and recommendation systems (Haiming, Kaili, Yunyun & Xuefeng, 2021).

In recent years, deep learning techniques have found applications in various domains, including the classification of medical images. Researchers have successfully applied deep learning models to automatically analyze and interpret medical images, aiding in the diagnosis of diseases and the identification of anomalies (Mehta, Aggarwal, Koundal et al., 2021). By leveraging the hierarchical learning capabilities of deep learning, these models can extract intricate features and patterns from medical images, enabling accurate and efficient analysis (Li, Zhao, et al., 2021).

The growth of deep learning in big data and artificial intelligence has been fueled by its ability to automatically extract meaningful features from data, surpassing traditional machine learning approaches. Its impact spans across diverse domains, including image recognition, machine translation, speech recognition, and online advertising. These advancements have not only achieved unprecedented levels of accuracy but also opened new possibilities for solving complex problems and driving innovation in various industries (Zhang, Yang, Chen & Li, 2017).

### 2.3.2 Advantages of deep learning in recommendation systems