Journal of Informatics and Web Engineering

Vol. 5 No. 1 (February 2026) eISSN: 2821-370X

A Conceptual Framework on Development of Sign Language Chatbot for E-Commerce

Salma Jahan Nisha[1], Nabhan Salih[2*], Wan-Noorshahida Mohd-Isa[3]

[1,3]Faculty of Computing and Informatics, Multimedia University, Jalan Multimedia, 63100 Cyberjaya, Malaysia.

[2]Department of Computer Science and Engineering, Ohio State University, Columbus, OH 43210-1210, USA.

[3]Centre for Image and Vision Computing, Multimedia University, Jalan Multimedia, 63100 Cyberjaya, Malaysia.

**Corresponding Author: (salih.43@osu.edu; ORCID: 0000-0001-7906-4071)*

Abstract—This study proposes and validates a conceptual framework using a Sign Language (SL) chatbot in the e-commerce domain to improve accessibility. Currently, the use of SL is a less-explored area in online platforms. It is difficult for any communication-challenged person to interact online for buying products and services, using a chatbot, particularly for people using SL. The objective of this study is to introduce a novel hybrid architecture using SL Recognition with a conversational-based chatbot agent via a custom Application Programming Interface (API) in e-commerce platforms. Our work proposes a combined hybrid chatbot framework model using Convolutional Neural Network (CNN) and Natural Language Processing (NLP). Python libraries such as Keras, OpenCV, and MediaPipe frameworks were used to read the signs in the system. To test this study at this initial stage, a preliminary feasibility experiment was conducted. Purposive sampling has been used to select 8 participants familiar with American Sign Language (ASL), who were tested under various conditions, including different lighting and clothing. The SL recognition module's initial performance data has been analysed using precision, recall, and F1 scores to assess ASL recognition and achieved an accuracy of 98%. This whole work showcases a blueprint to develop inclusive e-commerce platforms to encourage accessibility.

Keywords— Deep learning, Artificial Intelligence, Communication, Hybrid system, Convolutional Neural Network, Natural Language Processing.

Received: 29 October 2025; Accepted: 13 December 2025; Published: 16 February 2026

1. INTRODUCTION

Sign Language (SL) serves as the primary mode of communication for over 430 million individuals with disabling hearing loss worldwide. Unlike spoken languages, SL is a visual-spatial language governed by unique syntax, grammar, and structural rules involving hand motions, facial expressions, and body postures. However, the rapid digitization of the commercial sector has created a 'digital divide.' While e-commerce platforms have increasingly adopted conversational Artificial Intelligence (AI) chatbots to automate customer service, these interfaces mainly depend on text or voice-based processing of conversations. This dependency systematically excludes the Deaf and

Journal of Informatics and Web Engineering
https://doi.org/10.33093/jiwe.2026.5.1.24

Published by MMU Press. URL: https://journals.mmupress.com/jiwe

Hard-of-Hearing community, for whom text-based interaction can be a significant barrier due to varying literacy levels in spoken languages.

SLs are fully developed languages, like any other language with syntax, grammar, and so on. SLs, such as spoken language, vary significantly over the world [1]. For example, American Sign Language (ASL) is mainly used in the United States and some parts of Canada; however, British Sign Language (BSL) is used in the UK, and Indian Sign Language (ISL) is used by the Indian population in India. SLs are used by people with hearing impairments on a daily basis face-to-face. However, the application of SL on a generalized basis for commercial deployment online is not easily accessible because of the lack of methods or software for identifying and understanding SL interactions successfully [2].

A chatbot is a computer program that uses text, images, or voice recognition to conduct a conversation like a human one [3]. Such systems are very popular nowadays and are frequently developed to replicate human assistance with various online-based tasks. An example of such a system is ChatGPT. Chatbots are developed mainly using Natural Language Processing (NLP) and Natural Language Understanding (NLU), and it is also used in various platforms, or various frameworks to replace human agents in different domains and fields and come to the help of people online. In particular, in the e-commerce sector, sales and inquiry chatbots have gained popularity in the past few years for handling customer queries.

Technology has developed massively in e-commerce platforms, where we can utilize chatbots and other linguistic methods to communicate, purchase, make transactions, etc. AI chatbots are everywhere in e-commerce, but there is still a big problem: most of them only use NLP for text or voice inputs. This design effectively leaves the Deaf and Hard-of-Hearing community, who may prefer signing to writing because they can't read or write well or because they feel more comfortable doing so. There are independent SL recognition systems; however, they are not often used in business workflows. This project fills this gap by putting a visual pipeline based on a Convolutional Neural Network (CNN) right into the chatbot architecture. Unlike traditional text-based agents, this hybrid approach allows the system to visually perceive and interpret user intent through ASL, creating a truly inclusive e-commerce loop. Our current e-commerce solutions or stores do not have any specific way to meet the needs of the customer base who use SL. Even with all advancements, we still do not possess any proper system for individuals with communication barrier issues who can easily access online stores and services with ease. The implementation of a chatbot agent capable of real-time translation of SL into speech represents a pivotal advancement in fostering effective communication between individuals using SL and their interactions within e-commerce settings, thus retaining customer satisfaction by providing a good user experience. Customer satisfaction refers to a person's feelings after comparing the results he receives to his expectations. This ensures that the level of satisfaction is proportional to comparing the performance experienced with having said expectations. If expectations fail to be met, causing customers to be disappointed, the level of satisfaction will go down [4]. By harnessing NLP techniques and Deep Learning (DL) models, the chatbot agent can facilitate seamless and accurate translation of ASL gestures into spoken language. This work will make use of the NLP and DL method's algorithm CNN. CNN is a type of fully connected layer of neural networks that performs operations to solve problems related to computer vision [5]. This framework will make use of CNN to create the dynamic recognition solution, thereby bridging the communication gap and enabling inclusive e-commerce interactions for individuals with communication barriers.

This work presents visual recognition of SL using hand gestures with the integration of a chatbot to form a framework that can understand ASL in real time, send responses based on user queries, and help with e-commerce purchases and concerns.

2. LITERATURE REVIEW

The development of a sign language-enabled chatbot for e-commerce depends highly on the intersection of three well-researched areas, although often they are disconnected fields researched individually: conversational AI in the e-commerce domain, SL recognition technology, and digital accessibility for users. A thorough synthesis and analysis of the literature highlight the gap that while the individual components are well-created, as a holistic system it is not fully merged, creating a barrier to digital inclusion.

2.1 The Rise and Impact of Conversational AI in E-Commerce

Chatbots have recently gained popularity in performing tasks in various businesses by replacing human agents. Many businesses, including healthcare, education, and e-commerce, have adapted to use different types of chatbots to ease their tasks and focus on customer relationship growth. Eliza is one of the oldest and best-known chatbots, which was created by Joseph Weizenbaum in the AI Laboratory at MIT. The date for this chatbot is 1964-6. This program later inspired various developers in the field of AI [6].

The impact of chatbot usage has been evaluated by various researchers in the field of e-commerce. This is particularly true in terms of customer satisfaction. Katherine and Luis examined how customers feel about anthropomorphic chatbots in the food e-commerce industry. The chatbots' positive correlation with customer satisfaction is mediated by the factors of enjoyment, attitude, and trust [7]. Research has been performed on how different language styles used by chatbots can influence customer intention and attitude towards receiving services from any associated brand. This study focused on two scenario-based informal language styles explained by the concept of parasocial interaction [8].

The impact of chatbots on customers' purchase intentions in the e-commerce context was explored. According to this investigation, chatbots act as a medium for appealing and ingratiation cues, thus influencing customer purchasing decisions. Online experiments and surveys were conducted to examine the effects of chatbot cues on customer engagement and purchase intentions. These findings on chatbot integration proved to be valuable for e-commerce [9]. Another chatbot for the e-commerce domain was proposed to enhance the online shopping experience by using a virtual assistant. The sales assistant's role was to guide customers and resolve their doubts. User acceptance tests were conducted to validate the effectiveness of the bot in terms of customer service and interactions on e-commerce platforms [10].

Researchers have developed various e-commerce sales chatbots using different frameworks and programming languages. Using a modular chatbot framework and other programming languages, a chatbot was developed to provide support to customers for a variety of services. Likewise, to provide support in the e-commerce domain, the Django platform, Hypertext Markup Language (HTML), Cascading Style Sheet (CSS), JavaScript, and Bootstrap were utilized. The system was built to be quick and reliable and to support better judgments. "Hebron" was developed for a community mall in a university with the use of Python and React.js for programming and MySQL for databases to provide a comfortable shopping experience for everyone in the university [11-13].

For user responses, a chatbot was designed using Artificial Intelligence Markup Language (AIML), which resulted in an average response time of 0.3 seconds over 300 trials. Parsing, pattern matching, and crawling data were performed using AIML [14]. Thus, in [15] it was reported that AI chatbots have an impact on online consumer happiness and experience. Customer experience is affected by chatbot usability and reactivity.

Prior to our daily lives, chatbots have been utilized in a variety of contexts and capacities. For example, consider chatbots found on banking and e-commerce websites, feedback forms, social media platforms, voice-activated, information-seeking, contextual, and many other websites [16]. An inquiry bot for college students was developed using Rasa Core and Rasa NLU packages. A Recurrent Neural Network (RNN) is used to build a probabilistic model that achieved an accuracy of 0.725 and a precision of 0.628 [17]. Another chatbot was built using Discord. The chatbot was a rule-based system built to demonstrate how a bot can be integrated with other web-based platforms and can be utilized for teaching and learning [18].

Overall, based on the findings, it can be reported that AI chatbots positively impact online consumer interactions through their usability. However, it relies heavily on text or sometimes voice, inherently excluding users whose primary mode of communication is sign language, creating an accessibility gap.

2.2 Technical Maturity of SL Recognition

SL recognition technology has been an interesting research topic for a long time. According to the World Health Organization, over 430 million people, approximately 5% of the world's population, suffer from hearing loss, and this number is expected to increase to over 700 million by 2050 [19]. These people who are facing such communication barriers use SL as their primary mode of communication. For this demographic, SL serves as the primary mode of communication, necessitating digital tools that bridge the gap between signers and non-signers. Early studies has been done to address this integration focused on basic chatbot, for instance, Pardasani et al. used computer vision techniques to translate ASL inputs into audio-text outputs, though this Python-based system achieved an accuracy of 87% [20]. Improving upon user interaction, more research integrated Italian Sign Language (LIS) videos into chatbot interfaces, showing that visual aids significantly modify user engagement and social inclusivity for the hearing-impaired community [21].

In terms of data acquisition, visual recognition remains the predominant modality due to the ubiquity of cameras compared to wearable sensors. Consequently, the majority of modern SL recognition systems rely on two-dimensional (2D) camera inputs to parse video frames [22]. A vision-based 8-layer CNN technique for SL identification was suggested, with a validation accuracy of 99.34% on a self-created dataset of 52 classes. This was done using the ISL [23]. A study was done to search for static signs in SL recognition using DL-based CNNs. Pre-trained architectures were used to achieve this sign feature extraction, and the model used for this experiment were the modified versions of the AlexNet and VGG16 for classification. With the technique of leave-one-subject-out and random 70-30 cross-validation options, several layer features were tested for higher performance, resulting in a very good score of 99.82% identification accuracy [24].

To further refine feature extraction, researchers have explored advanced architectures that go beyond simple static analysis. Multi-stream CNNs utilizing Spatial Pyramidal Pooling (SPP) have been proven to effectively gather multi-scale data, achieving varying degrees of success across datasets: 99.6% on ASL Fingerspelling and 99.93% on ISL, though performance dropped to 93.75% on complex RGB-Depth datasets [25]. An ensemble model was proposed to extract capital features using the inception model, which is a CNN-based method. The temporal feature uses RNN on a self-created video dataset in ASL. After converting videos per frame, the inception model and RNN algorithms were applied to the outputs from the softmax and max pooling layers for classification. The accuracy achieved was 90% for the first 10 signs and 91% for the next 150 [26]. This accuracy was achieved using a softmax layer.

Recent optimizations have also focused on handling diverse input conditions and edge deployment. For regional languages like Bengali, normalization techniques using HSV and YCbCr colour segmentation have enabled CNN models to achieve 99.2% accuracy in character recognition [27]. For alphabets and numbers, a YOLOv5 model was proposed to recognize signs. An alteration to the existing ASL was performed to prepare a self-created dataset, and the YOLOv5 algorithm was applied to test recognition. The achieved mean average precision (mAP) of accuracy was 98%. This has a deployment advantage for edge devices such as a light YOLOv5 model [28]. The preprocessing techniques suggested by researchers include a histogram of gradients, principal component analysis, and local binary patterns. The model in this study was created using Canny edge detection, Oriented FAST and Rotated BRIEF (ORB), and the bag-of-words technique. The data is passed through different models of classifiers, including Random Forests (RF), Support Vector Machines (SVM), Naive Bayes (NB), Logistic Regression (LR), K-Nearest Neighbours (KNN), and Multilayer Perceptron (MP), to draw effective results [29]. AI-powered technology was introduced to recognize SL to allow signers and non-signers to communicate in both directions. This paper proposes a method of sentence recognition based on the DL model's recognition of 50 words and 20 sentences. The algorithm recognized 20 phrases and 50 words with an accuracy rate of 86.67% [30].

The three feature extraction streams included in one of the research studies are the CNN, Gabor filter, and ORB feature descriptor. It is mentioned that specific features were extracted from the hand gesture photographs of each stream separately during this experiment, using these three structures individually. The proposed algorithm achieved three different results on different datasets, attaining average precision rates of 99.92%, 99.8%, and 99.80% on the Massey, ASL Alphabet, and ASL databases, respectively [31]. Static signs in the ISL were recognized using a DL approach. Based on a dataset of 35,000 photos and using CNNs. The accuracy achieved was 99.72% for the colour images and 99.90% for the grayscale images [32].

Finally, the focus of modern SL recognition research is increasing robustness and user-based usefulness. Novel complex CNN models have been developed to resolve communication barriers, improving SL identification which increases by significant margins of around 9% over baseline models [33]. An effective 3DCNN method for hand gesture identification was studied and presented. This technique uses a transfer learning method to achieve high identification rates across various datasets and shows a promise for the development of precise and reliable SL recognition system [34].

Another study used unique DL models for hand gesture identification of emergency signs in ISL to assist hearing-impaired people in delivering urgent messages during emergencies. The study achieved high accuracy in recognizing emergency signs by combining classification and object detection methods, with the VGG + LSTM model excelling in classification and the object detection model successfully distinguishing dynamic motions. This study mainly puts its focuses on the ability of using DL technology to improve communication barrier among hearing-impaired people in important circumstances, showing a solid method to improve their safety and well-being [35].

2.3 The Unmet Need for Integrated Digital Accessibility

Despite the improvements of both chatbot and SL recognition technologies, a significant gap exists in their integration to address the needs of the deaf and hard-of-hearing community. The literature review done on digital accessibility highlights the issues people with disabilities face online and stresses the need for a systematic approach to eliminate these barriers.

A study was conducted based on 27 articles to review facilitation and communication improvement between the deaf and hearing people. Various solutions have been discussed, including gesture recognition and online content. However, based on the study, most of the technologies were on the prototyping level, indicating a further need to build systems or technologies based on SLs for the world's deaf communities [36]. A study conducted on accessibility to digital technology shows the persistent issues people with disabilities encounter while using digital technology. This research stresses the dynamic character of digital accessibility, viewing it as a complex process driven by technological, legal, economic, and social institutions. By investigating the impact of technical standards, hardware and software options, and legislative frameworks on accessibility, this study emphasizes the necessity of a systematic approach to eliminate barriers to digital inclusion and make communication accessible for people with disabilities [37].

Studies have shown that it is important to include accessibility considerations into industrial practices and government legislation in order to achieve universal accessibility for all types of users. They have mentioned the importance of inclusion of accessibility for people with disabilities so that they can be included more digitally [38-40]. According to findings, digital inclusion efforts not only empower people with disabilities by giving them access to e-commerce opportunities, but they also help them with their social inclusion, skill development, and autonomy when navigating online markets [41]. One study explains the goal of creating an accessible e-commerce platform to promote easy and happy buying experiences for people with disabilities and also potentially increase the everyday participation and increase the confidence of people with special needs [42]. Initiatives that provide accessible software are very important for delivering digital services and information to all types of users, including those with disabilities. Accessibility is a crucial component of usability and critical to the quality of software solutions. However, both academics and industry have made limited attempts to help developers create accessible web pages for deaf people [43].

Apart from the technological advancement of SL recognition algorithms, contemporary work has shifted its emphasis to Multimodal Human-Computer Interaction (HCI). Recent research in 2025 contends that recognition accuracy alone is inadequate for substantive digital inclusion; systems must additionally exhibit semantic intelligence to discern user intent within a transactional framework [44]. This view is supported by new studies on accessible conversational systems, which say that hybrid architectures that combine visual gestural inputs with NLU are important for making it easier for DHH users to use traditional platforms [45]. Moreover, building confidence in artificial intelligence-facilitated communication is essential, as users are inclined to embrace assistive technologies that exhibit reliable, low-latency performance in practical applications [46-47]. Table 1 summarizes related literature reviews, including their datasets and other aspects.

Unlike existing studies that focus primarily on translation accuracy in isolation, the novelty of this work lies in the hybrid integration of a lightweight CNN vision pipeline, which is the MediaPipe framework, with a domain-specific e-commerce intent engine. This approach allows for low-latency, browser-based deployment without requiring heavy external sensors, for example, data gloves, motion sensors, etc., specifically optimized for customer service interactions. The objective of this study is to design a combined hybrid framework for chatbot and SL for the e-commerce domain to drive towards the motivation of building an e-commerce framework that will support digital inclusivity by empowering all types of users, including users with hearing impairments, to use the system. As the e-commerce sector grows, the usage of SL recognition chatbots will become increasingly crucial because it improves user experience and supports diversity for all kinds of customers. Overall, the literature review indicated that increasing accessibility to digital technology is critical for encouraging social inclusion, economic growth, and creativity in the digital economy. A comparison of the research ideas and the proposed idea is presented in Table 2.

Table 1. Summary of Related Literature Review

Author	Purpose	Methodology	Dataset	Accuracy	Limitations
Sindhu et al., 2023 [9]	To understand the impact of chatbot customer purchase intent.	-Online experimental survey. -Structural equation model.	Actual dataset collected from the survey.	Chatbots make inspirational messages favourably affect buying intentions.	Real time human-chatbot interaction is missing.
Khan, 2020 [11]	Sales Chatbot	- Modular Framework - NLP - NLU	N/A	N/A	Do not mention specific datasets and accuracy.
Mostaq et al., 2022 [12]	Build an e-commerce sales chatbot.	-Django -HTML -CSS -JavaScript -Bootstrap	Product-based demo questions and answers.	Chatbot could send responses successfully.	Limited to specific users.
Nursetyo et al., 2018 [14]	Smart Chatbot for e-commerce Assistance.	- AIML - XML	1500 Questions	100% accuracy on formal words and sentences.	N/A
Chen et al., 2021 [15]	Influence of chatbot usability on customer experience.	-Quantitative approach. -Online questionnaires.	425 online questionnaires.	Positively influences customer experience based on responsiveness of chatbot.	N/A
Meshram et al., 2021 [17]	Build a college enquiry chatbot.	- RASA framework - NLU - RNN	488 sample data	98%-99%	Questionnaire failure.
M. G. C. P et al., 2021 [18]	Conversational Agent Chatbot.	- NLU - Discord Platform	450 educational conversations.	86.66% accuracy.	Limited to only simple questions.
Abul et al., 2021 [24]	Static SL recognition	- CNN - AlexNet - VGG16	ASL dataset	Recognition accuracy rate of 99.82%	Recognition performance is limited using simple differentiating motions. Cost-effectiveness was proved using a basic CPU system.
Singla et al., 2023 [25]	Static hand gesture recognition.	-CNN -SPP -Gabor Filter -Local Binary Pattern (LBP)	-ASL dataset -Massey University ASL Dataset -ISL Dataset -OpenGesture3D (RGB and Depth) Dataset	-99.6% -99.67% -99.93% -93.75%	Classifies only static SLs. Do not include dynamic SL recognition.
Bantupalli et al., 2018 [26]	Vision-based application to translate SL to text.	- CNN - LSTM - RNN	100 custom ASL datasets.	91%-93% - Softmax Layer 55% - 58% - Pooling Layer	Accuracy drops based on skin tones, different facial feature challenges, and different clothing challenges.
Tasmere et al., 2020 [27]	Bangla Sign Language (BSL) hand gesture detection	- CNN - HSV and YCbCr colour spaces	Custom dataset of 3219 images from 6 people.	99.22% in recognizing 37 characters	N/A
Damaneh et	Static Hand	- CNN	Datasets of	99.92%, 99.8%, and	The large number of

al., 2023 [31]	Gesture detection in SL	- Gabor Filter - ORB feature descriptor	ASL, Massey.	99.80% on the Massey, ASL Alphabet, and ASL databases	parameters is not feasible for implementation on mobile phones or microcontrollers.
Singla 2023 [43]	ASL recognition from image.	CNN Models -VGG16 -InceptionV3 -MobileNetV2	ASL Dataset	98% for MobileNetV2, 96% for InceptionV3, 91% for VGG16	Accuracy drops in different lighting conditions, hand shapes, and orientations.
Hassan et al., 2016 [48]	Sensor-based ASL recognition.	-KNN -HMM -Data Gloves -Polhemus G4 tracker	ASL dataset	97% recognition rate	A motion tracker is required, which can affect user experience.

Table 2. Comparison Between Proposed and Research Ideas

ASPECTS	RESEARCH FINDINGS	PROPOSED IDEA
Chatbot Framework	• Limited to serving only the able community, and interaction happens through written language text based.	Adept at both text-based communication and sign-to-text communication. Hence, the proposed idea will execute both forms of communication.
User Interaction for SL System	• Limited to understanding and displaying gestures or letters only. • Displaying signs from reading videos.	The proposed idea includes sending SL in real time through the app for practical use in the e-commerce domain for purchasing or inquiring about products.
User Interaction for Chatbot	• Limited to people with no hearing impairments. • Limited to reading video-based SL	The proposed idea is a combined system; hence, every type of user, such as abled and hearing-impaired users, can use it. Real-time SL is included.
e-Commerce Integration	• Based on findings, such a system does not exist; however, there is a need for such a system for digital inclusion.	This is a novel approach to integrate the system with e-commerce websites for better user experience for people with hearing impairments as well as people who use regular languages.
Scope of Application	• SL System Limited to Understanding words • Chatbot limited to understanding text or voice commands only.	Chatbot will be integrated into the SL system, which can understand ASL.
Overall System Limitations	• Limited to either the SL system or only the chatbot model.	A combined framework of a system of SL and chatbot.

3. PROPOSED CONCEPTUAL FRAMEWORK

To improve digital accessibility within the e-commerce domain, we have proposed this novel hybrid conceptual framework of an SL chatbot. The purpose of this framework focuses on allowing a smooth flow of conversation for users who use ASL. This is a multimodal structure that can be used in both text and ASL. This said proposed hybrid framework will have NLP integration with a CNN-based SL recognition module supported by Google's MediaPipe framework backbone structure. The whole proposed framework is designed in a modular way with the help of object-oriented principles for the development method. This allows the system to be flexible for future scalability and robustness.

3.1 Proposed System Architecture

The proposed technical architecture of the framework, as shown in Figure 1, displays a series of steps that form a pipeline to convert raw ASL signs to text and then return e-commerce responses. This whole process pipeline takes a systematic approach for overall data flow, business logic flow, and also system flow to create one unified framework.

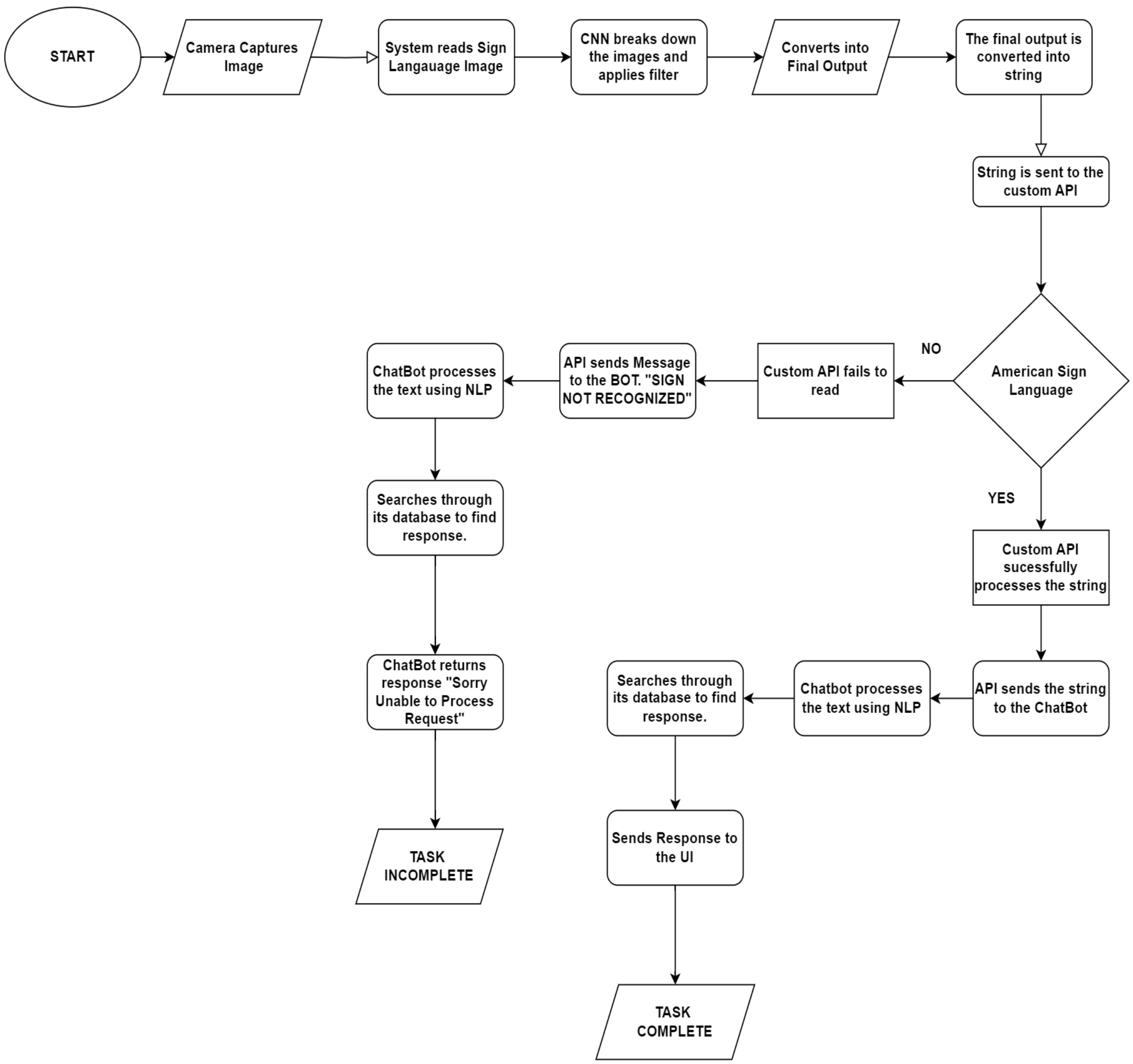

Figure 1. Technical Architecture

The technical process flow architecture in Figure 1 of the SL recognition chatbot system framework consists of a set of systematically arranged steps for the proposed methodology. They are described below:

Input Capture: The process starts by first capturing the users' sign inputs in real-time with the front-end interface using a webcam/camera and sending those captured signs to the next module to be broken down further to extract the signs of ASL.

Preprocessing: Every image that is captured from the real-time input step undergoes a next-level preprocessing process to collect the hand landmarks, and the image is normalized so that the analysis is carried out under the same conditions.

Signs Classification: The processed hand landmarks data taken from the previous step after normalization is sent to the core SL recognition module using the MediaPipe Framework with the core of the CNN algorithm, where it breaks down the image to apply small filters and recognize the hand signs.

Context Tracking and API Integration: Once the text-based input is received, it is then sent to the next module using an API in the form of a JSON payload for further NLP processing. The payload sample is given below:


```
{
  "module": "SL_Module",
  "jsonData": {
    "recognizedText": "A",
    "confidenceScore": 0.985,
  }
```


The recognized text is formatted into a standardized data structure (e.g., JSON) and transmitted via a custom API to the backend.

Chatbot Integration and System Response: The backend receives the text, engages the NLP module to understand intent, and generates an appropriate response.

Performance Evaluation: The framework follows the standard process of measuring accuracy of the detection and responses by evaluating accuracy, precision, recall, and F1-score for the response generation process of the system.

To manage the interaction between the functional units that make up the system, the general architecture of the system is envisioned as a multi-tiered, modular structure. The design ensures a pure separation of concerns, and thus each component is created, tested, and optimized separately. Figure 2 shows a block diagram consisting of the four main modules involved in the framework as well as the overall workflows from a user point of view.

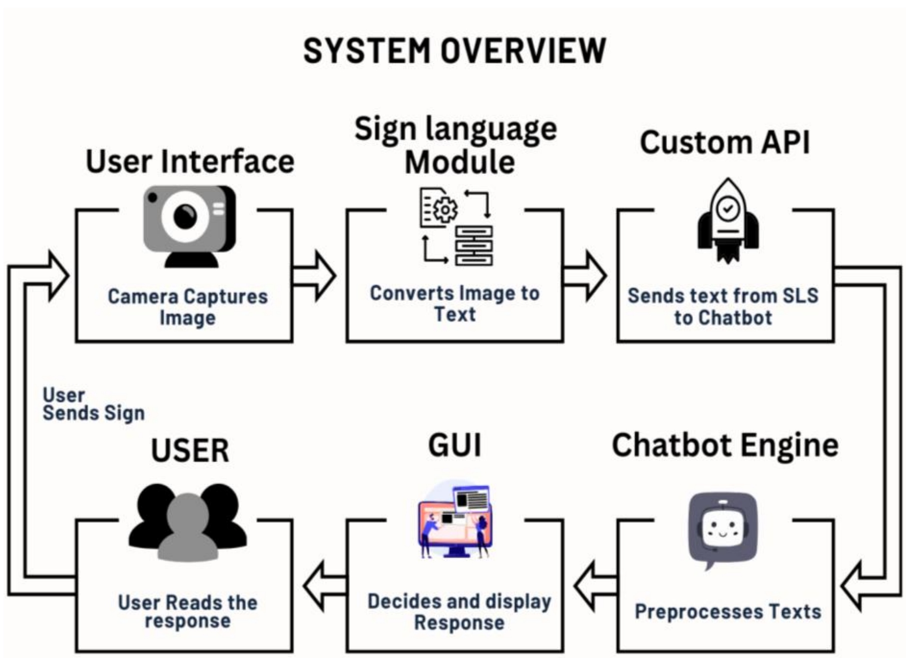

Figure 2. A Block Diagram of System Architecture

3.2 System Integration and Communication Pipeline

1. Frontend User Interface: This is the user interface, which will be the point of interaction for the client to use the system. The user will use the e-commerce front-end UI to provide signs in front of the camera for real-time interaction to query about e-commerce products and services and receive a real-time response for an interactive user experience.
2. SL Recognition Module**:** This is one of the most important components of the overall framework. This module will receive the detected image frames from the UI via HTTP request from the page. Once the frames are sent, the MediaPipe framework in the SL recognition module will find the hand signs for detection and classification to finally convert the frames received into text strings.
3. Backend Server & Chatbot Engine: This is the next important module within the hybrid SL framework. In this module, Flask is the main area of orchestration to support the chatbot engine. In this module we will be using the RESTful API to communicate between the MediaPipe structured SL recognition module and the chatbot engine connected to the UI for the processing of user queries and responses and interfering with the mock database of the e-commerce platform.
4. NLP Module: This is the module we will be using for the semantic analysis of the raw text received. This module will determine the intent and also classify the user's goal to generate a response in the UI.

3.3 Core Technologies

The technological foundation of the proposed framework rests on four synergistic pillars. The SL recognition module serves as the primary sensory input, responsible for the real-time detection and classification of visual gestures. This module interfaces with a robust backend API engine, which acts as the central orchestrator for data transmission and state management. To derive meaning from the converted inputs, the NLP module performs intent analysis, while the principle of multi-modal fusion ensures that both ASL and textual data streams are synthesized seamlessly to maintain a coherent e-commerce dialogue.

3.4 Formalization using Object-Oriented Design

1. To formalize the methodology, we propose a modular, hybrid framework designed for real-time translation of ASL communication into text within the e-commerce chatbot for users. The system architecture presented in Figure 3 shows the framework for the SL recognition hybrid system that is designed using the object-oriented design model [49]. It integrates design entities, properties, and relationships, while incorporating key performance metrics as variables for evaluation. The framework concept model is defined as

   A (SL Framework) = (E, Properties(E), Rel (E)). Where:
2. A: Represents the SL recognition module Framework in a whole manner.
3. E: This is mainly the representation of entities of the framework, which are CameraFeed, HandLandmarks, GestureClassifier, and HistoryBuffer.
4. Properties (E): The properties are mainly the characteristics that each entity will have. For example, CameraFeed entity properties are mainly Resolution and FrameRate. For HandLandmarks, the main properties are based on the data structure, including the 21 key points of the hand joints to detect HandLandmarks.
5. Rel (E): This is the final process to combine the entities and their relationships to determine proper logical workflows.

Figure 3 details the logical data structure of the backend's chatbot engine, which is essential for managing stateful, coherent conversations. This Entity-Relationship Diagram (ERD) illustrates the key entities and their relationships, forming the backbone for the system's conversational memory and e-commerce functionality. The independent and dependent variables are also assigned, and their connections are portrayed to represent the framework design, starting from the first capture to the result of evaluation. This follows the technical process flow diagram based on the referenced paper [50].

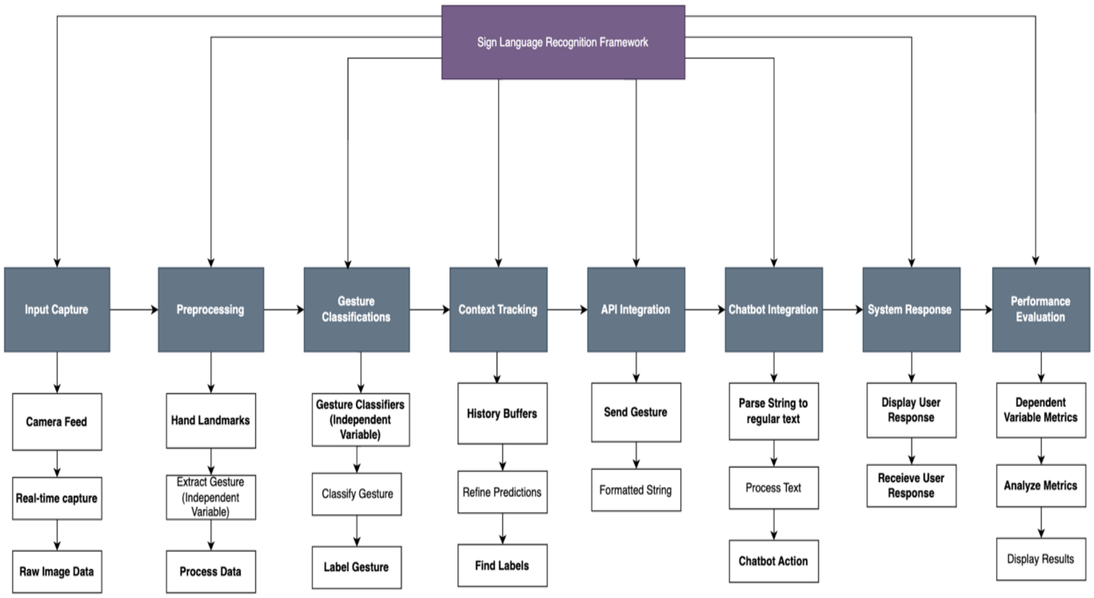

Figure 3. Process Flow Diagram

This integration method confirms that textual and ASL are considered together during user communication by enabling real-time communication within the modules. By explicitly connecting the SL recognition module, NLP, and backend components, the system creates a workflow for multimodal interaction, overcoming the limitations of accessibility in e-commerce.

4. RESEARCH METHODOLOGY

To validate the conceptual hybrid framework proposed in Section 3, the research methodology focuses on the practical implementation of the core SL recognition module and its integration with the backend API. While the conceptual framework encompasses a full e-commerce loop, this methodology specifically targets the development and feasibility testing of the vision-to-text pipeline as the primary enabler of accessibility. In this part of the research, we outlined the practical method of the implementation of the current system prototype, which follows a specific set of formal prototypes to validate the feasibility experiment of the SL recognition module, and we also discussed the challenges that were faced during the development.

4.1 System Development

The initial feasible prototype development was done through a structured multi-stage process. The whole development was done to convert the conceptual framework core module into a preliminary functional testable system to validate the whole framework development capability in future development. All the individual modules were developed to create a full-stage development of the conceptual hybrid framework.

4.1.1 Dataset Preparation

The first step of the development was to collect a dataset and prepare it for model training. The pre-trained framework has been trained by the small dataset curated for better recognition capabilities. A custom dataset was finalized based on making it work as a gesture classifier by annotating it and processing it.

1. Data Source and Refinement: As a foundation for the model training, a dataset of 1500 images were created based on the publicly available dataset from Kaggle. The dataset was the SL MNIST dataset, which consisted of the source images containing the letters and numbers of ASL. Upon collecting this dataset, a thorough cleaning and refinement was performed to remove vague or poorly framed data from the whole sample, and then it was annotated accordingly to collect it as a high-level dataset.
2. Feature Extraction: These 1500 images collected were processed into the SL recognition module using the MediaPipe framework. They were used to train this pre-trained framework to collect the 21 hand landmark coordinates for each image of the dataset. Through this, we collected geometric hand points for the MediaPipe framework. These landmark data were divided into 3 sets during the development process. The training set consisted of 1,200 landmark sets (80%), the validation set consisted of 150 landmark sets (10%), and the testing set was 150 landmark sets (10%).

4.1.2 SL Recognition Module Development

One of the main components of this hybrid system, which was also tested for feasibility, is the SL recognition module. This was built following a hand tracking process pipeline to achieve high performance during real-time recognition for a better user experience during e-commerce shopping experiences by users. The recognition module uses a sequential pipeline to quickly turn raw video frames into semantic class labels. There are three strict steps in the architecture: detection, preprocessing, and classification. The main framework for this step is MediaPipe, which is used to construct a pipeline. This pipeline is made up of two stages: CNN is used to find and evaluate the landmarks, localize them, and find key places [51]. CNNs have many levels, like activation and convolutional layers. Each layer does a different job on the input data, which in this case is the image. Convolutional layers "extract features" for picture categorization. The first layers pull out simple features, and the last levels pull out more complex semantic information [52]. CNNs use several layers of convolutions to find features and identify patterns that are similar in images [53]. This framework guarantees both precision and rapidity in this conceptual advancement for enhanced recognizing skills.

The job of the palm detector is to first find the area of the hand that is captured. Secondly, the Hand Landmark Model guesses where 21 specific hand joints are located from the hand of the captured image. This system uses 2D coordinates (x, y), which means that the feature set is small to reduce the amount of work on the computer must do.

To make sure they are strong against changes in user position, the raw landmarks are pre-processed and normalized. To make translation invariant, you change absolute coordinates to relative coordinates by taking the wrist location landmark 0 away from all the other landmarks. Then, to make sure that the scale is the same, these relative coordinates are normalized by the highest absolute value in the collection, which maps the hand geometry to a standard range of [-1, 1].

The data is turned into a 42-dimensional feature vector (21 points x 2 axes) that is used as input for a multi-layer perceptron built with TensorFlow. The architecture of the classifier is made to be very light so that it can work on a CPU. It includes:

1. Input Layer: 42 units with a dropout layer (rate=0.2) to keep noise from overfitting.
2. Hidden Layers: A completely linked layer with 20 units (with ReLU activation), then a second dropout layer with a rate of 0.4, and then a second dense layer with 10 units of ReLU activation.
3. Output Layer: The last dense layer, which uses SoftMax activation, shows the sign classes.

The model was trained utilizing the Adam optimizer and the sparse categorical crossentropy loss function. The batch size was 128, and early stopping was set to patience = 20 to make sure the model converged without overfitting.

In this final step of the SL recognition module development, we classify the ASL signs based on the geometric data, which we receive from the hand landmark step as 21 landmarks as shown in Figure 4. These data are then passed to the lightweight machine classifier, which has been trained on the 1200 images of the final dataset to recognize spatial patterns from the ASL signs and provide an input.

Figure 4. Real-time Extraction of the 21-point Skeletal Landmarks

4.1.3 Key Steps in Image Recognition

To make sure that the system operates in real-time, the image recognition process follows a specific process flow. It begins with palm detection first, which scans the entire frame to separate the hand since this step is very important for filtering out background noise and focusing computational power where it's needed. Once the hand is located, the landmark detection model starts the next task which is using a regression-based CNN to map the coordinates of 21 specific joints and knuckles of the hand image. These landmarks are then passed into the final classification stage, which analyses the spatial arrangement of the points to identify the correct ASL letter or digit. Table 3 explains the conceptual architecture of the MediaPipe sign detection pipeline.

Table 3. Conceptual Architecture of the MediaPipe Hands CNN Pipeline

Stage	Core CNN Model	Primary Function	Input	Output
1	Palm Detector CNN	Object Detection	Full Image Frame 224x224x3	Bounding Box of the Palm
2	Hand Landmark CNN	Keypoint Regression	Cropped Hand Image (from Stage 1)	21 Hand Landmark Coordinates (x, y, z)

4.1.4 Module Development

The primary feasibility study for this conceptual framework was done mainly on the SL recognition module; however, the NLP module was also designed for better understanding the whole framework goal on a system level. The main objective of this module is to receive the string formatted as JSON from the SL recognition module once recognition is successful. Once received, the NLP module understands the intent of the strings and based on that, provides a relevant response within the e-commerce platform.

To achieve the whole structure of the framework in this module, we implemented the NLP pipeline using NLTK and Scikit-learn for text preprocessing, tokenization based on the input query, feature extraction using Term Frequency-Inverse Document Frequency, and also classifying the intent using the Support Vector Machine mechanism.

4.1.5 Backend Development and API Design

The backend is the main orchestrator that communicates between each layer of the system and the frontend. It carries the whole cycle of the workflow and back to the user.

The backend was designed to be lightweight; hence, Flask was chosen for development. A RESTful API was designed to be added as the primary endpoint in the system from the front-end for accepting the images as a Base64 encoded string. This was done using the POST request method for the API. The APIs serve as interfaces for reusable programs. They are not standalone software entities; rather, they are packaged and built with software libraries, frameworks, or web services to support them [54].

The backend supports the orchestration of the workflow, starting from image decoding, then passing it to the SL recognition module to determine the hand landmark extraction and classification. After that, the backend API forwards it to the NLP module to finally finalize the JSON as a final string and send it back to the UI as a response to the client query.

4.2 Technologies and Tools Used

This initial prototype is built and tested using a MacBook Pro by using its standard camera input at 30 frames per second. By making use of the macOS environment, we set up the core logic using Python version 3.8 for the backend, while React handled the development of user-facing interface. The computer vision pipeline pairs Google's MediaPipe, which manages skeletal tracking, with OpenCV for capturing and processing live video frames. On the machine learning side for the chatbot interaction, we used a mix of NLTK, Scikit-learn, and TensorFlow to handle tasks ranging from feature normalization to the actual classification of signs and response generation as well. Finally, a Flask web server acts as the bridge, connecting these vision components to the user interface and managing the data flow between them.

4.3 Experimental Feasibility Study

To validate the conceptual framework at the preliminary level, a feasibility study was conducted for real-time performance monitoring. Participants were recruited to test and validate the system, and the ethical code was approved by the Research Ethics Committee at Multimedia University (Approval Code: EA0532024). A total of 8 participants helped perform this testing after initial development in real time in front of the camera with various lighting, clothing, and background conditions.

4.4 Developmental Challenges and Solutions

During the development phase, several key challenges were encountered, for which solutions have been provided as well below:

Challenge 1: To ensure that performance is real-time, a high frame rate for the smooth user experience was necessary to accomplish.

Solution: This challenge was addressed by using the core technology of the MediaPipe framework, which is a two-stage-based CNN detector, and it is lightweight enough to be used on standard hardware, which allows users to use the system without any extra hassle.

Challenge 2: For a smooth workflow of the whole hybrid system, it was important that the communication between the frontend UI and the backend have a good architecture so that the whole process remains error-free.

Solution: The development of the RESTful API and the use of the JSON format to communicate data between the Flask backend and the client-facing frontend made sure that all modules interacted with each other smoothly without having any issues.

5. Evaluation

5.1 Results

At this initial stage, to justify the viability of the system, a minimal feasibility implementation of the framework is performed for the core module of the SL recognition system. The SL Module processes the gestures read from the input of ASL using MediaPipe and OpenCV. It allows the processing of images and videos to recognize gestures and writings.

In our study, we evaluated the SL chatbot's ability to recognize signs in terms of letters and digits. Performance was quantified using standard evaluation metrics, including accuracy, precision, recall, and F1-score. For this preliminary verification, accuracy is defined as a binary per-frame metric: a recognition instance is considered 'accurate' only if the predicted class label matches the ground truth exactly. To ensure fairness, a weighted average was applied so that categories with more test samples had a proportional impact on the final score. Furthermore, in terms of real-time performance, the reported processing speed reflects the end-to-end system latency, encompassing the entire pipeline from image capture and MediaPipe landmark extraction to the CNN-based classification and API response generation.

The model performed very well quantitatively, achieving 98% overall accuracy in recognition of signs and letters tested. These results were further validated based on the help of the 8 participants to support the evaluation using a small qualitative test. Table 4 presents the results more clearly, based on the support values and calculations.

Table 4. Accuracy Rate for SL Module Detection

	Precision	Recall	F1-Score	Accuracy	Support
A	0.99 (99%)	1 (100%)	0.99 (99%)	0.99 (99%)	402
B	0.98 (98%)	0.95 (95%)	0.96 (96%)	0.96 (96%)	366
1	0.95 (95%)	0.98 (98%)	0.96 (96%)	0.97 (97%)	343
9	1 (100%)	0.99 (99%)	0.96 (96%)	0.99 (99%)	86
Average	N/A	N/A	0.97 (97%)	0.98 (98%)	1197
Macro Avg	0.98 (98%)	0.98 (98%)	0.98 (98%)	0.98 (98%)	1197
Weighted Avg	0.98 (98%)	0.98 (98%)	0.98 (98%)	0.98 (98%)	1197

To supplement the initial preliminary quantitative evaluation, we carried out a small-scale user study in which 8 users performed SL gestures in real time. They had given verbal and written consent for participating in the experiment. The consents were approved by the Research Ethics Committee at Multimedia University. The approval code is EA0532024. During the experiment, the participants were asked to wear distinct types of clothing to test the system for better calculation of accuracy. They were given letters and digits of the ASL to perform the signs for preliminary testing.

Every participant expressed great satisfaction with the SL recognition capabilities and how well it can detect the signs. This sets an overall base point for the viability of this conceptual framework.

However, there are a few limitations to consider. Future research should involve a larger and more diverse dataset for users to try out sentence-level SL recognition to capture a wider range of interaction patterns. To address these challenges, the eventual goal is to enhance the CNN architecture with a self-supervised pre-training process and increase the training dataset, especially complex signs. Additionally, we will implement offline reinforcement learning, and an ongoing A/B testing method will be implemented to continuously keep updating the e-commerce experience based on user acceptance.

Unlike other SL systems based on research, the conceptual SL recognition chatbot combines SL recognition and NLP to offer a real-time multi-modal conversation and context-based e-commerce framework for better accessibility for hearing-impaired users. This multimodal method creates a more flexible e-commerce experience in real time with good accuracy.

5.2 User Feedback

We carried out a user study with 8 participants, gathering feedback through a 5-point Likert scale to evaluate different parts of the system's performance. Participants were asked questions to rate usability, accuracy, and overall satisfaction.

1. "How well did the system recognize the signs? Could you describe your experience with its accuracy and speed?"
   - This question was designed to assess how well the system processes both signs to generate an accurate response in real-time at a good speed.
2. "What did you think of the visual output on the screen, like the skeletal overlay of your hand?"
   - This question helped understand the HCI part of the preliminary prototype tested to showcase if the users were satisfied with the real-time hand tracking and other aspects of the system in the core module.
3. "Based on your experience today with how it recognizes individual letters, can you imagine using a more advanced version of this technology for a task like e-commerce?"
   - This question was designed to evaluate the overall user experience of using the SL module in terms of recognition and usability before we move on to the final future development of the system.

5.3 Quantitative Results

The feedback results are summarized below based on a 5-scale set by us according to the Likert scale rating as shown in Table 5:

1. Effectiveness and Understanding: After trying the system, participants gave the system an average score of 4.7 out of 5, showing that most users found the preliminary system effective in understanding ASL letters.
2. Accuracy of Recommendation: Based on user testing, the system received an average rating of 4.5 out of 5 for the recognition of signs.
3. Overall: In an overall system-level usage, most users were satisfied and gave an average score of 4.6 out of 5, indicating a positive experience during their recognition process and also validating further future usage once the whole system is complete.

Table 5. Likert Scale Ratings

Question	Average Rating	Interpretation
Effectiveness in understanding image and text input	4.5/5	Most users found the system effective at processing SL accurately with good speed.
Satisfaction with the visual output and input	4.6/5	High user satisfaction with the system and its interface.
Likelihood of recommending the system to others after final version	4.4/5	Users were generally satisfied with sign recognition but requested sentence-level recognition.

5.4 Qualitative Feedback

Feedback based on user acceptance is shown in Table 6.

Table 6. User Response

Feedback Area	Participants' Ratings	Number of Participants
Ease of Use	4.7/5	8 participants
Recommendation Accuracy	4.5/5	8 participants
Overall Satisfaction	4.6/5	8 participants

1. Real-time SL module: We tested the system with 8 participants for the real-time recognition module. Their feedback was based on the ASL input, and they mentioned the response generated was faster with both low and high lighting due to real-time recognition.
2. Value of Intuitive Visual Feedback: The other main feedback was on the skeletal overlay of the framework for a better intuitive response. This was critical because users said it helped them have confidence that the sign was being seen correctly and the tracking was real-time. Which in turn made the whole recognition and tracking process much easier and more intuitive for them.

The feedback that we gathered from the users for the study shows that users are very satisfied with the preliminary module of the conceptual system, especially with how easy it is to use. The accuracy and overall experience have also been praised. Participants liked the idea of the multimodal approach method combining text and SL input. According to the initial users, it is intuitive and offers communication accessibility, which made the experience easy.

5.5 Robustness and Error Analysis

A qualitative error analysis was done throughout the feasibility study to see how reliable the system was beyond only accuracy measurements. This was done under different environmental conditions.

1. Environmental Robustness: The system was very resilient to complicated backgrounds because of the first step in the vision pipeline, which cropped the ROI. However, lighting circumstances were very important. In low-light situations, which are less than 50 lux, confidence in landmark detection sometimes dipped below the threshold, which caused momentary tracking loss. On the other hand, changing the colour of the clothes had little effect on performance, which proved that the skeletal tracking method is more reliable than pixel-based colour segmentation methods.

2. Class-Specific Confusion: The overall accuracy was 98%, but an error analysis showed that not all the wrong classifications were evenly spread across. Most of the false negatives happened when the signs had a lot of finger occlusions or seemed very similar geometrically for letters such as M and N. The difference between the precision (0.95) and recall (0.98) seen in the numeric classes in Table 4 is due to these specific edge circumstances.

6. CONCLUSION

This research shows that a hybrid e-commerce framework can effectively remove the accessibility gap for the Deaf and Hard-of-Hearing community by giving them more accessibility. By combining a CNN-based SL recognition module with a conversational chatbot interface, the study shows that a standard, webcam-based setup is more than enough to interpret user intent without the need for specialized hardware from the consumer's end. A key achievement of this research is the validation of a lightweight MLP classifier, which reached a weighted average accuracy of 98% for static ASL gestures. Beyond the raw metrics, this framework also provides a scalable model for the future for inclusive online retail, reducing the need for text-based communication and helping users navigate digital marketplaces with more independence.

However, the study does have limitations. The current validation was restricted to a small group of eight participants and focused only on static signs, which does not capture the fluidity and complexity of natural conversation. Qualitative analysis also revealed that the system is sensitive to environmental conditions such as performance dropping slightly in low-light settings, specifically under 50 lux or when fingers were not clearly visible, hiding key geometric landmarks that are needed for detection.

In future development, the aim of this research is to modify this framework from recognizing individual characters to understanding full sentences of SL recognition. Future development will include stronger object detection backbones to track continuous signing while using Large Language Models (LLMs) [45] to convert those raw inputs into understandable, domain-specific text and detection of the e-commerce intent. We also plan to highly stress-test the system across different lighting and backgrounds to ensure reliability in real-world user-dependent scenarios. Finally, to move beyond technical feasibility and prove practical usability, we will conduct extensive usability trials using established HCI metrics. By measuring task completion times and perceived cognitive load, we aim to demonstrate that this multimodal approach genuinely improves the shopping experience and reduces the mental effort required for users.

ACKNOWLEDGEMENT

We would like to express our gratitude to the people who contributed to this research.

FUNDING STATEMENT

The authors received no funding from any party for the research and publication of this article.

AUTHOR CONTRIBUTIONS

Salma Jahan Nisha: Conceptualization, Data Curation, Methodology, Investigation, Writing – Original Draft Preparation;
Nabhan Salih: Methodology, Supervision, Writing —Review & Editing;
Wan-Noorshahida Mohd-Isa: Methodology, Supervision, Validation, Writing – Review & Editing.

CONFLICT OF INTERESTS

No conflict of interest to disclose.

ETHICS STATEMENTS

Our publication ethics follow The Committee of Publication Ethics (COPE) guideline. https://publicationethics.org/.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon request.

REFERENCES

[1] D. Bragg *et al*., "Sign language recognition, generation, and translation: An interdisciplinary perspective," in *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility*, New York, NY, USA, pp. 16–31, 2019, doi:10.1145/3308561.3353774.

[2] M. Al-Qurishi, T. Khalid, and R. Souissi, "Deep learning for sign language recognition: Current techniques, benchmarks, and open issues," *IEEE Access*, vol. 9, pp. 126917–126951, 2021, doi:10.1109/ACCESS.2021.3110912.

[3] D. W. Harahap, and L. Fitria, “Aplikasi chatbot berbasis web menggunakan metode dialogflow,” *J-Icom - Jurnal Informatika Dan Teknologi Komputer*, vol. 1, no. 1, pp. 6–13, 2020, doi:10.33059/j-icom.v1i1.2796.

[4] S. Richardson, F. Jovanka, P. K. Zabrina, M. P. Adiati, and D. Rosman, “The consequences of digital concierge chatbots acceptance on hotel guest experience and satisfaction at IHG hotel and resort,” in *2023 International Conference on Digital Applications, Transformation & Economy (ICDATE)*, Miri, Malaysia, pp. 1–5, 2023, doi:10.1109/ICDATE58146.2023.10248751.

[5] N. Ketkar, and J. Moolayil, “Convolutional neural networks,” in Deep Learning with Python: Learn best practices of deep learning models with PyTorch, *Apress*: Berkeley, CA, USA, pp. 197–242, 2021, doi:10.1007/978-1-4842-5364-9_6.

[6] T. Zemcik, “A brief history of chatbots,” *DEStech Transactions on Computer Science and Engineering*, 2019, doi:10.12783/dtcse/aicae2019/31439.

[7] K. Klein, and L. F. Martinez, “The impact of anthropomorphism on customer satisfaction in chatbot commerce: An experimental study in the food sector,” *Electronic Commerce Research*, vol. 23, no. 4, pp. 2789–2825, 2023, doi:10.1007/s10660-022-09562-8.

[8] M. Li, and R. Wang, “Chatbots in e-Commerce: The effect of chatbot language style on customers’ continuance usage intention and attitude toward brand,” *Journal of Retailing and Consumer Services*, vol. 71, Art. no. 103209, 2023, doi:10.1016/j.jretconser.2022.103209.

[9] P. Sindhu, and K. Bharti, “Influence of chatbots on purchase intention in social commerce,” *Behaviour & Information Technology*, pp. 1–22, 2023, doi:10.1080/0144929x.2022.2163188.

[10] J. M. Solis-Quispe, K. M. Quico-Cauti, and W. Ugarte, “Chatbot to simplify customer interaction in e-Commerce channels of retail companies,” in *HCI International 2021*, pp. 561–570, 2021, doi:10.1007/978-3-030-68285-9_52.

[11] M. M. Khan, “Development of an e-commerce sales chatbot,” in *2020 IEEE 17th International Conference on Smart Communities: Improving Quality of Life Using ICT, IoT and AI (HONET)*, Charlotte, NC, USA, pp. 173–176, 2020, doi:10.1109/HONET50430.2020.9322667.

[12] H. Mostaq *et al*., “Research and development of an e-commerce with sales chatbot,” in *2022 IEEE World AI IoT Congress (AIIoT)*, pp. 557–564, 2022, doi:10.1109/aiiot54504.2022.9817272.

[13] V. Oguntosin, and A. Olomo, “Development of an e-commerce chatbot for a university shopping mall,” *Applied Computational Intelligence and Soft Computing*, vol. 2021, no. 1, 2021, doi:10.1155/2021/6630326.

[14] A. Nursetyo, D. R. I. M. Setiadi, and E. R. Subhiyakto, “Smart chatbot system for e-commerce assistance based on AIML,” in *2018 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, Yogyakarta, Indonesia, pp. 641–645, 2018, doi:10.1109/ISRITI.2018.8864349.

[15] J.-S. Chen, T.-T. Le, and D. Florence, “Usability and responsiveness of artificial intelligence chatbot on online customer experience in e-retailing,” *International Journal of Retail & Distribution Management*, vol. 49, no. 11, pp. 1512–1531, 2021, doi:10.1108/IJRDM-08-2020-0312.

[16] R. Parkar *et al*., “AI and web-based interactive college enquiry chatbot,” in *2021 13th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, Pitesti, Romania, pp. 1–5, 2021, doi:10.1109/ECAI52376.2021.9515065.

[17] S. Meshram *et al*., “College enquiry chatbot using Rasa framework,” in *2021 Asian Conference on Innovation Technology (ASIANCON)*, Pune, India, pp. 1–8, 2021, doi:10.1109/ASIANCON51346.2021.9544650.

[18] M. G. C. P. Srivastava, A. Chakraborty, S. Ghosh, and H. Raj, “Development of information technology telecom chatbot: An artificial intelligence and machine learning approach,” in *2021 2nd International Conference on Intelligent Engineering and Management (ICIEM)*, London, UK, pp. 216–221, 2021, doi:10.1109/ICIEM51511.2021.9445354.

[19] World Health Organization, “Deafness and hearing loss,” Mar. 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss.

[20] A. Pardasani, A. K. Sharma, S. Banerjee, V. Garg, and D. S. Roy, “Enhancing the ability to communicate by synthesizing American sign language using image recognition in a chatbot for differently abled,” *in 2018 7th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, Noida, India, pp. 529–532, 2018, doi:10.1109/ICRITO.2018.8748590.

[21] C. Apuzzo, and G. Burresi, “Designing accessible chatbots for deaf people,” in *2022 11th Mediterranean Conference on Embedded Computing (MECO)*, pp. 1–4, 2022, doi:10.1109/MECO55406.2022.9797093.

[22] M. Mohandes, M. Deriche, and J. Liu, “Image-based and sensor-based approaches to Arabic sign language recognition,” *IEEE Transaction on Human-Machine System*, vol. 44, no. 4, pp. 551–557, 2014, doi:10.1109/THMS.2014.2318280.

[23] R. Sreemathy, M. Turuk, J. Jagdale, A. Agarwal, and V. Kumar, “Indian sign language interpretation using convolutional neural networks,” in *2023 10th International Conference on Signal Processing and Integrated Networks (SPIN)*, Noida, India, pp. 789–794, 2023, doi:10.1109/SPIN57001.2023.10116204.

[24] A. Abul *et al*., “CNN based feature extraction and classification for sign language,” *Multimedia Tools and Applications*, vol. 80, no. 2, pp. 3051–3069, 2021, doi:10.1007/S11042-020-09829-Y.

[25] N. Singla, M. Taneja, N. Goyal, and R. Jindal, “Feature fusion and multi-stream CNNs for scale-adaptive multimodal sign language recognition,” in *2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Coimbatore, India, pp. 1266–1273, 2023, doi:10.1109/ICACCS57279.2023.10112915.

[26] K. Bantupalli, and Y. Xie, “American sign language recognition using deep learning and computer vision,” in *2018 IEEE International Conference on Big Data*, Seattle, WA, USA, pp. 4896–4899, 2018, doi:10.1109/BigData.2018.8622141.

[27] D. Tasmere, and B. Ahmed, “Hand gesture recognition for Bangla sign language using deep convolution neural network,” in *2020 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI),* Dhaka, Bangladesh, pp. 1–5, 2020, doi:10.1109/STI50764.2020.9350438.

[28] T. F. Dima, and M. E. Ahmed, “Using YOLOv5 algorithm to detect and recognize American sign language,” in *2021 International Conference on Information Technology (ICIT)*, Amman, Jordan, pp. 603–607, 2021, doi:10.1109/ICIT52682.2021.9491642.

[29] A. Sharma, A. Mittal, S. Singh, and V. Awatramani, “Hand gesture recognition using image processing and feature extraction techniques,” *Procedia Computer Science*, vol. 173, pp. 181–190, 2020, doi:10.1016/j.procs.2020.06.022.

[30] W. Feng, Z. Zhang, T. He, and C. Lee, “AI enabled sign language recognition and VR space bidirectional communication using triboelectric smart glove,” *Nature Communications.*, vol. 12, no. 1, p. 5378, 2021, doi:10.1038/S41467-021-25637-W.

[31] M. M. Damaneh, F. Mohanna, and P. Jafari, “Static hand gesture recognition in sign language based on convolutional neural network with feature extraction method using ORB descriptor and Gabor filter,” *Expert Systems with Applications*, vol. 211, p. 118559, 2023, doi:10.1016/j.eswa.2022.118559.

[32] A. Wadhawan and P. Kumar, “Deep learning-based sign language recognition system for static signs,” *Neural Computing and Applications*, vol. 32, no. 18, pp. 7957–7968, 2020, doi:10.1007/s00521-019-04691-y.

[33] M. M. Rahman *et al*., “A new benchmark on American sign language recognition using convolutional neural network,” in *2019 International Conference on Sustainable Technologies for Industry 4.0 (STI),* Dhaka, Bangladesh, pp. 1–6, 2019, doi:10.1109/STI47673.2019.9067974.

[34] M. Al-Hammadi *et al*., “Hand gesture recognition for sign language using 3DCNN,” *IEEE Access*, vol. 8, pp. 79491–79509, 2020, doi:10.1109/ACCESS.2020.2990434.

[35] Q. M. Areeb, M. Maryam, M. Nadeem, R. Alroobaea, and F. Anwer, “Helping hearing-impaired in emergency situations: A deep learning-based approach,” *IEEE Access*, vol. 10, pp. 8502–8517, 2022, doi:10.1109/ACCESS.2022.3142918.

[36] P. A. Rodriguez-Correa, A. Valencia-Arias, O. N. Patino-Toro, Y. Oblitas Diaz, and R. Teodori De la Puente, “Benefits and development of assistive technologies for deaf people's communication: A systematic review,” *Frontiers in Education*, vol. 8, 2023, doi:10.3389/feduc.2023.1112440.

[37] F. H. F. Botelho, “Accessibility to digital technology: Virtual barriers, real opportunities,” *Assistive Technology*, vol. 33, no. sup1, pp. 27–34, 2021, doi:10.1080/10400435.2021.1945705.

[38] S. Svensson *et al*., “The emerging ‘right to repair’ legislation in the EU and the US,” in *Proceedings of the Going green - Care Innovation 2018*, Vienna, Austria, 2018. https://api.semanticscholar.org/CorpusID:211235579.

[39] G. Astbrink, and W. Tibben, “The role of public procurement in improving accessibility to ICT,” *Telecommunications Journal of Australia*, vol. 63, no. 2, 2013, doi:10.7790/tja.v63i2.412.

[40] A. Foley, and B. A. Ferri, “Technology for people, not disabilities: Ensuring access and inclusion,” *Journal of Research in Special Educational Needs*, vol. 12, no. 4, pp. 192–200, 2012, doi:10.1111/j.1471-3802.2011.01230.x.

[41] D. T. S. Warnars, L. P. Putra, H. L. H. S. Warnars, and W. H. Utomo, “Intelligent e-commerce for special needs,” in *2019 7th International Conference on Cyber and IT Service Management (CITSM)*, Jakarta, Indonesia, pp. 1–5, 2019, doi:10.1109/CITSM47753.2019.8965341.

[42] G. M. Carmo, D. B. Paiva, and M. I. Cagnin, “Incorporating accessibility in web interface development for deaf users,” in *Proceedings of the XX Brazilian Symposium on Software Quality (SBQS)*, New York, NY, USA, pp. 1–10, 2021, doi:10.1145/3493244.3493246.

[43] N. Singla, “American sign language letter recognition from images using CNN,” in *2023 2nd International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT),* Trichirappalli, India, pp. 1–9, 2023, doi:10.1109/ICEEICT56924.2023.10156922.

[44] S. Janagiraman, P. Mithun, J. Sandhiya, and S. Priyanka, “Enabling inclusive communication: A sign language chatbot for diverse audiences,” in *2025 International Conference on Automation and Computation (AUTOCOM)*, IEEE, pp. 33–39, 2025, doi:10.1109/AUTOCOM64127.2025.10957203.

[45] F. Fang, Y. A, and S. Wang, “SignLLM: Sign language production large language models,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pp. 2500–2509, 2025, doi:10.48550/arXiv.2405.10718.

[46] R. Zabin, M. R. T. Hossain, K. F. Haque, and K. M. Rumman, “SignNet-Nano: Efficient sign language recognition for real-time edge deployment,” in *2025 2nd International Conference on Computing, Applications and Systems (COMPAS)*, Kushtia, Bangladesh, Oct. 2025, doi:10.13140/RG.2.2.35509.54249.

[47] S. Mukherjee, A. Chatterjee, K. Chandrakar, and S. Saxena, “Artificial Intelligence powered chatbots for the speech disabled and hearing impaired using Feedforward Neural Network,” in *2024 1st International Conference on Advances in Computing, Communication and Networking (ICAC2N)*, Greater Noida, India, pp. 1071–1075, 2024, doi:10.1109/ICAC2N63387.2024.10894803.

[48] M. Hassan, K. Assaleh, and T. Shanableh, “User-dependent sign language recognition using motion detection,” in *2016 International Conference on Computational Science and Computational Intelligence (CSCI),* Las Vegas, NV, USA, pp. 852–856, 2016, doi:10.1109/CSCI.2016.0165.

[49] C. Serban, and F. Bota, “A framework for software fault prediction using neural networks,” in *Modelling and Development of Intelligent Systems (MDIS 2019)*, D. Simian and L. Stoica, Eds. Cham: Springer, 2020, pp. 137–148, doi:10.1007/978-3-030-39237-6_12.

[50] M. Tomanek, R. Cermak, and Z. Smutny, “A framework for web development projects based on project management and agile development principles,” *10th European Conference on Management Leadership and Governance (ECMLG)*, 2014, doi:10.13140/2.1.1262.4165.

[51] W. Utomo, Y. Suhanda, H. Ar-Rasyid, and A. Dharmalau, “Indonesian language sign detection using Mediapipe with Long Short-Term Memory (LSTM) algorithm,” *Journal of Informatics and Web Engineering*, vol. 4, no. 3, pp. 245–258, 2025, doi:10.33093/jiwe.2025.4.3.15.

[52] Z. J. Wang *et al*., “CNN explainer: Learning convolutional neural networks with interactive visualization,” in *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 2, pp. 1396–1406, 2021, doi:10.1109/TVCG.2020.3030418.

[53] J.-R. Lee, K.-W. Ng, and Y.-J. Yoong, “Face and facial expressions recognition system for blind people using ResNet50 architecture and CNN,” *Journal of Informatics and Web Engineering*, vol. 2, no. 2, pp. 284–298, 2023, doi:10.33093/jiwe.2023.2.2.20

[54] M. Lamothe, Y.-G. Gueheneuc, and W. Shang, “A systematic review of API evolution literature,” *ACM Computing Surveys*, vol. 54, no. 8, p. 171, 2021, doi:10.1145/3470133.

BIOGRAPHIES OF AUTHORS

	**Salma Jahan Nisha** is currently pursuing her Master of Information Technology (Research) at Multimedia University, Malaysia. She holds a bachelor’s degree in software engineering with multimedia and is a software engineer with professional experience in web development and project management. Her research focuses on *sign language chatbots in the e-commerce domain.* She has delivered technology projects in education, including the Training Management Portal and the Brain-Based Learning application, and also taught a UX/UI course as a guest lecturer in Singapore for Educare Global Academy. Her research interests include sign language recognition, human–computer interaction, NLP, and real-time AI systems. She recently presented her work *“Sentence-based Sign Language Recognition using CNN and NLP”* at MECON 2025. She can be contacted by email at salmajn359@gmail.com or 1221404902@student.mmu.edu.my.
	**Dr. Nbhan D. Salih** is currently a lecturer at Ohio State University, USA, with prior academic experience as an assistant professor in the Faculty of Computing and Informatics at Multimedia University (MMU), Malaysia. He holds a Ph.D. in Engineering, and his research interests lie at the intersection of computer vision, digital image processing, and machine learning. Dr. Salih has been actively involved in advancing technologies that enable intelligent image analysis and pattern recognition, contributing to both the academic and applied aspects of visual computing. He can be contacted at email: salih.43@osu.edu.

	**Dr. Wan-Noorshahida Mohd-Isa** is currently an Assistant Professor at the Faculty of Computing and Informatics, Multimedia University, Cyberjaya campus, Malaysia. She is also the chairperson for the Research Centre for Image and Vision Computing of Multimedia University. She is an engineering alumnus of Vanderbilt University, USA with a bachelor's degree in electrical engineering (magna cum laude) who then ventured into the computing field. She completed her master's degree at the University of Southampton in Great Britain, and she went on to pursue her doctorate degree at Multimedia University. She was a laureate of the France-Malaysia Collaboration Programme for Joint Research 2022 (MyTIGER 2022), which is a research program under the Embassy of France to Malaysia. Her specific research area is on visual analytics, where she designs and develops machine learning models for videos and images. She can be contacted by email at wan.noorshahida.isa@mmu.edu.my.