DES646: End-Term Project Report

AI-Vastra

A Multimodal AI Assistant for E-Commerce Trust and Fit

Team Members:

Ayushmaan Singh [220276]
Prathamesh Baviskar [220285]
Dheeraj Dagar [220355]
Rishith [240484]
Rajesh Inapakurti [240463]

Indian Institute of Technology Kanpur

November 2025

Project Links:

GitHub Repository

Abstract

AI-Vastra is a prototype web application designed to enhance **trust**, **accessibility**, and **visualization** for new online shoppers. The system integrates multiple AI/ML APIs into a unified, multimodal experience. It aims to solve key trust and visualization challenges by allowing users to interact via voice or text, receive AI-generated product insights, and visualize products using **Virtual Try-On (VTO)** and **AR-based scale estimation** tools. The application integrates advanced Large Language Models (LLMs), speech recognition (ASR/TTS), and computer vision APIs in one seamless interface.

Keywords: Human-Computer Interaction, E-Commerce, Generative AI, Multimodal Interfaces, Virtual Try-On (VTO), Augmented Reality (AR), Trust in AI, Accessibility.

Contents

1 Introduction

E-commerce productivity, especially for new-to-internet users, is often hampered by a lack of trust, difficulty in visualizing products, and accessibility barriers. Language and low digital literacy can aggravate this challenge. AI-Vastra addresses this by providing a digital shopping assistant that is accessible, multimodal, and visualization-driven. It empowers users to get help in their preferred language (English/Hindi), interact hands-free via voice, and, most importantly, visualize product fit and scale through integrated AR and VTO tools, bridging the "trust gap."

2 Objectives and Motivation

- **Implement Multimodal Interaction:** Use Azure Cognitive Services for a voice-based, hands-free shopping assistant.
- **Enable AI-Powered Insights:** Use LLMs (Perplexity) to generate answers and summarize product insights.
- **Integrate Visualization Tools:** Combine a Streamlit-based **Virtual Try-On (VTO)** and an **AR Size Visualizer** (ClipDrop API).
- **Deliver a Cohesive, Multilingual Prototype:** Seamlessly connect all features within a single web interface supporting English and Hindi.

3 Problem Statement

Despite the growth of e-commerce, many users struggle with key "trust" questions: "Will this fit me?", "Is this review accurate?", "Is this product a high-quality item?". Standard e-commerce portals lack conversational support, have no easy way to visualize a product's true scale or fit, and require users to sift through hundreds of text-based reviews. There is a critical need for an intelligent assistant that is approachable (multimodal, multilingual) and provides visualization tools to democratize access to trustworthy e-commerce.

4 System Overview

AI-Vastra is a client-side Single Page Application (SPA) prototype. It is architected for minimal backend dependence, instead relying on rich front-end logic and external API integrations. It brings together a host of features:

4.1 Major Features

1. **Multimodal & Multilingual Assistant:** Robust interface (via a floating action button) for text and voice-based queries. Supports Speech-to-Speech (S2S) conversation using Azure STT/TTS and provides AI-powered answers from Perplexity.
2. **AR-Lite Visualization Suite:**
    - **Virtual Try-On (VTO):** An integrated Streamlit app to visualize garment fit using 2D pose estimation.
    - **AR Size Visualizer:** A tool to display a product's true scale in the user's environment (or next to a hand image).

3. **Static Trust Signal:** A "Verified" badge on select product cards to visually signal trust, quality, and authenticity.
4. **AI-Powered Review Summarization:** The LLM assistant can answer user queries by drawing context from product details and reviews.
5. **Language Switcher:** Dynamic text and voice translation between English and Hindi.

5 Technical Architecture

Table 1: AI-Vastra Technology Stack

Component	Technology / API
Frontend	HTML5, CSS3, Vanilla JavaScript
Speech AI	Azure Cognitive Services (Speech-to-Text / Text-to-Speech)
LLM Backend	Perplexity API
Virtual Try-On	Streamlit App (Pose Estimation + Garment Warping)
AR Size Visualizer	ClipDrop Computer Vision API / Netlify App
Language Support	JS-based module for multilingual UI switching

5.1 Frontend

- HTML5, CSS3, Vanilla JavaScript.
- SPA-like architecture (toggling between "Home" and "Product Detail" views).
- Responsive design for desktop and mobile.

5.2 Backend and API Integrations

- **Perplexity API:** For web-indexed, context-aware LLM text queries.
- **Azure Cognitive Services:** Speech-to-text and text-to-speech for multiple languages (English, Hindi).
- **Streamlit (External):** Hosts the VTO module.
- **Netlify/ClipDrop (External):** Hosts the AR Size Visualizer.

5.3 Client-Side Logic

- State management for the chat modal.
- DOM manipulation for switching between product and home pages.
- Voice dialog controls managing Azure ASR/TTS SDK.
- Language and AI model switching mechanics.

6 System Pipeline and Workflow

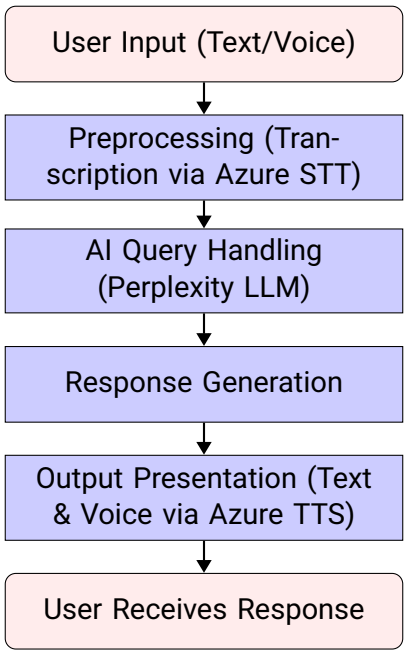

Figure 1: AI-Vastra System Pipeline

Description: The pipeline describes how user inputs (text or voice) are pre-processed (transcribed if voice), passed to the Perplexity LLM for a response, and then formatted for presentation (either as text or synthesized back into voice).

7 User Query to AI Response Flowchart

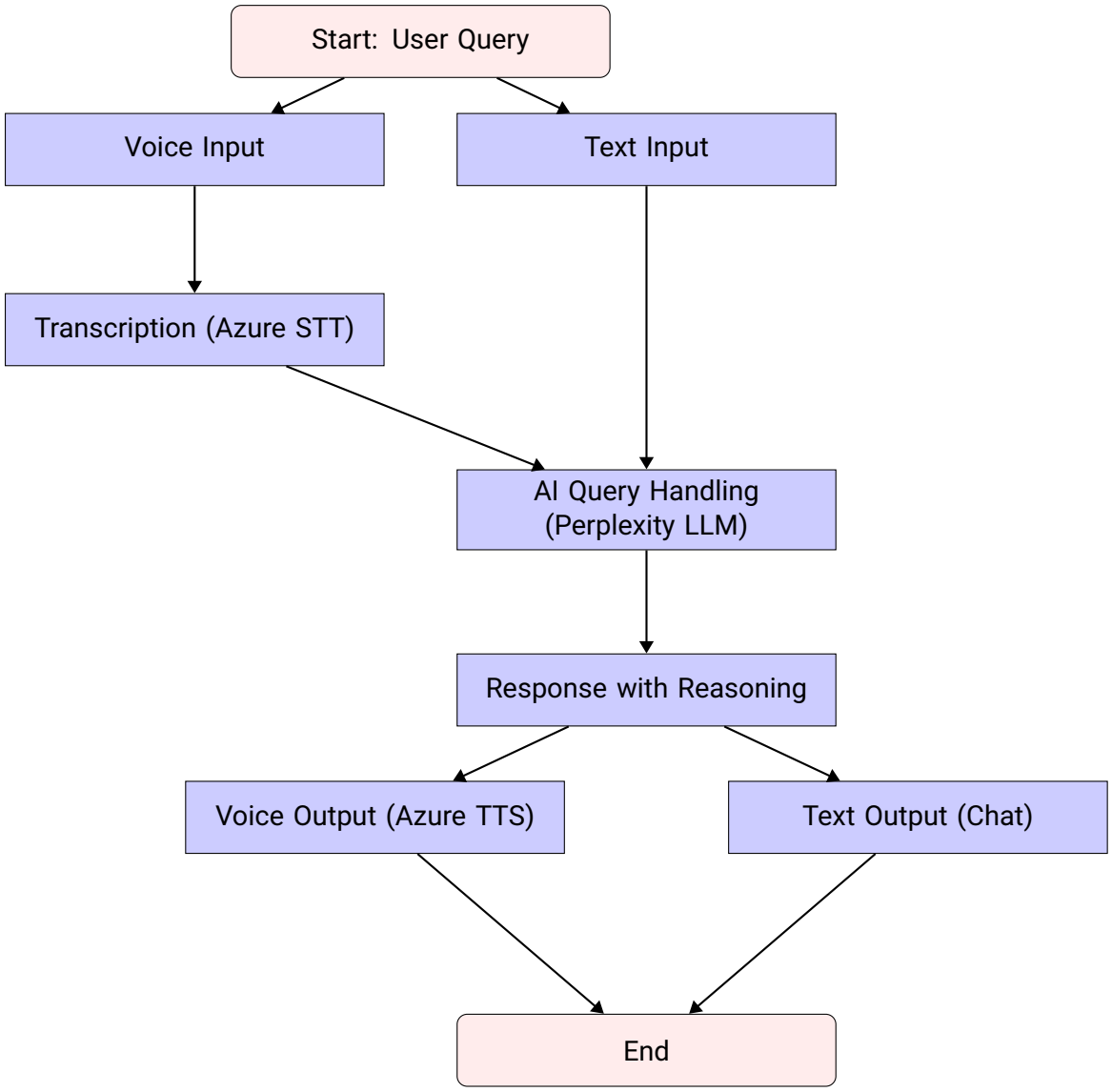

Figure 2: User Query to AI Response Flowchart

8 Key Modules and Components

- **AI Assistant Modal:** The central chat interface handling text and voice communication.
- **Floating Action Button (FAB):** The main entry point to launch the AI assistant.
- **Product Grid & Detail Pages:** The core e-commerce UI, built with static HTML/CSS, which provides context for the AI.
- **Speech Utilities:** JavaScript module implementing ASR and TTS using the Azure Speech SDK.
- **Immersive Zone:** Component on the product page linking to external VTO and AR Visualizer tools.

9 User Experience and Interface

AI-Vastra features a clean, modern e-commerce UI:

- A purple floating microphone button (FAB) provides persistent, non-intrusive access to the AI assistant.
- The assistant opens in a modal window, which is familiar to users and keeps context.
- Clear visual separation between user messages and assistant messages in the chat.
- Real-time feedback during processing ("Listening...", "Thinking...").
- "Immersive Zone" on product pages clearly invites users to "Try-On" or "Visualize" products.

10 Implementation Highlights

- **Multi-API Integration:** Successfully streamlined multiple, disparate APIs (Azure, Perplexity, Streamlit) into a single, cohesive-feeling JS application.
- **Multilingual Voice:** Robust implementation of Azure SDK for seamless speech-to-speech conversation in both English and Hindi.
- **Client-Side SPA:** The entire site runs as a client-side application, with JS managing page state and view toggling without full page reloads.
- **Contextual Prompting:** The AI prompt is dynamically built to be aware of the user's context (e.g., current product page).

11 Evaluation and Results

The deployed prototype, accessible via the GitHub repository, demonstrates:

- Seamless switching between text and voice interaction modes and languages.
- Context-aware, relevant AI responses from Perplexity.
- Successful integration and proof-of-concept for the VTO and AR Size Visualizer modules.
- High accessibility for users with diverse digital literacy levels via the voice-first interface.
- Smooth voice features with high speech recognition and natural voice synthesis quality.

12 Dissemination & Impact

The innovation of this project lies in the **system integration and inclusive design** for e-commerce, not in novel algorithms. Therefore, academic publication is preferred over patenting. Intended submissions to **HCI, Design Research, and Tech-for-Social-Good** conferences aim to share the novelty of combining multimodal AI with visualization tools to solve trust and accessibility challenges in e-commerce.

13 Conclusion

AI-Vastra exemplifies the potential of integrating recent AI advancements (LLMs, Speech, Vision) with thoughtful multimodal interface design. It creates an impactful tool for enhancing trust and accessibility in India's diverse e-commerce landscape. The project's architectural and interaction design can serve as a blueprint for similar digital inclusion efforts aimed at bridging the gap between technology and new-to-internet users.