Project Title:

smolVLM: A Lightweight Vision-Language Assistant for Sustainable and Inclusive Digital Navigation

Team Name (optional): TechNirvana

College Name: Anurag University

Chosen SDG Target

SDG 9:

By 2030, upgrade infrastructure and retrofit industries to make them sustainable, with increased resource-use efficiency and greater adoption of clean and environmentally sound technologies and industrial processes, with all countries taking action in accordance with their respective capabilities.

Problem Statement

Many users, especially those with limited digital literacy, face difficulty navigating modern digital systems. This is compounded by the digital divide in underserved or rural areas, where access to advanced devices and reliable internet connectivity is limited. Additionally, rapid tech obsolescence contributes to growing e-waste, driven by dependency on cloud-based and resource-heavy applications.

According to the UN, over 53.6 million metric tons of e-waste was generated in 2019 alone, a figure projected to rise sharply. Furthermore, around 2.9 billion people globally remain offline, many due to accessibility barriers. These realities call for sustainable, lightweight, and inclusive tech solutions.

Abstract

The proposed solution addresses the intersection of digital accessibility and sustainability. Our system, **smolVLM**, is a lightweight, locally-running vision-language assistant that provides real-time navigation support by interpreting screen content. It is designed to help users—especially elderly, digitally underserved communities, and persons with disabilities—interact with complex interfaces through contextual visual and textual assistance.

By operating entirely on local hardware, **smolVLM** avoids high-energy cloud systems, thus reducing its carbon footprint. Its ability to run on older or low-power devices promotes longer device lifespans, helping reduce e-waste. This dual impact—environmental and social—makes it a powerful solution for both sustainability and inclusive digital transformation.

Proposed Solution & Objectives

Proposed Solution:

smolVLM is a vision-language model integrated into a lightweight system that captures screen content and provides real-time, guided feedback using AI. It acts like a digital co-pilot for users who struggle with navigating apps or websites.

Objectives:

- Enhance digital accessibility for underserved and elderly users.
- Reduce dependence on energy-intensive cloud infrastructure.
- Promote reuse of older devices and reduce electronic waste.
- Build a scalable, offline-first AI assistant for broader applicability.

Detailed Explanation of the Solution

Technical Overview

- Uses a pre-trained small Vision-Language Model (e.g., TinyVLM or custom fine-tuned transformer)
- Real-time screen capture and inference
- Edge-device compatible: designed to run on devices with <4GB RAM

- Provides voice/textual guidance overlays for navigation support

Architecture (Simplified)

Input: Screen feed → VLM processor → Text/voice output
Optional interaction with assistive touch or keyboard interface

Interdisciplinary Aspects

- AI/ML (Vision-Language models)
- HCI (Human-Computer Interaction)
- Embedded Systems and OS-level hooks
- Environmental sustainability and digital sociology

Design Thinking Application

- Empathy interviews with elderly and rural users
- Defined user personas
- Iterative low-fidelity prototyping
- Feedback-led improvements

Proof of Concept (if applicable)

Prototype Description

- Prototype developed using Python + OpenCV for screen capture
- ONNX-compatible Vision-Language model embedded
- Offline mode tested on Raspberry Pi 4 and 2015 laptops

Technologies Used

- Python, PyTorch, OpenCV, ONNX
- HuggingFace vision-language models
- Local text-to-speech & OCR modules

Initial Results

- Successfully guided users through email login and form filling
- Demonstrated on 4GB RAM system with smooth operation

Images/Sketches: (Attach if available; placeholder for now)

Expected Outcomes

Short-Term

- Improved tech accessibility for elderly and digitally underserved
- Working low-cost prototype tested on old hardware

Long-Term

- Reduction in e-waste by promoting reuse of older devices
- Scalable model for offline education and digital inclusion
- Potential for commercialization or open-source community adoption

Environmental Impact: Decreased reliance on cloud infrastructure reduces CO_2 emissions

Social Impact: Promotes digital equity

Economic Impact: Cost-effective solution for NGOs and rural institutions

---

Resources Required

Materials & Equipment

- Raspberry Pi, old laptops
- Basic hardware peripherals (microphone, speakers)

Technologies

- VLM models (TinyVLM, LLaVA-lite), TTS, OCR, local UI capture

Budget Estimate: 15,000–20,000

Mentorship Areas Needed

- Optimization of AI models for edge devices
- Human-centered design validation

Team Details

Name	Departments	Year	Role in Project
Siddu G	AIML	3rd	Model Development & Testing
Vighneshwar K	ECE	3rd	Hardware Integration
Dharun	CSE	3rd	Front-end + User Testing
Amulya	CSE	3rd	Design Thinking Lead
Kuldeep	CSE	3rd	Documentation & Outreach

Faculty Mentor

Name: Dr. Narendhar Singh
Position: Associate Professor, ECE
Email: narendarsinghece@anurag.edu.in
