The Digital Cortex (ViDocX): Open-Source Cognitive Infrastructure for Accessible STEM Diagrams

A Neuro-Symbolic Architecture for Accessible Diagram Understanding

1. Scientific Foundation: The "Two-Streams" Hypothesis

Current Vision-Language Models (VLMs) like GPT-4 process images as a unified field of pixels, frequently leading to "hallucinations" where the AI invents connections that do not exist. To solve this, ViDocX mimics the human visual cortex, specifically drawing on the **Two-Streams Hypothesis** (Goodale & Milner, 1992).

Our architecture explicitly splits the processing pipeline into two biological mimics:

- **The Dorsal Stream mimic ("The Where"):** A geometric parser that handles spatial localization, topology, and motion (arrows/flow).
- **The Ventral Stream mimic ("The What"):** A semantic classifier that identifies object labels and text content.

By decoupling "Structure" from "Semantics," we ensure that the system generates a **Verifiable Scene Graph**, preventing the probabilistic errors common in standard AI.

2. System Architecture Breakdown

Module A: The "Artificial Retina" (Visual Perception Layer)

This module functions as a domain-independent feature extractor. It does not "guess" meaning; it extracts raw primitives.

- **Stage 1: Segmentation:** Utilizes a fine-tuned object detection model (e.g., YOLOv8 / Faster R-CNN) to identify generic primitives: `[Node, Connector, Text_Block, Cluster]`.
- **Stage 2: The "Dorsal" Geometric Check:** A rule-based algorithm calculates vector relationships (Intersection over Union, Nearest Neighbor) to mathematically prove connectivity (e.g., *Arrow_Tail(x,y) is inside Box_A(x,y)*).
- **Stage 3: Graph Construction:** Outputs a raw **Scene Graph** (JSON/RDF) representing the diagram's topology.

Module B: The Cognitive Diagram Ontology (CDO)

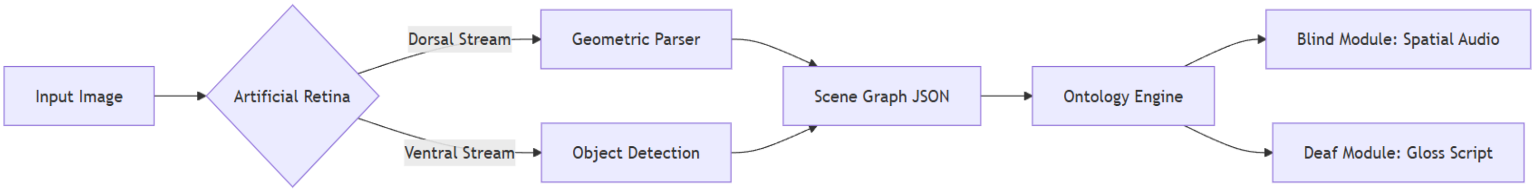
To ensure the system is **Domain Independent**, we map the raw Scene Graph to a standardized Logic Layer.

- **The Ontology:** We will publish an open OWL (Web Ontology Language) standard defining universal relationships (e.g., `flowsTo`, `isParentOf`, `contains`).
- **The Benefit:** This separates the *visual representation* from the *logical meaning*, allowing the same engine to eventually support Biology or Engineering diagrams by simply swapping the ontology rules.

Module C: The "Artificial Broca's Area" (Expression Layer)

This layer translates the internal logic into human-centered outputs.

- **For Deaf Users (Linguistic Translation):** Instead of an "Audio" pipeline, this module functions as a language production center. It converts the Scene Graph into **Sign Language Gloss** by applying spatial grammar rules (e.g., converting `Node A -> Node B` into `NODE-A (Setup:Left) -> NODE-B (Setup:Right) -> CONNECT`).
- **For Blind Users (Spatial Navigation):** This module indexes the graph for **Non-Linear Traversal**, allowing users to navigate "Parent-to-Child" using keyboard inputs and spatial audio panning.



3. Research Strategy & Scope Management

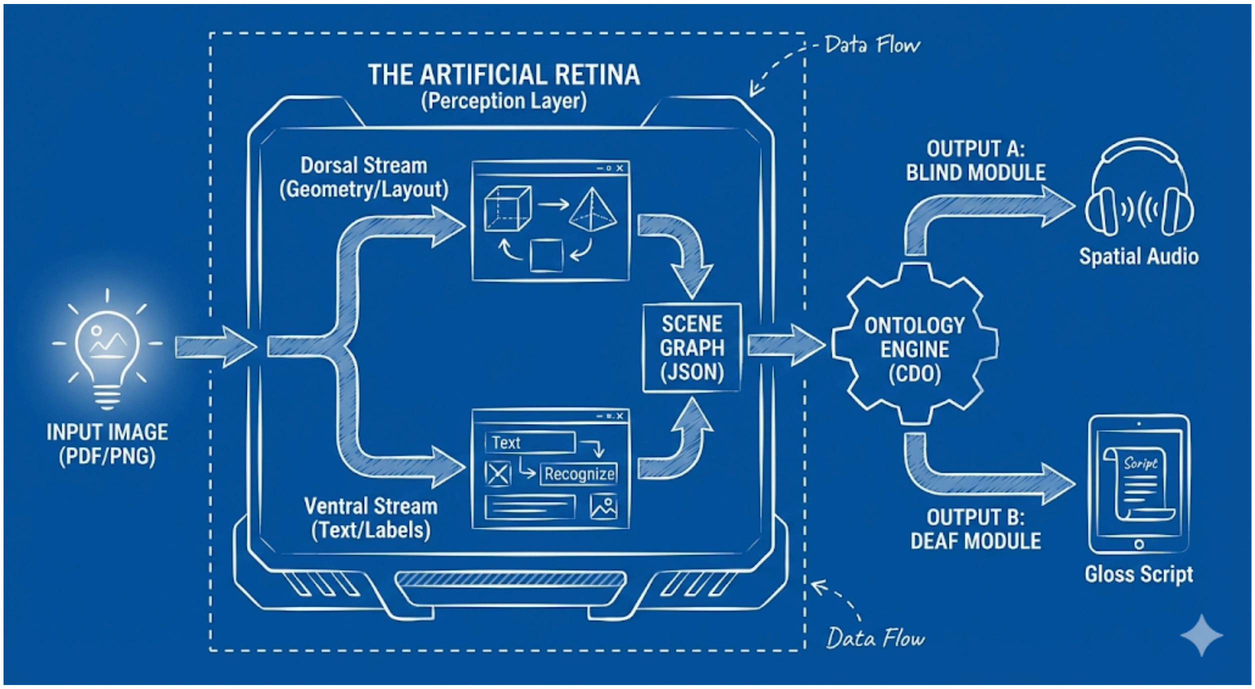
The "Universal Engine" Strategy

While our goal is a domain-independent architecture, trying to solve *all* diagrams in 6 months is unrealistic.

- **The Constraint:** We will build the **Universal Core Engine** but validate it strictly on **Flowcharts and Finite State Automata**.
- **Why Flowcharts?** They provide a rigorous "stress test" for branching logic and recursion without the noise of artistic variation found in biological diagrams. This ensures we deliver a stable, working prototype within the grant window.

4. User Interaction Design (Use Cases)

User Profile	Current Reality	The ViDocX Intervention
The Blind "Explorer"	Listens to a screen reader say "Image: Flowchart." Passive and confusing.	**Active Navigation:** Presses 'Down' to follow a path. Hears "Decision Diamond" in center audio, then "Yes Path" in right-ear audio. Builds a mental map.
The Deaf "Visualizer"	Struggles to decode dense English paragraphs explaining the image (High Cognitive Load).	**Linguistic Equity:** Reads a clean **Gloss Script** below the image. Logic is stripped of English filler and presented in native Sign syntax: START -> CHECK-FAIL -> RESTART.



5. Technical Implementation Plan

Phase	Deliverable	Technical Goal
1. Vision (Months 1-2)	**The "Retina" Pipeline**	Reliable detection of Arrows vs. Lines in Python (OpenCV/PyTorch).
2. Logic (Months 3-4)	**The Ontology (OWL)**	Defining the .OWL standard and JSON-LD mapping.
3. Access (Months 4-5)	**The Prototypes**	(1) Web-based "Blindwalker" navigator. (2) Text-based Gloss Generator.
4. Validation (Month 6)	**User Report**	Usability testing with 20 Blind & Deaf students at KDU (Paid Study).

6. Budget Justification & Open Source Promise

Budget Category	Activity & Reliability Upgrade (Justification)	Est. Effort / Allocation	Cost (€)
1. Core Architecture & AI Training	**Activity:** Finalizing the Domain-Independent Ontology (CDO) and training "Artificial Retina" vision models. **Reliability Upgrade:** Covers **Cloud GPU training costs** for high-accuracy detection and implementation of	~150 Hours + Cloud Compute Costs	**€5,500**

	CI/CD pipelines to ensure a regression-free, stable engine.		
2. Accessibility Modules (The Novelty)	**Activity:** Developing the "English-to-Gloss" rule engine (Deaf) and Vector-Based Spatial Navigation (Blind). **Reliability Upgrade:** Covers engineering for **Edge Cases** (e.g., nested loops) and optimizing **low-latency API response** times essential for real-time interaction.	~170 Hours	**€6,000**
3. Rigorous User Validation & Ethics	**Activity:** Conducting formal usability trials with Blind and Deaf participant groups. **Reliability Upgrade:** Scales study to **20 paid participants** (statistically significant). Specifically covers **ethical compensation (stipends)** and creating an "Accessibility Impact Report."	~60 Hours + Participant Stipends	**€2,000**

4. Open Source Compliance	**Activity:** Dockerization, API Reference creation, and Community Onboarding materials. **Reliability Upgrade:** Ensures the repo is **"Adoption Ready"** for the NGI community, including comprehensive "Getting Started" guides and architecture documentation.	~50 Hours	**€1,500**
TOTAL		**~430 Hours**	**€15,000**

- **Regional Efficiency:** In the context of Sri Lanka, the requested €15,000 allows for **400+ hours of dedicated engineering time**, maximizing the ROI for NLnet.
- **Licensing:** All code will be released under **GPLv3**. The CDO Ontology will be **CC-BY-SA**.
- **Sustainability:** By building a standard API, we aim to allow platforms like Moodle and Wikipedia to integrate "The Digital Cortex" as their default accessibility layer.