

Modelling Cognitive Brain Processes

CS298 Master's Defense

Adithya Harish Kumar

Department of Computer Science
San José State University

Advisors

Dr. Chris Pollett

Dr. Robert Chun

Dr. Sayma Akther

Presentation Roadmap

- **Project Overview**
 - Motivation, Project Goals and Impact
- **Background**
 - Related work, multimodal models, agent orchestration
- **Preliminary Work**
 - CrewAI baseline, Topology Manager design, Housefly simulation
- **Implementation**
 - Phi-3.5-mini-instruct finetuning
 - Multimodal processing
 - Visual cortex modeling (dorsal-ventral, V2-V4-V5 pathways)
 - Saliency visualizer
- **Experimental Results**
 - Reasoning benchmarks
 - CLEVR + COD10K Experiments
 - Visual Saliency evaluation
- **Conclusion & Future Work**
 - Key findings, impact, and research directions

Project Overview

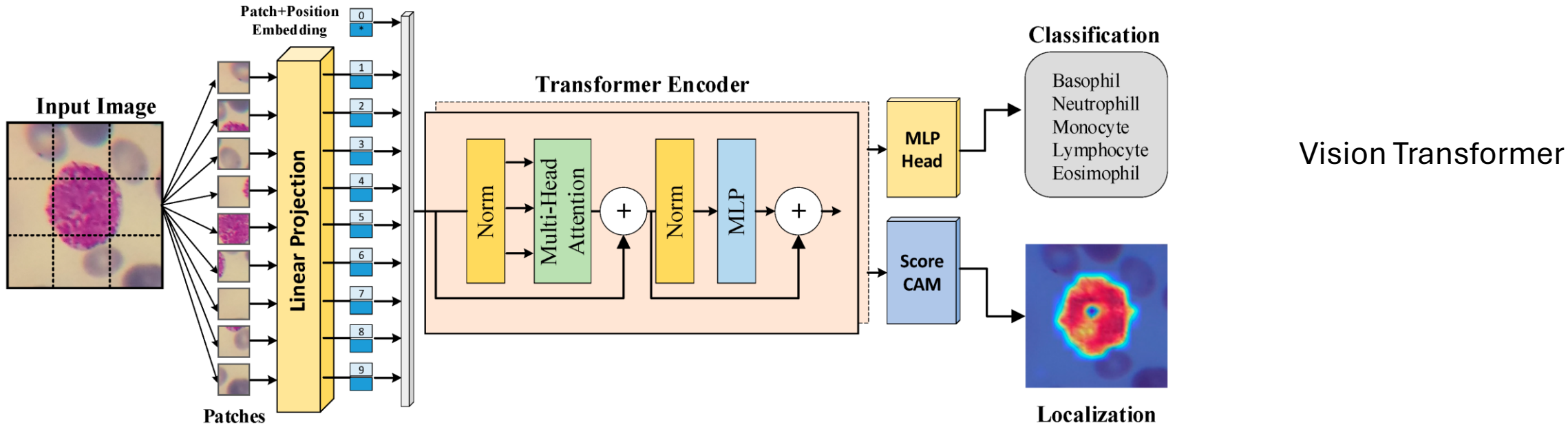
- **Motivation**
 - Simulate complex cognitive and perceptual brain processes using AI agent systems
 - Leverage Large Language Models (LLMs) to model human-like reasoning and multimodal perception
- **Project Goals**
 - Build modular agent-based simulations of the visual cortex using LLM topologies
 - Enable cognitive functions like contour detection, motion, and color perception via coordinated agents
 - Develop the Topology Manager framework for flexible deployment and interaction of LLM agents
 - Fine-tune advanced models (Phi-3.5-mini-instruct) for logical reasoning and integrate with multimodal tasks
- **Impact**
 - Demonstrates the potential of multi-agent LLM systems to provide interpretable, human-inspired cognition
 - Lays groundwork for future developments in brain-like AI architectures and multimodal understanding

Background - Related Work

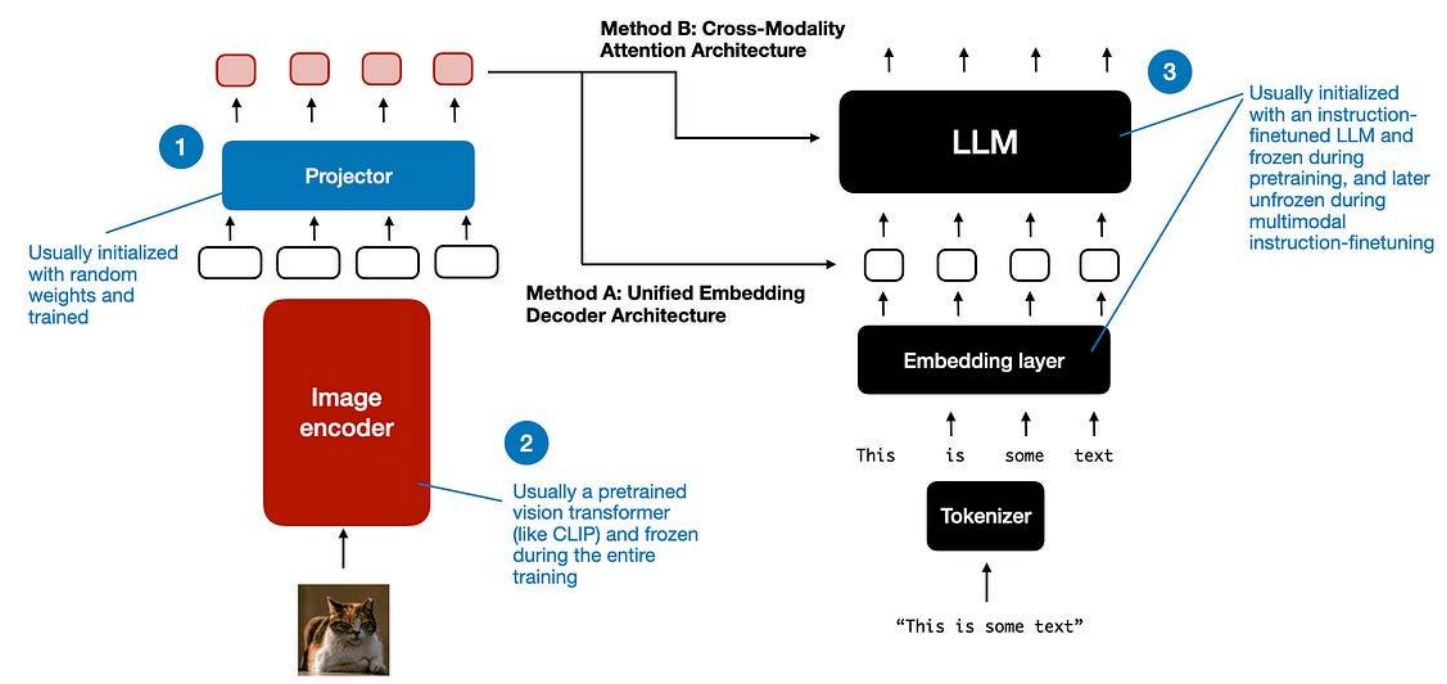
- **Dennett et al. [1] Proposed the Multiple Drafts Model (MDM) of consciousness**
Rejects a single “central observer”; consciousness arises from parallel, competing narrative drafts. Explains subjective experience via distributed, heterarchical processing rather than a unified stream; accounts for phenomena like temporal illusions, confabulation, and post-hoc editing through decentralized interpretive processes
- **Chen et al. [5] built a large-scale V1 model** integrating anatomical + neurophysiological data. Replicated visual processing traits: **robustness to noise**, temporal stability, distributed feature detection
- **Wang et al. [6] CLIP-trained ResNet50** explained **~79% of voxel response variance** in hold-out brain data. Outperformed models trained only on images or text
- **Wang et al. [8] Video saliency model** trained on **DHF1K dataset** (1,000 clips, 250Hz eyetracker). Improved spatio-temporal saliency predictions with **DeepLabV3**; analyzed **age-related fixation patterns (20–28 yrs)**
- **Hulse et al. [9] →** Mapped the full synapse-level connectome of the Drosophila central complex. Revealed circuit motifs supporting flexible navigation & action selection
Identified recurrent loops, fan-in/fan-out pathways, and compass–sensory integration channels enabling context-dependent behavior; provides a high-resolution neural architecture for biologically grounded navigation models

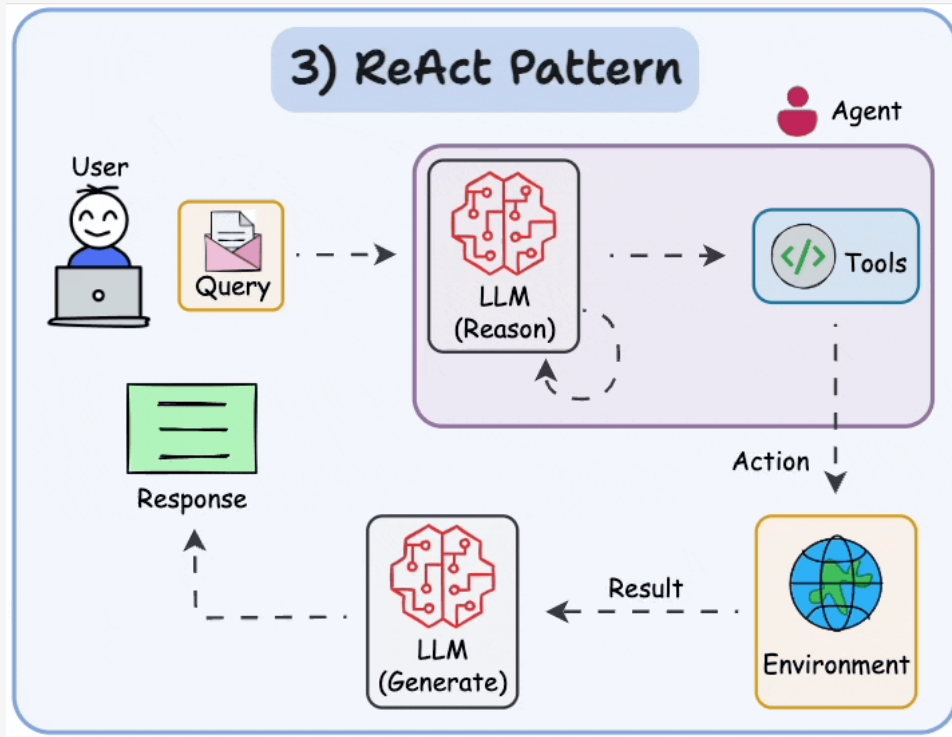
Background - Multimodal Language Models

- **From text-only → Vision + Language reasoning**
- Key systems: **GPT-4, Gemini, LLaVA** → combine pretrained LLMs with **Vision Transformers (ViTs)**
- **Applications:** Image captioning, VQA, multimodal reasoning
- **Impactful models:**
 - **DALL-E** → Generates realistic images from text prompts → simulates “mental imagery.”
 - **Flamingo** → Few-shot multimodal model → learns cross-modal tasks with minimal supervision
- **Scientific Relevance:**
 - Replicates **dorsal (where) + ventral (what) streams** of the visual cortex
 - Provides **better saliency prediction & scene understanding** than unimodal approaches



Multimodal Architecture





Background - Agent Orchestration

- **Multi-agent LLMs** = multiple specialized models collaborating
- **Frameworks:**
 - **Crew AI, AutoGPT** → agents with specific roles & goals
 - Limitation: Communication often **opaque**
- **Advances:**
 - **LangGraph, ReAct, Chain-of-Thought (CoT)** → structured, step-by-step reasoning flows
 - **Retrieval-Augmented Generation (RAG)** → connects agents to external vector DBs (**FAISS, LanceDB**) for dynamic domain knowledge
- **Why it matters:**
 - Supports **modular, interpretable, distributed reasoning**
 - Enables brain-inspired systems where **agents = cortical regions** with controlled pathways

Preliminary Work - Dennett's Multiple Drafts Model

- **Core Ideas**

- Proposed by Daniel Dennett in *Consciousness Explained* (1991)
- Denies any central “Cartesian Theater” (no single place or moment in the brain where consciousness “happens”)
- Instead, there are **parallel, multitrack** processes: different subsystems of the brain make content-fixations (detecting features, interpreting sensory inputs, etc.) at different places and times

- **Mechanisms & Implications**

- These content-fixations are continuously edited (“drafts”)—new interpretations, revisions, additions—across time
- Not all “drafts” reach what we consider full conscious awareness; only some become “famous” or influential enough to affect behavior, memory, report
- Consciousness is thus dynamic, distributed, and emergent—tied to behavior and the capacity of content to influence actions rather than to the presence of a single, fixed observer

- **Why It Matters for Cognitive Modeling**

- Motivates designing systems with **modular, parallel sub-agents** rather than monolithic sequential pipelines
- Suggests focus on how drafts (partial interpretations) are generated, edited, and selected
- Offers a framework to compare artificial systems to human-like consciousness in terms of integration, revision, and competition among representations

Preliminary Work – CrewAI Implementation

- Create a **baseline system** that follows Dennett’s model: multiple agents (“drafts”) working in parallel, then integrated, to simulate aspects of conscious processing
- **Framework:** CrewAI multi-agent LLM system
- **Agents:**
 - **LLM1 (Quantitative Focus)** — extracts and summarizes numerical/quantity information
 - **LLM2 (Logical Relationships Focus)** — identifies logical relations, interactions, structures in the input
 - **Primary LLM (Integrator)** — takes outputs from LLM1 and LLM2 and produces final answer
- **Procedure**
 - Feed 20 input questions/tasks: arithmetic (quantities), logic, everyday reasoning
 - Agents work in parallel, produce their “drafts,” then integrator synthesizes into final output
 - Observe and document not just correctness, but how partial interpretations differ, what is preserved or discarded, whether context or realism enters into the integrator’s answer (beyond raw logical quantity)

Preliminary Work – CrewAI Implementation

Task	Quantitative Agent (LLM1) Output	Logical Agent (LLM2) Output	Final Answer (Integrator)
<i>“I have 100 apples. I gave 20 to my sister and 10 to my brother. How many are left?”</i>	Summarized quantities: 100, 20, 10	Logical: gave away, subtract, remaining	70 apples
<i>“There are 10 crows on an electric line. If a hunter shoots the leftmost crow ... how many remain?”</i>	Quantities: 10 initially, 1 shot	Relations: removal, context of noise scaring away others	0 crows (realism); or 9 if only raw subtraction considered
<i>Logical syllogism (“All trucks fly...” etc.)</i>	Identifies premises’ quantities/attributes	Derives relations among categories; checks validity	Correctly rejects invalid conclusions; selects only supported ones

Preliminary Work – Topology Manager

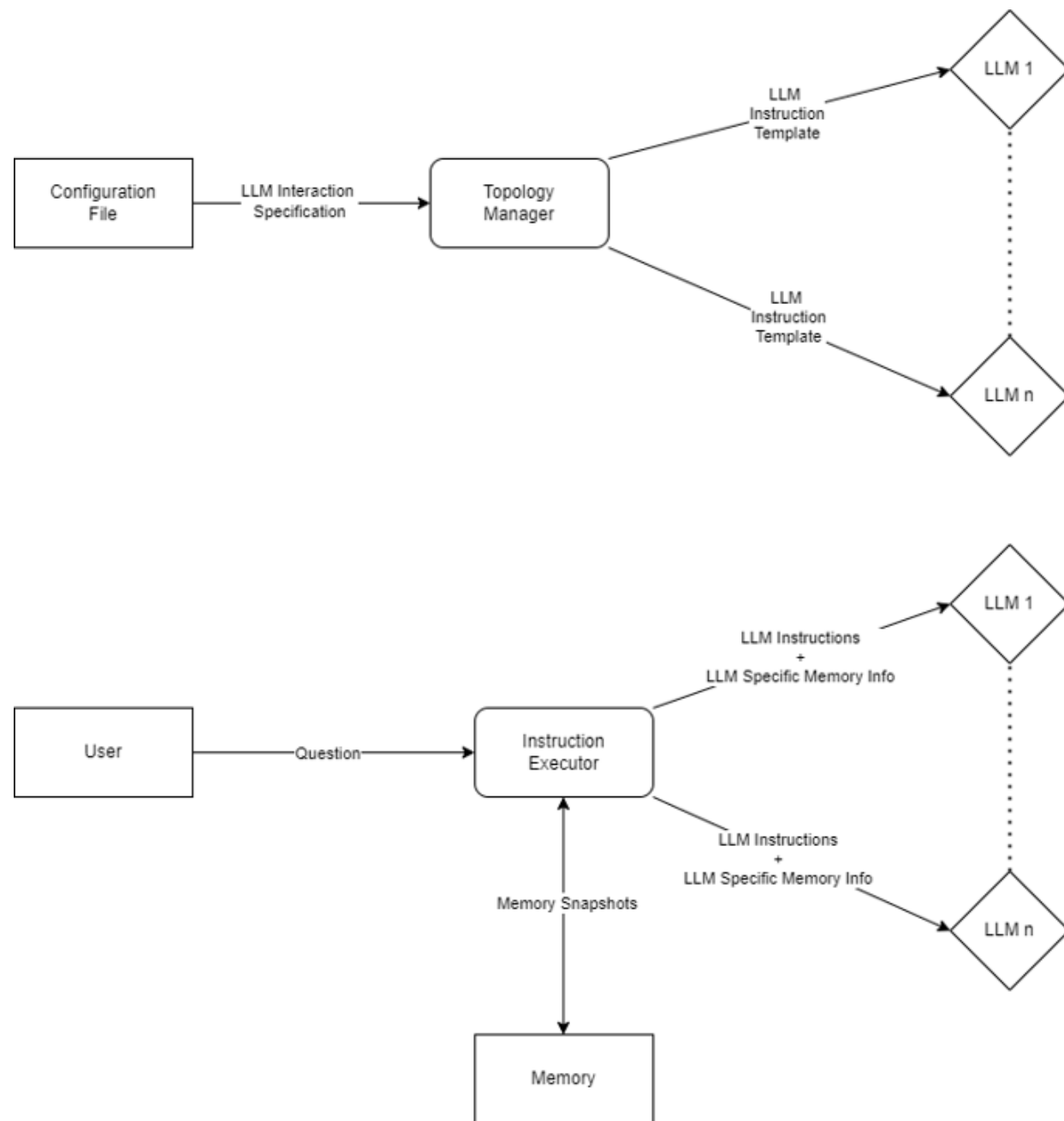
Purpose

- Framework to **orchestrate LLM interactions** with fine-grained control
- Goes beyond CrewAI → lets us **decide which agents talk, how data flows, and what is shared**
- Inspired by **Infrastructure-as-Code (IaC)** → reproducible, configurable, and scalable

Key Features

- **Declarative config** → Graph-based agent topologies defined in JSON.
- **Version-controlled & reproducible** → Easy rollback & deployment
- **Automated deployment & scaling** → Minimal human intervention
- **Consistency across environments** → Prevents human error

Preliminary Work – Topology Manager



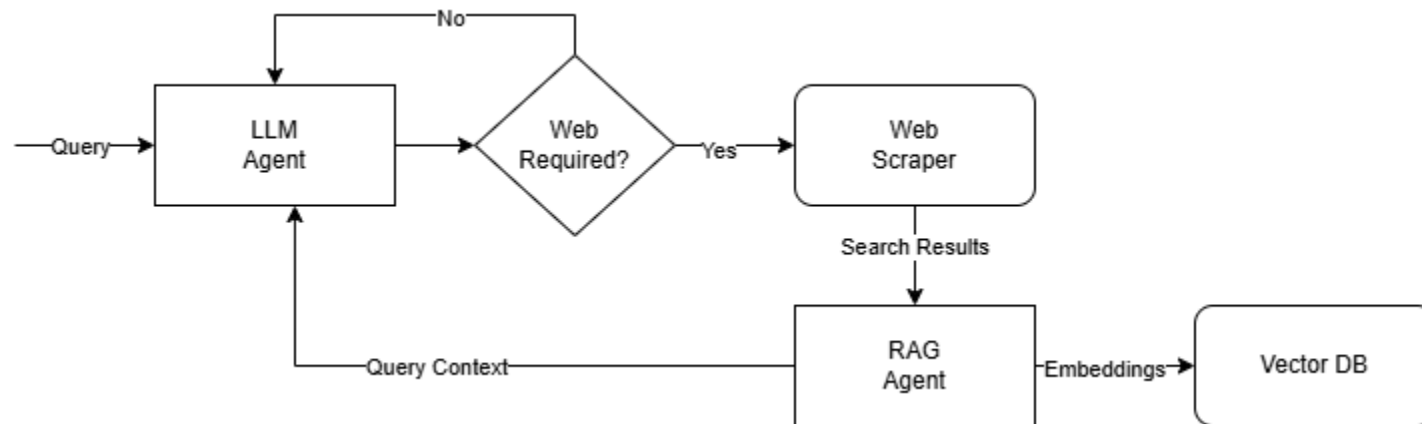
Preliminary Work – Topology Manager

- **Architecture**
 - **Configuration File** → Blueprint of agents + interaction graph
 - **Topology Manager** → Builds hierarchy, topological sorting, execution order
 - **Instruction Executor** → Central processor, query handler, memory snapshot support
 - **Memory Module** → Context persistence via snapshots
 - **LLM Modules** → Cloud (OpenAI, Gemini, OpenRouter) + local (Ollama, LM Studio)
- **Supporting Modules**
 - **Config** (validation & cycle checks)
 - **Core** (module interaction, async comms)
 - **DB** (RAG, vector DB abstraction)
 - **Manager** (agent orchestration in topological order)
 - **Model** (LLM abstraction layer)

Preliminary Work – Topology Manager (RAG)

- **Purpose**
 - Enhance LLM inference with **query-relevant retrieval**
 - Enables **offline execution** without cloud LLMs
 - Improves accuracy, adaptability, and contextual reasoning
- **Pipeline Components**
 - **Web Scraping** → Async Python crawler for latest info
 - **Vector DB (LanceDB)** → In-memory multimodal storage & retrieval
 - **Embedding Model** → *nomic embed v1.5* (384-dim multimodal embeddings)
 - **On-Device LLM (Ollama)** → Consumes retrieved results; cloud-free inference
- **Benefits**
 - **Contextual Understanding** via efficient retrieval
 - **Improved Accuracy** using latest, web-backed info
 - **Adaptive Learning** → continuously updating knowledge base

Preliminary Work – Topology Manager (RAG)



Preliminary Work – Topology Manager

Question	LLM 1	LLM 2	Primary	Answer
I have 100 apples. I gave 20 apples to my sister and 10 apples to my brother. How many apples do I have left?	You started with 100 apples. You gave 20 to your sister and 10 to your brother. Subtracting gives 70 apples remaining.	To find how many apples are left, subtract 20 (sister) and 10 (brother) from 100 → 70 apples left.	You would have 70 apples left. Calculation confirms: $100 - 20 - 10 = 70$.	70 apples
There are 10 crows on an electric line. If a hunter shoots the leftmost crow and it falls to the ground, how many crows will be left on the electric line?	The question describes 10 crows. After the hunter shoots one, 9 remain on the line.	Logical relation: one crow is removed, so 9 remain on the line.	There will be 9 crows left on the line.	9 crows
Statements: All trucks fly. Some scooters fly. Conclusions: All trucks are scooters. Some scooters do not fly. Which of the conclusions are correct?	“All trucks are scooters” is not correct. “Some scooters do not fly” is correct as per the statements.	From “All trucks fly” + “Some scooters fly,” the correct conclusion is that some scooters do not fly; all trucks are scooters is unsupported.	Correct conclusion: “Some scooters do not fly.” Incorrect: “All trucks are scooters.”	Conclusion 2 is correct; Conclusion 1 is incorrect

Preliminary Work – Topology Manager

Accuracy on 20 Questions

CrewAI	Topology Manager
80%	75%

Preliminary Work – Housefly Vision Simulation

- **Goal**

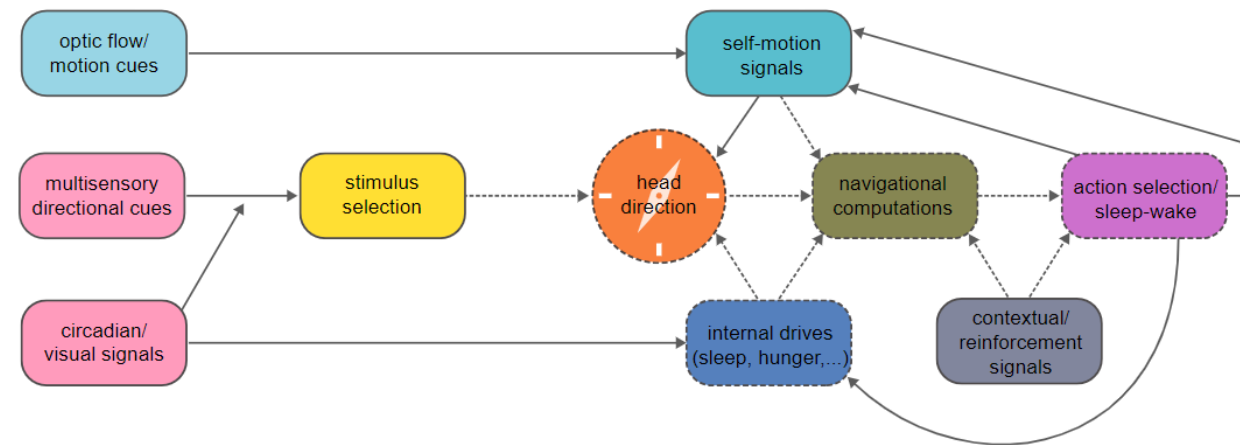
- Model the **sensorimotor pathway of a housefly** using the **LLM Topology Manager**
- Proof-of-concept for extending to more complex **human-like visual systems**

- **Subsystems Implemented (7 total)**

- **Stimulus Selection**
- **Head Direction**
- **Internal Drives**
- **Circadian Rhythm Simulation**
- **Multisensory Integration**
- **Self-Motion Processing**
- **Optic Flow Simulation**

- **Key Idea**

- Each subsystem = modular LLM agent with **biologically inspired inputs & outputs**
- Collectively simulate decision-making → perception-driven behavior



Housefly Brain Process

Preliminary Work – Housefly Vision Simulation

- **Implementation**
 - Added **conversational memory** to store previous outputs as dictionaries
 - Used **summarization** to prevent cascading context overload
 - Enabled **episodic generation** → agents aware of past states, not just current stimuli
- **Impact**
 - Simulated **continuous & adaptive behavior**, closer to real insect responses
 - Actions depended on **history + current input**, not isolated events
 - More realistic framework for studying **how organisms adapt** to dynamic environments
- **Outcome**
 - Successfully reproduced adaptive fly-like behavior
 - Establishes foundation for **modelling higher-order visual systems**



Preliminary Work – Housefly Vision Simulation

```
{  
  "query": "Behave like you are this subsystem and generate the outputs as specified and provide one of the outputs you are created to",  
  "memory": {  
    "4": "Circadian Signals: prepare for sleep",  
    "7": "Internal Drives: seek sleep.",  
    "3": "Multisensory Cues: head towards the sound, \nMultisensory Cues: orient to light source.",  
    "5": "Stimulus Selection: orient to light source.",  
    "1": "Optic Flow: turning right",  
    "2": "Self-Motion: adjust heading slightly left",  
    "6": "Output: Maintain a stable heading, adjust heading slightly left."  
  }  
}
```

Vision Simulation for a particular Timestep

Implementation - Finetuning Phi-3.5-mini-instruct

- **Model Details**

- **Phi-3.5-mini-instruct:** Decoder-only transformer, **3.8B parameters**
- Pre-trained for **instruction-following, reasoning, and multimodal tasks**
- Trained on **high-quality reasoning datasets, synthetic data, and image-caption pairs**

- **Purpose of Finetuning**

- Adapt the model for **instruction execution** and **multimodal reasoning**
- Resolve **competing outputs** from modular LLM subsystems, inspired by biological attention mechanisms
- Serve as the **key reasoning layer** for the overall cognitive modeling framework

- **Implementation Setup**

- **Platform:** Google Colab, consumer-grade GPU
- **Memory Optimization:** 4-bit quantization to fit model on limited VRAM
- **Optimization Techniques:** Unsloth → reduces memory & training overhead without affecting accuracy
- **Pipeline:** Base64 image integration → support visual queries

Implementation - Finetuning Datasets and Metrics

- **Datasets Used**

- **ReClor** – 6,000+ multiple-choice reasoning questions from graduate-level exams
 - Each datapoint: prior context + question + 1 correct answer
 - Trains logical inference & decision-making
- **LogiQA** – 2,442 independent reasoning questions
 - Used to prevent overfitting to ReClor
 - Ensures generalization beyond one dataset

- **Training & Monitoring**

- Monitored **training & validation loss** on ReClor; noted potential overfitting
- Validated periodically on **LogiQA**
- Stopped training around **100 iterations** when LogiQA accuracy plateaued

- **Results**

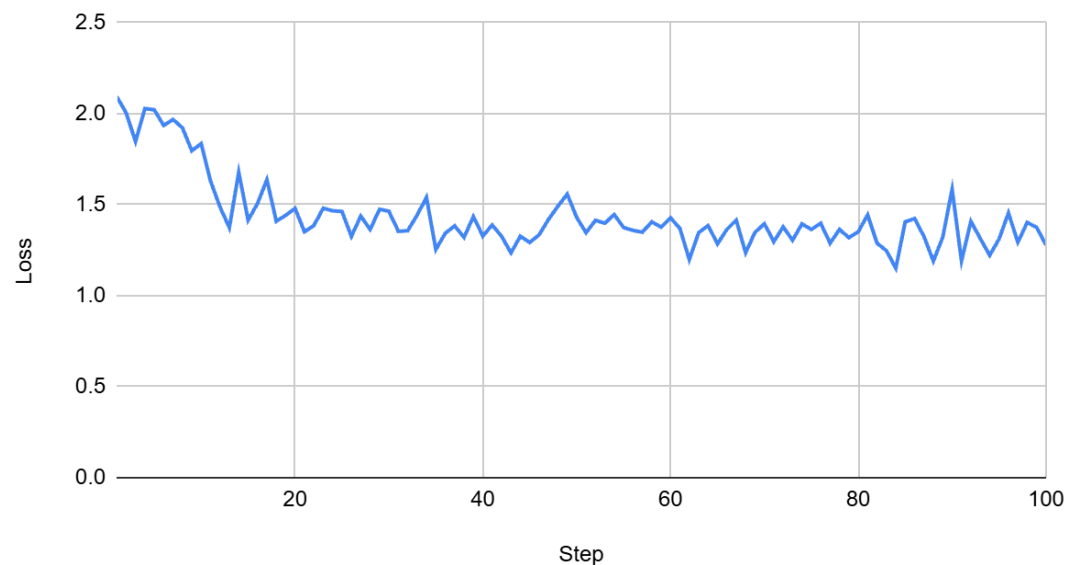
- **ReClor:** Steady loss decrease, model learned dataset-specific reasoning
- **LogiQA:** Final accuracy **~72%**, indicating balanced reasoning generalization
- **Outcome:** Phi-3.5-mini-instruct can integrate multimodal inputs and act as a **decision arbiter** for modular LLM subsystems

- **Impact**

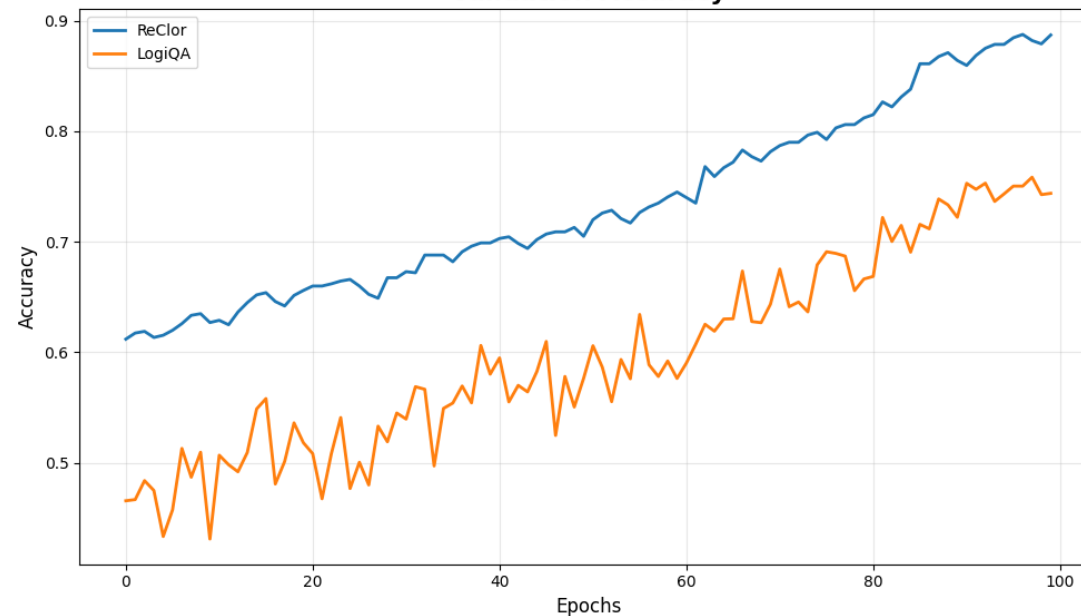
- Provides **robust reasoning layer** for hierarchical LLM setups
- Supports **distributed attention simulation** in visual and cognitive tasks

Implementation - Finetuning Phi-3.5-mini-instruct

Testing Loss



Validation Accuracy



Implementation - Multimodal Support

- **Objective**
 - Extend LLM Topology Manager to handle **both text and image inputs**
 - Simulate **visual cortex-inspired networks** where agents process multimodal information in parallel
- **Key Enhancements**
 - **Multimodal Input Handling**
 - Queries now accept **text + base64-encoded images**
 - Encoded images stored in **conversation memory** to avoid recomputation
 - Supports **seamless integration** with cloud APIs or local LLMs
 - **Instruction & Memory Updates**
 - Instruction class now supports **dict-based text-image objects**
 - Memory module stores **historical multimodal data**, enabling context-aware reasoning
 - **Architecture Impact**
 - Compatible with **existing text-based reasoning** → backward compatible
 - Modular design allows **future extensions** for other data types
 - Enhances LLMs' ability to **reason over visual context + textual information** simultaneously
- **Outcome**
 - Supports **multimodal networks** for integrated perception tasks
 - Enables **structured, reusable, and parallel reasoning** across text and visual streams

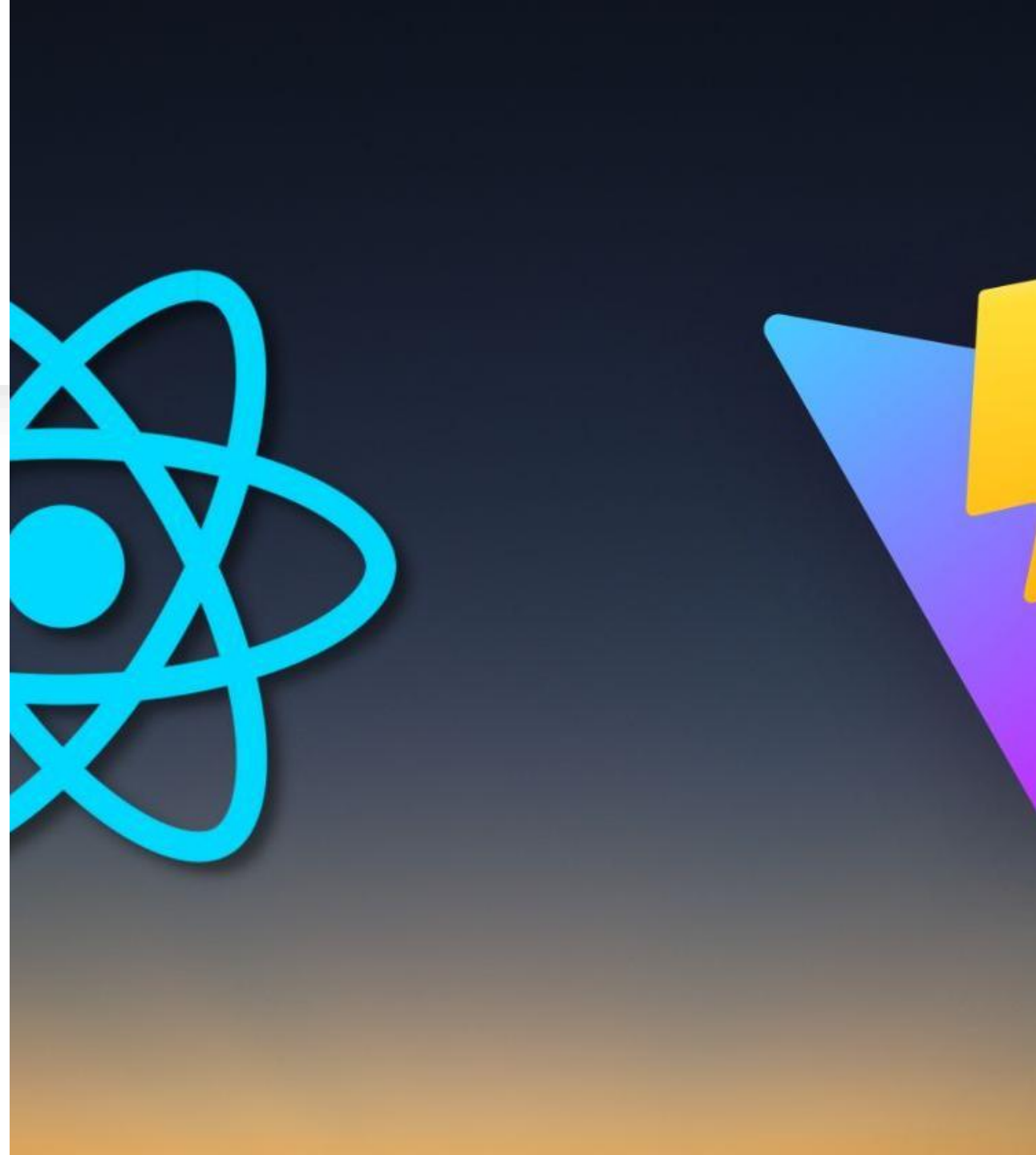
Implementation – Saliency Visualizer

- Built using **Vite + React** as a **single-page application**
- Uses **reusable components**:
 - Image upload
 - Image card with subsystem outputs
 - Description and conversation display

Two-way bound image preview on upload

Service layer:

- Formats input for backend
- Handles multipart requests
- Parses JSON responses for display



Implementation – Saliency Visualizer

- **Output Structure**

- Subsystem name
- Base64-encoded image
- Explanation of focus region

```
{  
  "names": [...],      // Subsystem names  
  "outputs": [...]     // Base64 images + descriptions per subsystem  
}
```

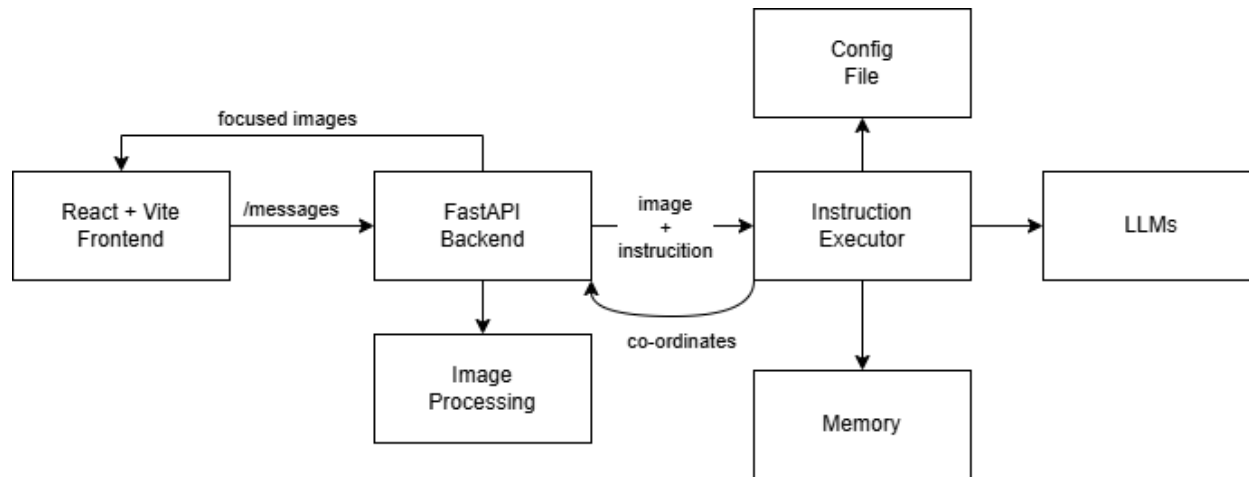
Implementation – Saliency Visualizer

- **Backend Design**

- Built using **FastAPI + Uvicorn** for **asynchronous requests**
- Initializes **InstructionExecutor** on startup with:
 - Graph-based configuration files
 - Agent communication rules
- Provides `/messages` endpoint to handle image inputs

- **Processing Pipeline**

- **Instruction Execution:**
 - Validates inputs
 - Runs image through modular LLM agents
 - Each subsystem predicts regions of focus
- **Image Processing:**
 - Circular focus regions extracted based on coordinates
 - **Gaussian blur** applied to non-focused regions (25x25 kernel)
 - Overlay focused regions onto blurred image to simulate saliency
- **Response:**
 - Structured output per subsystem
 - Includes **image + explanation**
 - Returned to frontend for **interactive visualization**



High Level Architecture

Implementation – Visual Cortex

- **System Design**

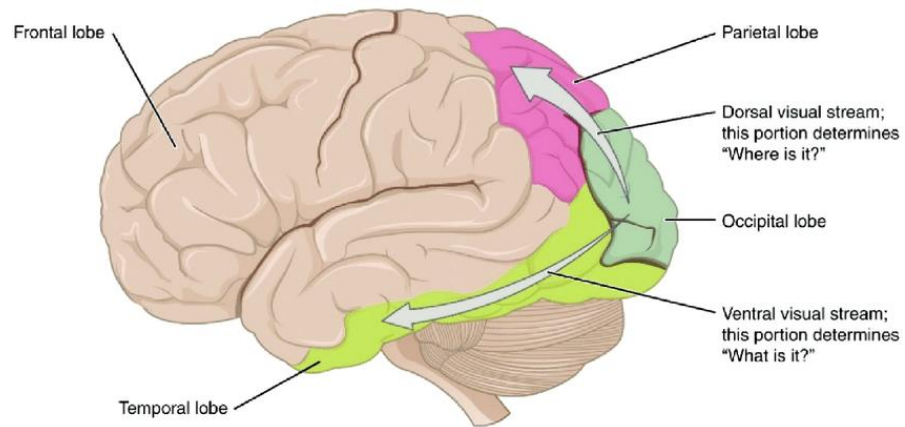
- Integrated previously developed components into a **single interactive system**
- **InstructionExecutor framework:**
 - Auto-deploys **modular agent-task inference patterns**
 - Subsystems are **nodes in a directed graph**, each producing focused outputs
 - No need to modify LLM internals; functionality is **prompt-driven**
- **Output Processing:**
 - Subsystem outputs are **visualized via image pipeline** to highlight attention
 - Enables a **scalable and modular simulation** of visual processing
- **Key Feature:** Parallel, modular, prompt-driven subsystems emulate **biologically inspired visual cortex behavior**

Implementation - Visual Cortex

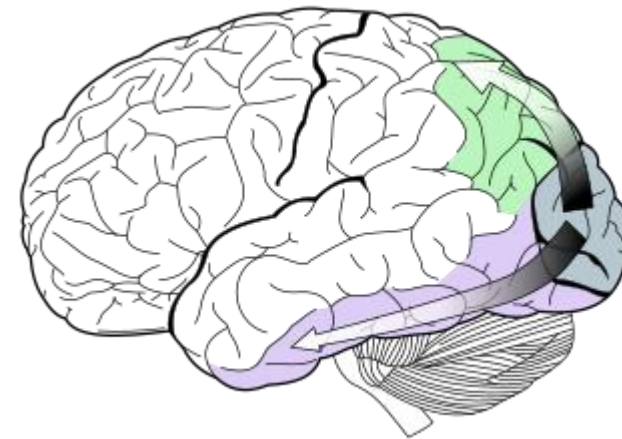
Dorsal-Ventral Pathway Simulation

- **Objective:** Model the classical "what" vs "where" visual pathways
 - **Ventral Pathway (“What”):**
 - Performs **object recognition** and **feature extraction**
 - Detects **textures, colors, shapes** in the scene
 - **Dorsal Pathway (“Where”):**
 - Processes **spatial relationships, motion, and object trajectories**
- **Implementation Approach:**
 - Separate **prompts** for dorsal vs ventral tasks
 - Both paths executed in **parallel via Topology Manager**
 - Modular agents allow flexible downstream integration and attention visualization

Implementation - Visual Cortex

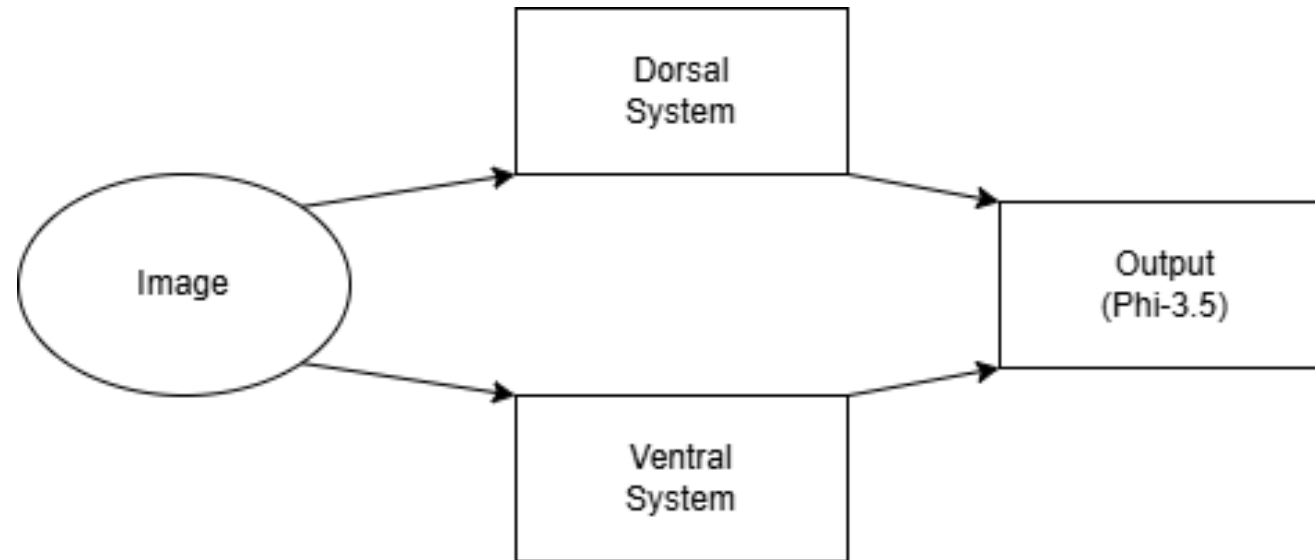


Dorsal-Ventral Stream



The dorsal stream (green) and ventral stream (purple) are shown. They originate from primary visual cortex.

Implementation - Visual Cortex



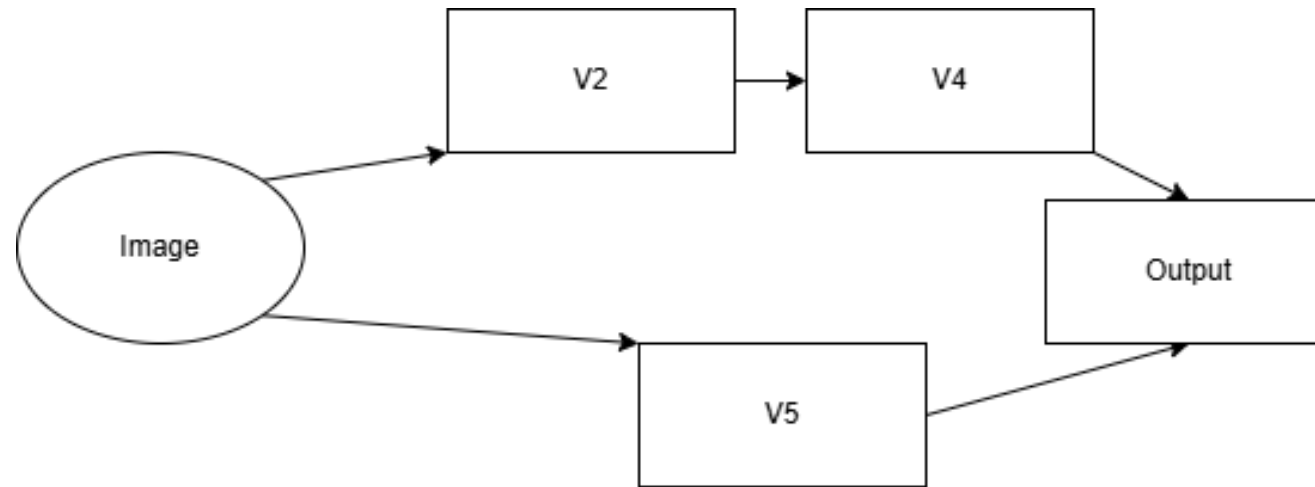
Dorsal-Ventral Subsystem

Implementation - Visual Cortex

V2-V4-V5 Hierarchical Subsystem

- **Objective:** Simulate an agentic visual hierarchy in human cortex
 - **V2 Agent:**
 - Extracts **edges, shapes, contours, and anomalies** from input images
 - **V4 Agent:**
 - Focuses on **color, contrast, texture, and object properties**
 - **V5 Agent:**
 - Responsible for **motion detection, object tracking, and dynamic scene analysis**
- **Architecture & Flow:**
 - Prompt-driven, multi-agent setup, **comparable in scale to dorsal-ventral system**
 - Outputs processed via **image post-processing pipeline**
 - Each agent produces **focused image + textual description** for attention visualization



Implementation - Visual Cortex



V2-V4-V5 system





Experiments - CLEVR

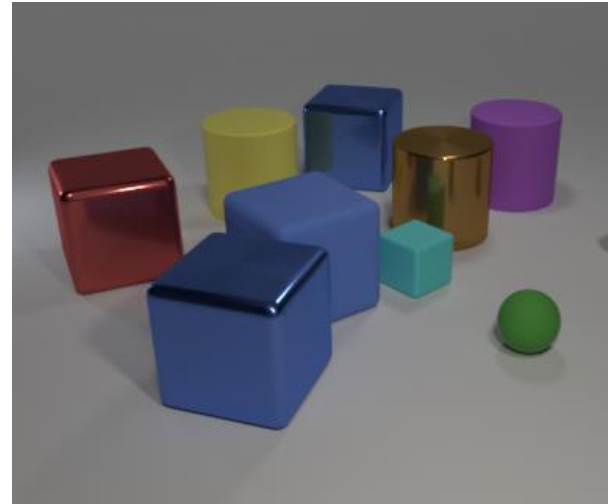
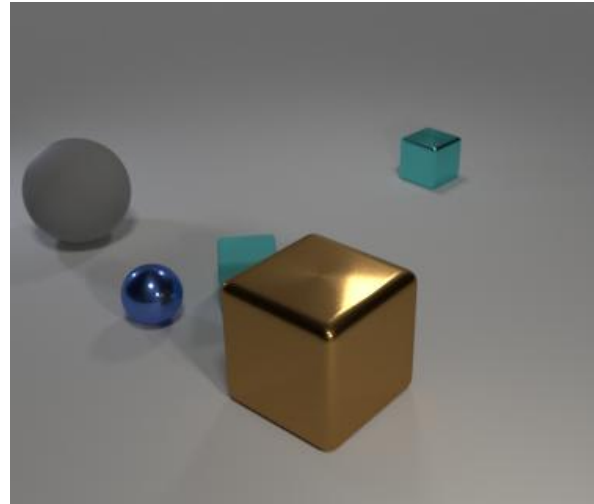
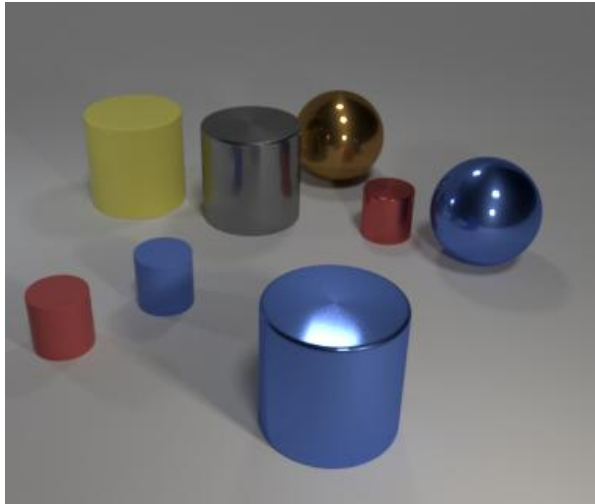
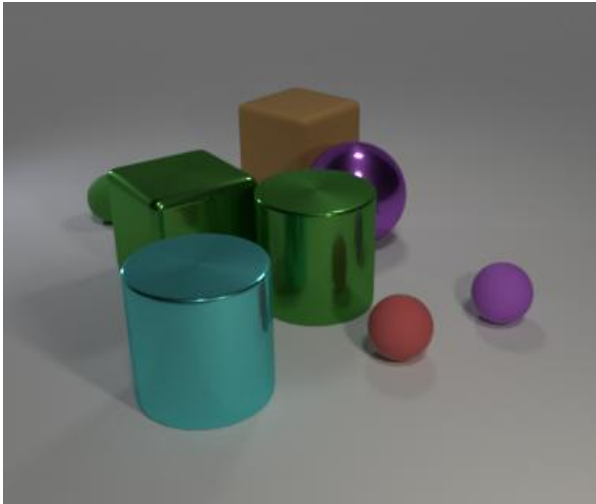
- The **human visual system** processes information through two complementary pathways
 - Ventral stream — object recognition (“what”)
 - Dorsal stream — spatial estimation and motion (“where”)
 - **Hypothesis** - A **dual-stream agent** that separates object recognition (ventral) from spatial estimation (dorsal) will produce **object distance judgments** closer to human annotations (i.e., **lower Mean Absolute Error**) than a single multimodal LLM that reasons over the whole image jointly.
 - **Goal** - Test whether this modular perception architecture leads to **more human-like spatial reasoning**
- 
- 

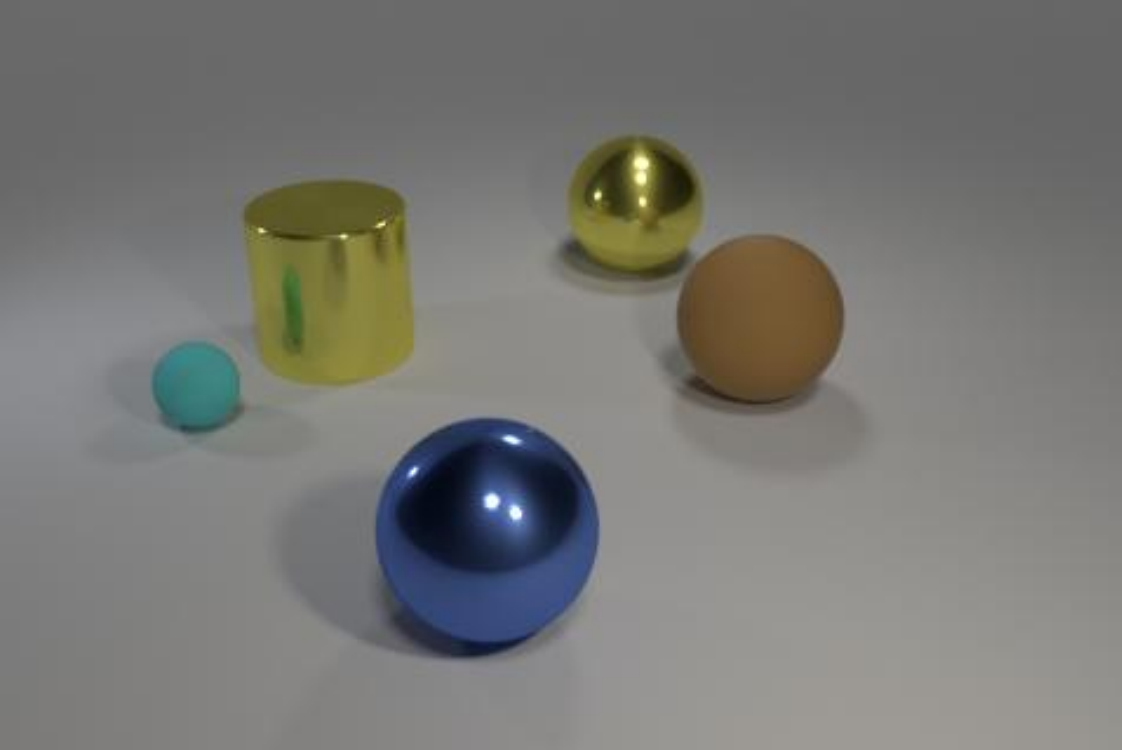


Experiments – CLEVR

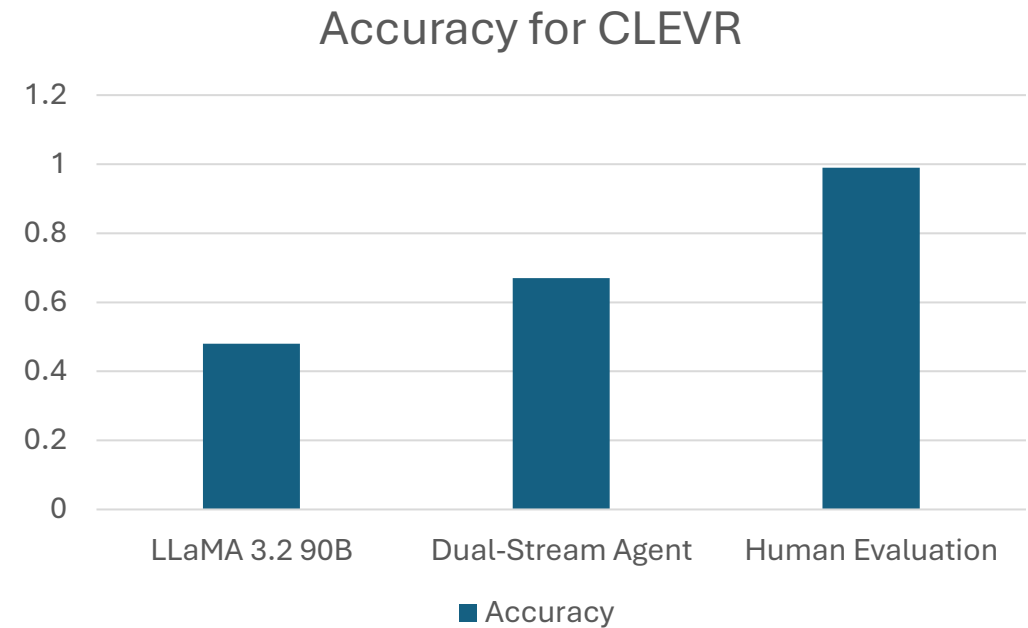
- ~100,000 Synthetic 3D scenes with **simple objects** (spheres, cubes, cylinders).
 - Each object annotated with **shape, color, size, material, and position**.
 - Enables **precise measurement** of object distances and spatial relationships.
 - Objects are **visually distinct but spatially complex**, mirroring the need for both
 - Ventral to recognize and identify objects.
 - Dorsal to estimate positions and relative distances.
 - Controlled nature of CLEVR allows testing **modularity of perception** — whether decoupling recognition and spatial estimation yields better spatial understanding.
 - Sampled **100 CLEVR scenes**.
 - Collected object distance judgments from
 - Single Multimodal LLM (LLaMA 3.2 90B)
 - Dual-Stream (Dorsal–Ventral) Agent
 - Human annotations
- 
- 

Experiments – CLEVR





CLEVR Dataset



Question	LLaMA 3.2 90B	Dual Stream Agent
Are there any other things that have the same material as the large brown object?	<p>Yes, based on the image, the following objects have the same material:</p> <ul style="list-style-type: none">The small green sphereThe large yellow cylinder <p>They all have a shiny, reflective, metallic appearance.</p>	<p>No, there are no other objects in the image that have the same material as the large brown object. The large brown object has a matte, non-reflective surface, while the other objects have different materials—several are shiny and metallic, and one is matte but a different color.</p>

Experiments – COD10K

Hierarchical Visual Processing Enhances Camouflage Detection

The **human visual cortex** detects camouflaged or low-contrast objects through a **hierarchical cascade** of visual areas:

- **V2** — accentuates **edges, orientation contrasts, and boundaries**, revealing subtle contours.
 - **V4** — integrates **shape and color constancy**, distinguishing objects from textured or noisy backgrounds.
 - **V5 (MT)** — contributes **motion and dynamic texture sensitivity**, aiding detection when objects or backgrounds have temporal variation.
- **Hypothesis**
 - A system modeled on the **V2–V4–V5 hierarchy** will outperform a generic multimodal LLM baseline in **camouflage detection tasks**, due to its progressive refinement from **low-level to high-level** visual cues.
 - **Goal**

Evaluate if biologically inspired hierarchical feature extraction yields **improved detection accuracy** for camouflaged objects.

Experiments – COD10K

MM-CamObj Dataset for Camouflage Detection

Dataset Summary:

- Contains ~12,000 **synthetic and real images** of **camouflaged objects** (animals, shapes, or patterns) blended into complex or noisy backgrounds.
- Each image includes **ground-truth masks or bounding boxes** for object locations.
- Balanced across **varied textures, lighting, and motion cues** (for temporal sequences).
- **Why This Dataset Tests the V2–V4–V5 System:**
 - **V2-level features:** Edge enhancement helps in revealing subtle contour discontinuities.
 - **V4-level processing:** Shape and color constancy disambiguate objects from chromatically similar surroundings.
 - **V5-level dynamics:** Detects motion differences between object and background, useful for temporal camouflage.
 - Together, these mimic the **human cortical pipeline**, making the dataset ideal to test **progressive cue integration** in camouflage detection.
- **Setup:**
 - Sampled **100 images** from the camouflage dataset.
 - Measured detection accuracy for:
 - **Baseline Multimodal LLM** (LLaMA 3.2 90B)
 - **V2–V4–V5 Hierarchical Agent**
 - **Human annotations**

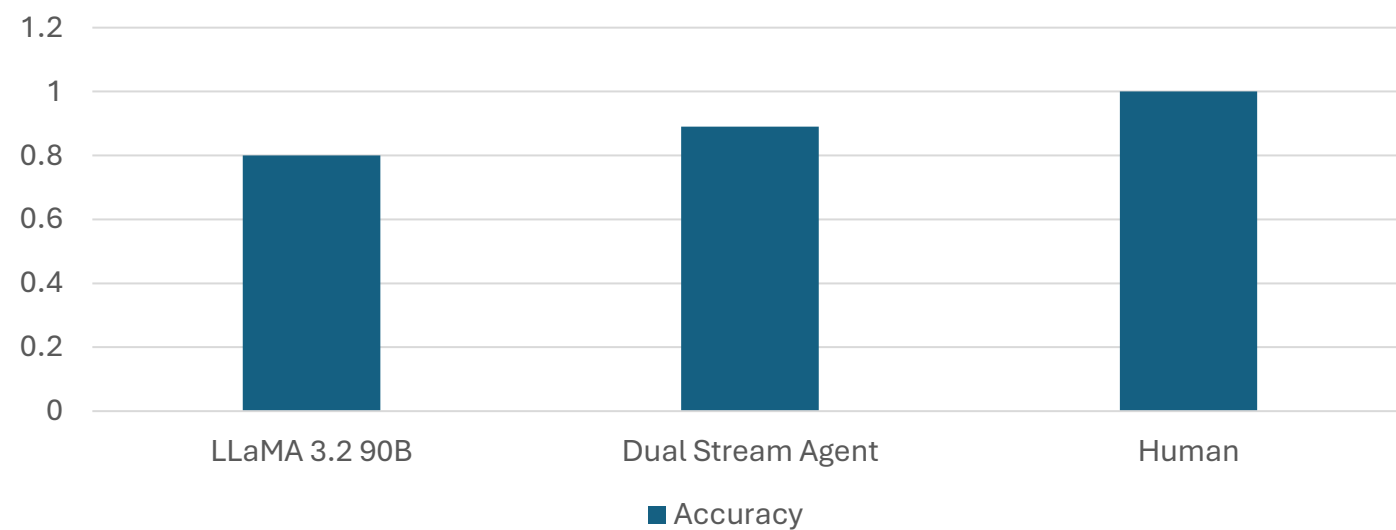


Experiments – COD10K



COD-10K Dataset

Accuracy



Experiments– Visual Saliency

- **Dataset:** Where People Look Dataset (Eye-Tracking) with ~1000 images
- **Type:** Eye-tracking dataset with fixation maps
- **Fixation Maps:** White regions indicate high human attention/intensity
- **Evaluation Method:**
 - Predicted coordinates from model compared against ground truth fixation maps
 - Weighted **precision score** computed for predicted regions
 - Single predicted point compared to multiple intensity regions in fixation maps
- **Purpose:** Assess ability of visual subsystems to predict human-like focus points in images
- **F1:** 0.59 (Using V2-V4-V5 Subsystem)

Experiments – Visual Saliency



Experiments – Visual Saliency

- **V2-V4-V5** excels at detecting intricate features, texture, contrast, and fine details
- **Dorsal-Ventral** captures general spatial regions and motion patterns effectively

F1 Scores of the 2 systems

VISION MODEL	F1 SCORE
Dorsal-Ventral System	0.41
V2-V4-V5 System	0.59

Conclusion

- Topology Manager is capable of orchestrating reasoning agents and achieved similar accuracy to CrewAI (~5% difference)
- Phi-3.5-mini-instruct, after finetuning, achieved 89% accuracy on ReClor and 72% on LogiQA reasoning tasks
- Dorsal-Ventral dual-stream agent reached 0.67 accuracy on CLEVR spatial estimation, compared to 0.48 for a baseline model
- V2-V4-V5 hierarchical agent detected camouflaged objects with 0.89 accuracy on COD10K, surpassing Dorsal-Ventral at 0.80
- Visual saliency prediction yielded an F1 score of 0.59 for V2-V4-V5 and 0.41 for Dorsal-Ventral streams

References

1. [Dennett1994] The Multiple Drafts Model of Consciousness. Daniel C. Dennett. In: Current Issues in the Philosophy of Mind. Oxford University Press. 1994.
2. [Sattin2021] Sattin, D.; Theoretical Models of Consciousness: A Scoping Review. Magnani, F.G.; Bartesaghi, L.; Caputo, M.; Fittipaldo, A.V.; Cacciatore, M.; Picozzi, M.; Leonardi, M. Brain Sci. 2021, 11, 535.
3. [Jeong2024] Domain-specialized LLM: Financial fine-tuning and utilization method using Mistral 7B. Journal of Intelligence and Information Systems, Jeong, C. (2024). Korea Intelligent Information System Society. 2024
4. [Ravuru2024] Agentic Retrieval-Augmented Generation for Time Series Analysis. Chidaksh Ravuru, Sagar Srinivas Sakhinana, and Venkataramana Runkana. arXiv. 2024.
5. Guozhang Chen et al., A data-based large-scale model for primary visual cortex enables brain-like robust and versatile visual processing. Sci. Adv.8, eabq7592(2022). DOI: 10.1126/sciadv.abq 7592
6. Wang, AY, Kay, K, Naselaris, T, Tarr, MJ & Wehbe, L 2023, 'Better models of human high-level visual cortex emerge from natural language supervision with a large and diverse dataset', *Nature Machine Intelligence*, vol. 5, no. 12, pp. 1415-1426. <https://doi.org/10.1038/s42256-023-00753-y>
7. Krassanakis V, Perreira Da Silva M, Ricordel V. Monitoring Human Visual Behavior during the Observation of Unmanned Aerial Vehicles (UAVs) Videos. *Drones*. 2018; 2(4):36. <https://doi.org/10.3390/drones2040036>
8. Wenguan Wang, Jianbing Shen, Jianwen Xie, Ming-Ming Cheng, Haibing Ling, & Ali Borji (2019). Revisiting Video Saliency Prediction in the Deep Learning Era. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1-1.
9. [Hulse2021] A connectome of the Drosophila central complex reveals network motifs suitable for flexible navigation and context-dependent action selection. Brad K. Hulse; Hannah Haberkern; Romain Franconville;
10. J. Ruan, W. Yuan, Z. Lin, N. Liao, Z. Li, F. Xiong, T. Liu, and Y. Fu, "MM-CamObj: A Comprehensive Multimodal Dataset for Camouflaged Object Scenarios," *arXiv preprint arXiv:2409.16084*, 2024, doi: 10.48550/arXiv.2409.16084.