BEGANS

"PIXELS TO PERFECTION"

STRATEGIC HANDLING AND KNOWLEDGE FOR TACTICAL INTELLIGENCE

SHAKTI AI

BEGANS

"PIXELS TO PERFECTION"

OUR TEAM

Ryan (Leader & Core AI Development)

Ronit (Core AI Development)

Udit (Backend & Web Developer)

Anukrati (Frontend & Design)

Murtuza (Data Collection)

Kushal (Data Testing & Validation)

BEGANS
"PIXELS TO PERFECTION"

THE CHALLENGE OF MODERN TACTICAL INTELLIGENCE

- Difficulties in integrating multimodal data for real-time decision-making.
- Lack of AI systems optimized for military applications (e.g., threat detection, situational awareness).
- Problems like false positives, inefficient data handling, and lack of resource-efficient solutions.
- Statistics or real-world examples (e.g., challenges faced in dynamic battlefield environments).

BEGANS

"PIXELS TO PERFECTION"

AI FOR DEFENSE & TACTICAL APPLICATIONS

OBJECTIVES

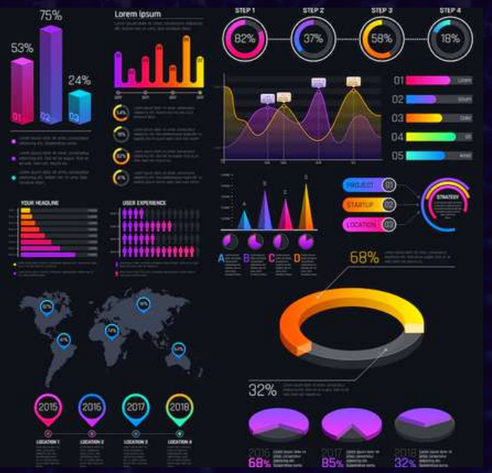
01 To develop a modular, resource-efficient AI capable of processing and integrating multimodal data.

02 Enhance decision-making, threat detection, and situational awareness in military contexts.

03 Minimize computational costs while maximizing tactical intelligence accuracy.

BEGANS

"PIXELS TO PERFECTION"



OVERVIEW

01

Modular AI system combining Drishti Vision Module (visual data), Chakravyuha Text Module (language processing), and Airavata Framework (pipeline management).

02

- Cross-hyper attention for vision-language fusion.
- Iterative and parallel reasoning for complex decision-making.
- Lightweight design optimized for hardware with limited resources.

03

- Cross-modal integration.
- Flash attention and hyper-attention for efficient processing.

BEGANS

"PIXELS TO PERFECTION"

CORE FEATURES

01 Iterative & Parallel Reasoning: Step-by-step and multi-stream processing.

02 Rotary Positional Encoding (RoPE): Improved spatial and temporal reasoning.

03 Camouflage Detection: Distinguishing hidden threats.

04 Friend/Foe Classification: Military-specific feature for battlefield clarity.

05 Supported RAG as well as long sequence queries along with OCR detection.

06 Low-Rank Adaptation (LoRA): Efficient model training and deployment.

BEGANS

"PIXELS TO PERFECTION"

Benchmark	Shakti (2B)	Top Model 1	Top Model 2	Top Model 3
Visual Question Answering				
VQAv2	80.1	GPT-4V (1.8T): 87.6	LLaVA-1.5 (7B): 85.4	Qwen-VL-Plus (4.4B): 79.8
OK-VQA	54.2	BLIP-2 (4.4B): 62.8	InstructBLIP (7B): 58.6	Kosmos-2 (2.7B): 56.7
TextVQA	62.9	GPT-4V (1.8T): 76.5	LLaVA-1.5 (7B): 73.2	MPLUG-Owl2 (7B): 68.4
Zero-Shot Multimodal				
MMB-EM	66.3	GPT-4V (1.8T): 76.8	Gemini Pro (175B): 72.5	LLaVA-1.5 (7B): 69.1
POPE	87.4	Qwen-VL-Plus (4.4B): 90.1	BLIP-2 (4.4B): 88.7	InstructBLIP (7B): 87.6
AI2D	63.2	GPT-4V (1.8T): 72.5	Gemini Pro (175B): 69.8	LLaVA-1.5 (7B): 67.3
Video Understanding				
NextQA	73.6	VideoChat2 (7B): 81.3	BLIP-2 (4.4B): 78.5	MPLUG-Owl2 (7B): 76.9
MLBU	60.1	VideoChat2 (7B): 68.4	BLIP-2 (4.4B): 65.7	LLaVA-1.5 (7B): 63.2
Multi-Image Benchmarks				
NLVR2	85.6	GPT-4V (1.8T): 92.3	LLaVA-1.5 (7B): 89.7	Qwen-VL-Plus (4.4B): 87.2
BLINK	42.7	BLIP-2 (4.4B): 51.6	InstructBLIP (7B): 48.9	Qwen-VL (4.4B): 45.3

Table 5: Multimodal Model Performance Comparison

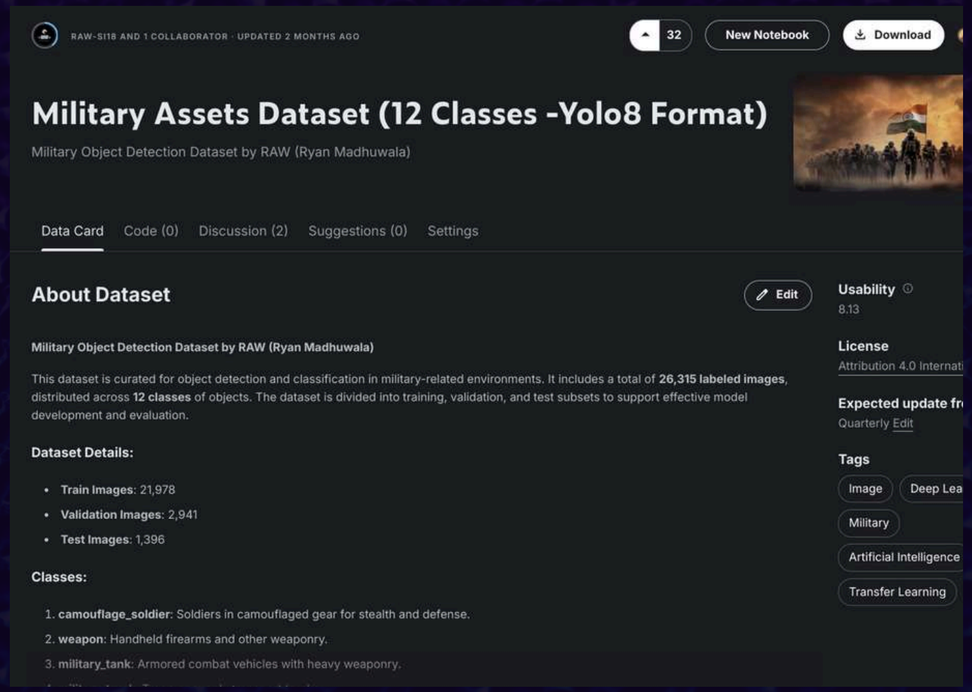
METRICS

- Image processing speed: 15-20 FPS for single images, 5-7 FPS for multi-image grids.
- Accuracy: 80%+ on vision-language benchmarks.
- Resource requirements: Runs on GPUs with as little as 5GB VRAM.
- Resource-efficient yet achieves good results in multi-image and video scenarios.

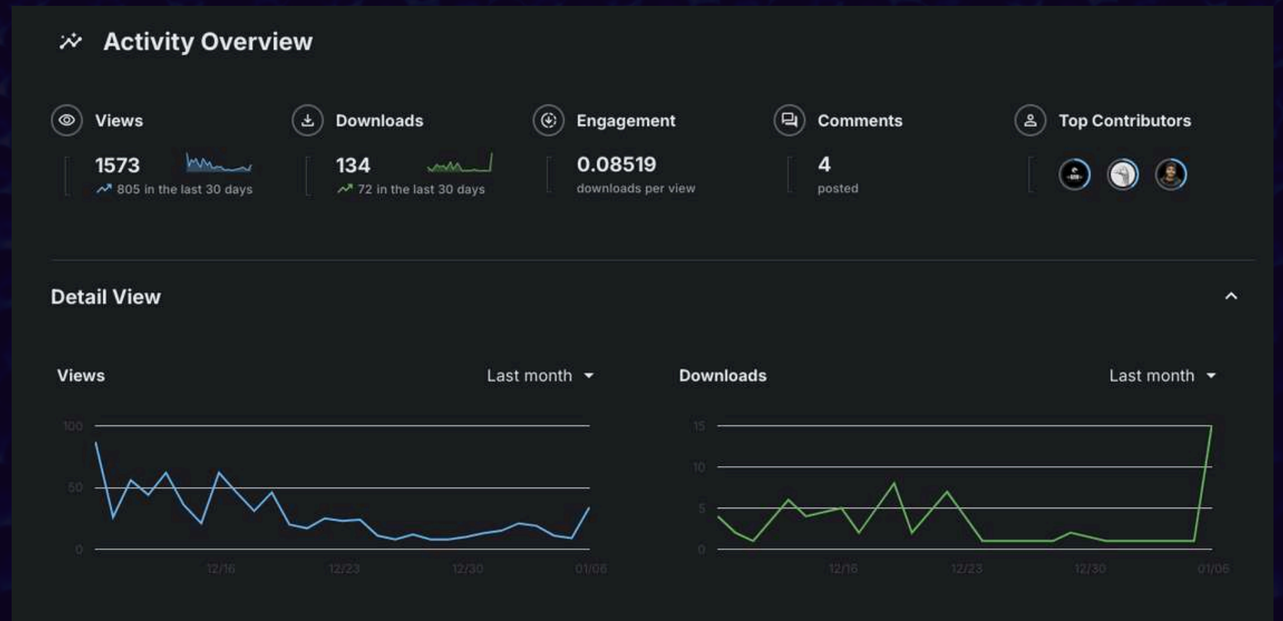
BEGANS

"PIXELS TO PERFECTION"

TRAINED DATASETS



- 17,000 Military Assets Images: Tanks, aircraft, ships, and other vehicles.
- 9,000 Camouflage Detection Images: Hidden objects for threat identification.
- 2,000 OCR Documents: Military text for document analysis.
- 1,500 Tactical Scenarios: Multi-modal cases combining images, videos, and text.



OWN MADE (KAGGLE BRONZE MEDAL) DATASET

BEGANS

"PIXELS TO PERFECTION"

CURRENT SOLUTIONS:

- Lack integration of multimodal data streams for holistic analysis.
- Focus on single-task analysis (e.g., object detection or text reasoning).
- High resource demands, making them impractical for on-field deployment.

SHAKTI AI SOLUTION:

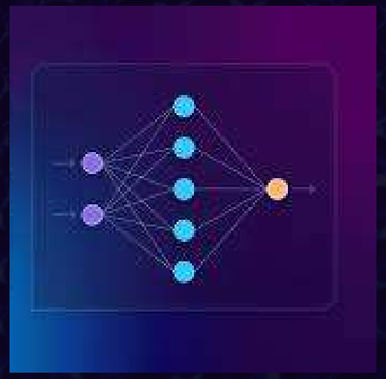
- Fully multimodal (images, videos, and text).
- Tailored for military applications with domain-specific optimizations.
- Lightweight and deployable in resource-constrained settings.
- Innovative attention mechanisms to ensure minimal latency and high accuracy.

BEGANS

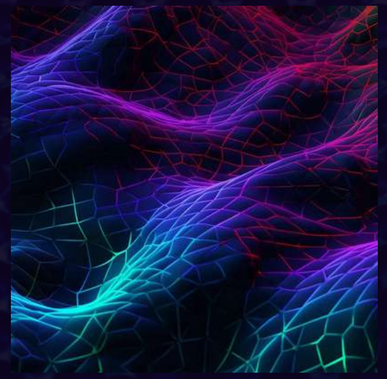
"PIXELS TO PERFECTION"

TECH STACK

Frameworks: PyTorch, TensorFlow, Hugging Face Transformers.



Models: Vision Transformers (Drishti), Text Reasoning Models (Chakravyuha).



Attention Mechanisms: Cross-Hyper Attention, Flash Attention, Iterative & Parallel Reasoning.

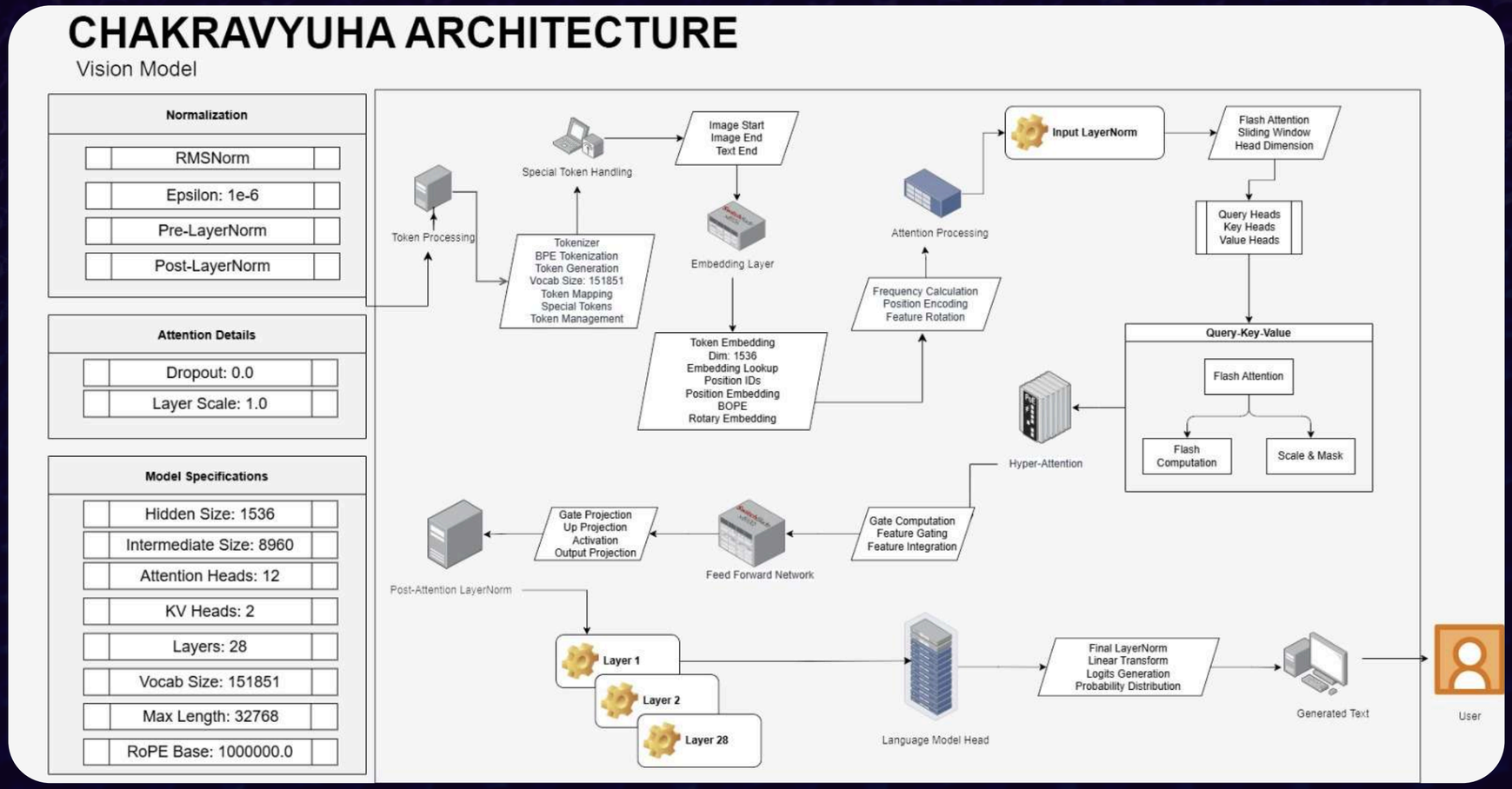
Optimization: Low-Rank Adaptation (LoRA), Rotary Positional Encoding (RoPE).

Backend & Infrastructure: Airavata Framework, Python APIs, NVIDIA GPUs (5GB+).

BEGANS

"PIXELS TO PERFECTION"

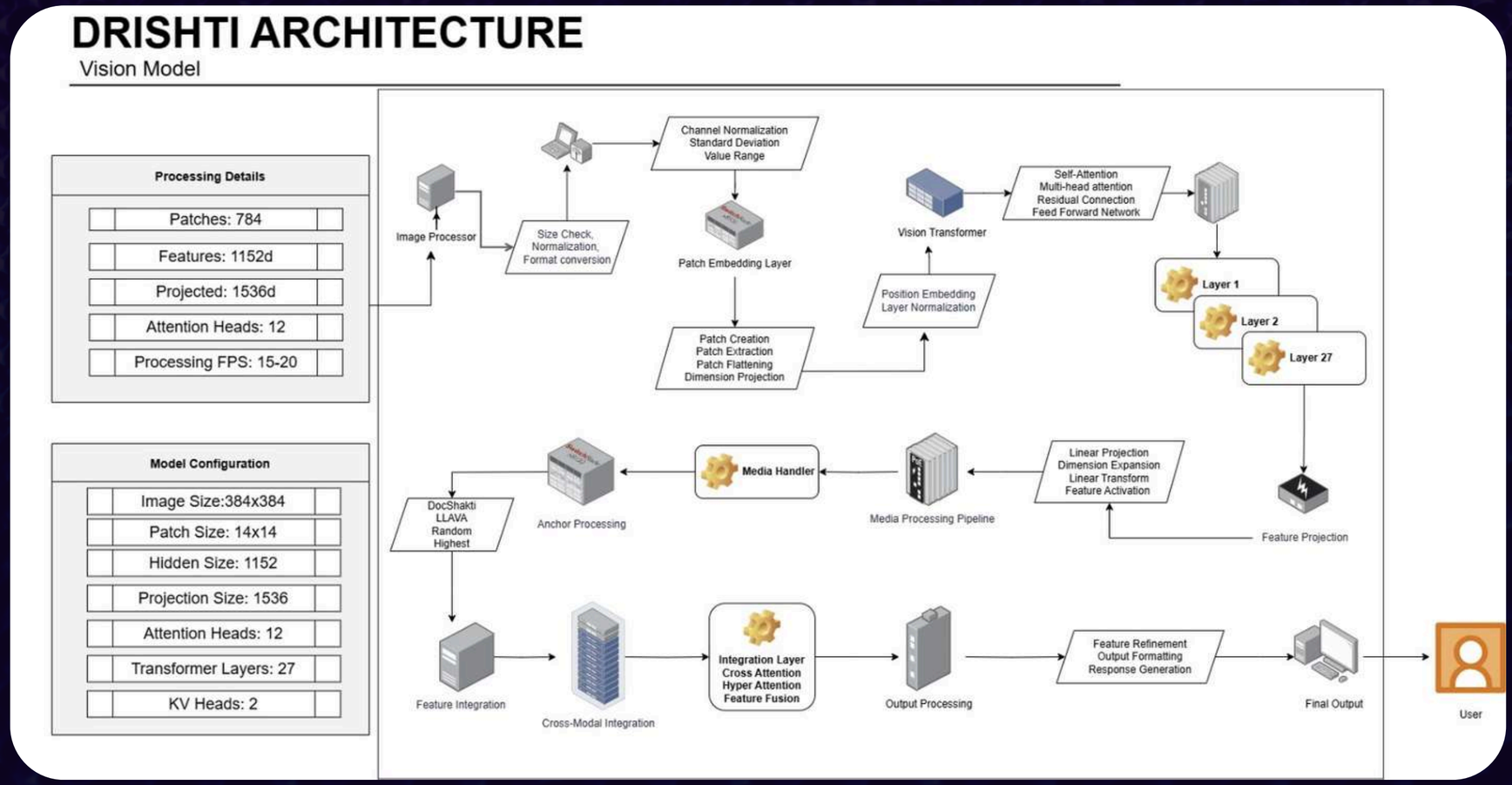
CHAKRAVYUHA MODEL



BEGANS

"PIXELS TO PERFECTION"

DRISHTI MODEL



BEGANS

"PIXELS TO PERFECTION"

WORKFLOW

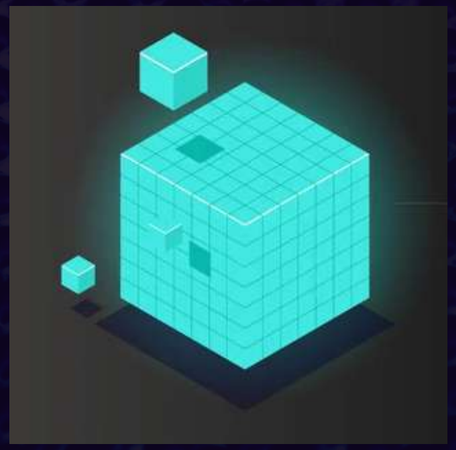
- Input data: Images, videos, pdf, and textual commands.
- Data processing: Vision and language fusion via hyper cross-modal layers.
- Output: Tactical insights like threat levels, camouflage detection, and decision support.

BEGANS

"PIXELS TO PERFECTION"

SPECIALITY

First-of-its-kind military-specific multimodal intelligence.



Modular and scalable architecture.

Lightweight design capable of running on 5GB VRAM GPUs.

Use cases in dynamic and resource-constrained environments

BEGANS

"PIXELS TO PERFECTION"

THANK YOU

DEV.BEGANS@GMAIL.COM