# Aryan Sharma

Riverside, CA aryan250403@gmail.com 1-951-830-5938 LinkedIn GitHub Google Scholar

## EDUCATION

**Master of Science in Computer Science,** University of California Riverside, CA **Sep 2025 – Present** (Expected Jun 2027)

**Bachelor of Technology in Computer Science and Engineering,** VIT Vellore, India **(GPA: 3.9/4)** **Sep 2021 – Jun 2025**

## WORK EXPERIENCE

**Cybersecurity and Data analysis Intern**, Redinent Innovations, Bengaluru **Sep 2023 – Dec 2023**

- Aggregated banner fingerprints from exposed interfaces of IOT devices using dorks, Censys, Shodan, and ZoomEye, expanded the database by over 1200+ entries, achieving 25% growth.
- Pinpointed 17 vertical parameters to uniquely identify login pages for specific subgroups of devices.
- Compiled stable attributes such as tuples through recursive analysis of login pages and reduced false matches by 25%, improving identification accuracy to 91%.
- Delivered a clean, usable dataset mapped to device identifiers and helped streamline downstream tasks like external attack surface monitoring, increasing the number of devices monitored by 35%.

**3D Modeling, Animations, and Game Development Intern**, Vifr Technologies, Vellore **Feb 2023 – May 2023**

- Architected and refined 25+ environmental, character, and animated assets every week, improving the assets library by 275%.
- Spearheaded 5-person team to deliver 10 core features, completing the project 3 weeks before schedule.
- Pioneered shader-based solutions in Unity to replace simulation-based methods for environmental objects; enhanced runtime efficiency by 60% and reduced draw calls by 150 per frame.
- Engineered pipelines integrating accessibility features within the game, resulting in a 65% surge in user engagement among children with neurodevelopmental disorders during initial user testing.

## TECHNICAL SKILLS

| | |
|---|---|
| **Programming Languages:** | Python (Advanced), C++ (Intermediate), Java (Intermediate) |
| **Technical Expertise:** | ML, DL, Data Science, Natural Language Processing, Computer Vision, GenAI, Retrieval Augmented Generation |
| **Tools And Libraries:** | LangChain, Tensorflow, Pandas, Numpy, Matplotlib, BeautifulSoup, Pytorch, Transformers |
| **Databases:** | SQL |
| **Front-end Tools:** | HTML, CSS, JavaScript, React, Vite |

## PUBLICATIONS

**A cross dataset meta-model for hepatitis C detection using multi-dimensional pre-clustering** **Mar 2025 (Published)**

***Aryan Sharma***, Tanmay Khade & Shashank Mouli Satapathy

*Scientific Reports (5-year IF: 4.3 (2024))*

## PROJECTS

**SmartSubs: Context-Aware Smart Subtitle Engine** | *Python, OpenCV, YOLOv8, PySceneDetect* **Feb 2026 - Present**

- Engineered an automated video processing pipeline that dynamically calculates optimal subtitle placement across a 6-zone spatial grid to prevent the occlusion of faces and critical action.
- Implemented YOLOv8 object detection with a strict <15% Intersection over Union (IoU) veto threshold to create dynamic "cost heatmaps" across 1080p video frames.
- Designed a temporal aggregation algorithm that samples every 5th frame (achieving an 80% reduction in processing time) to enforce subtitle stability and eliminate visual jitter within continuous scenes.
- Built a conversion engine to parse standard .srt files and output .ass format with precise, frame-accurate coordinate tags, maintaining a 20-pixel safety margin around detected subjects

**InfoFetch-AI: Production-Ready RAG System with FAISS Vector Database** **Sep 2024 – Nov 2024**

- Identified optimal k=10 clusters using elbow method and silhouette analysis.
- Reduced search space by 8x through k means pre-clustering before FAISS indexing.
- Reduced storage requirements by 35% through efficient numpy array indexing in FAISS.
- Achieved <500ms retrieval latency for FAISS vector search across 10K+ document embeddings.
- Improved context relevance score from 0.72 to 0.91 (27% improvement) compared to unfiltered retrieval
- Reduced hallucination rate by 40% through accurate source citation and context grounding.
- Achieved 85-92% answer correctness rate using retrieval-augmented context vs. 65% baseline without RAG.
- Outperformed naive keyword search by >50% in answer relevance scores and by over 3x in retrieval speed vs sequential document search.

**Amazon Product Page Entity Extraction** **Sep 2023 – Dec 2023**

- Cross-analyzed the efficacy of an MLLM pipeline to extract entity information from product page images.
- Trained this model on multidimensional text and image input, improving performance against a purely image input by 12%.
- Compared implementations against OCR, Transformer, and Regex-based implementations and observed a 10% improvement in generation time per image.