## Abhay Kumar

Bangalore,Karnataka,India

Email : **abhay8280@gmail.com** | Phone : **+91-9731-812-812**

[LinkedIn Profile](https://in.linkedin.com/in/awesomeabhay) | [GitHub Profile](https://github.com/abhaymise/abhaymise.github.io)

## About Me:

Passionate and accomplished Machine Learning Professional with a decade of hands-on expertise in **Generative AI**,**LLM**, **VLM**, **NLP**, **Computer Vision** and solving complex problems. Proven track record of leading teams, engaging with clients, and contributing impactful research to the field. A dedicated individual with a passion for innovation and a drive to deliver cutting-edge solutions.

## Professional Experience:

### Sr AI Engineer - Target Corporation (Feb 2021 – Present)

* Collaborated with cross-functional teams to design and implement solutions that address business needs and drive value for the organization.
* Led a team of Data Scientists and Machine Learning Engineers, providing mentorship and guidance to enhance their technical skills and professional development.
* Developed and implemented best practices for model development, evaluation, and deployment, ensuring high-quality and scalable Machine Learning solutions.
* Led the development of innovative Machine Learning models and algorithms, leveraging cutting-edge techniques and technologies to solve complex business problems.
* Conducted regular code reviews and technical discussions to ensure the team’s adherence to coding standards and best practices.

#### RAG powered Invoice Intelligence System

* Architected and deployed a scalable invoice search system combining LLM-powered RAG with high-performance vector search (Milvus) and Elasticsearch, handling over millions of financial documents.
* Engineered a hybrid search strategy, combining structured field-based filtering (Elasticsearch) with semantic vector search (Milvus) for high-precision invoice retrieval.
* Spearheaded the selection and tuning of models (OCR,CLIP, LayoutLM, SentenceTransformers) for extracting structured and unstructured metadata, reducing manual effort in invoice processing by 70%.
* Developed anomaly detection models to flag inconsistencies and fraudulent invoices, improving coverage of financial discrepancies and reducing revenue leakage.
* Provided technical leadership on ML architecture, model selection, and productionization, ensuring scalability and alignment with business objectives.

#### AI-Powered Asset Tagging & Retrieval System

* Designed and led the development of an enterprise-grade AI-powered asset tagging and search system, enabling semantic and hybrid search over millions of assets enabling search by images and natural language
* Developed a metadata enrichment pipeline using OCR, object detection, and brand/logo recognition models, generating structured metadata for improved filtering.
* Built a multi-modal retrieval pipeline, leveraging CLIP and Sentence-BERT embeddings for image and text understanding, enhancing searchability by 40%.
* Designed a scalable hybrid search architecture, leveraging Milvus for vector search and Elasticsearch for metadata-based filtering and reranking.
* Built a high-throughput ingestion pipeline for batch & incremental asset tagging using Kafka, Spark, and distributed deep learning inference.
* Led model optimization efforts, reducing embedding computation latency and improving retrieval speed for large-scale datasets.

### Senior Data Scientist - Embibe (May 2019 – Jan 2021)

#### Question Tagging and Topic Extraction System

* Designed and implemented an automated system to tag questions with subjects and extract relevant topics from text.
* Developed question tagging using both text and image content, enabling better content recommendation and search accuracy.
* Used BERT model for text embeddings extraction and Resnet based architecture for visual feature extraction.

#### Query Intent Classification and Routing

* Developed a system to classify queries based on intent, enabling efficient routing to the appropriate department.
* Identified questions subject and routed the customer query to the relevant SME or department.
* Improved customer query resolution time by 20% through efficient routing.

### Senior Data Scientist - ClipIndia (Sep 2018 – Apr 2019)

#### NSFW Video Classification

* Implemented a classification system to identify and filter out inappropriate content, reducing NSFW content
* Developed a video classification system, reducing inappropriate content by 28%.

#### Video Asset Tagging

* Developed a video asset tagging system to improve content discoverability and feed ranking.
* Enhanced content organization and searchability, increasing user engagement by 3%.

### Senior Data Scientist - Cogknit Semantics (Sep 2015 – Sep 2018)

#### Digital catalogue misplaced product detection

* Developed a computer vision-based system to detect and correct misplaced products in an e-commerce catalog using CNN models.
* Identified and reorganized 12% of misplaced products, resulting in substantial reductions in incorrect placements.

### Sr. Research Analyst - Edureka (Dec 2013 – Sep 2015)

* Led the initiative to create a comprehensive knowledge base, reducing the workload on the TA team and enhancing the overall efficiency of support and learning resources.

## Education:

* Bachelor in Computer Science, Visvesvaraya Technological University (VTU), 2013.

## Skills:

* **Generative AI**: Advanced Prompt engineering, LLM app development, Agents, RAG, LangChain, ChatGPT, Bert, LLAMA
* **Search Systems**: Elasticsearch, Milvus, FAISS
* **Computer Vision**: Object detection, image classification, image segmentation
* **NLP**: Text classification, topic modeling, Intent classification
* **Machine Learning**: Deep Learning (PyTorch, TensorFlow), Scikit-learn, Huggingface
* **Applied ML** : Model training, optimisation and deployment at scale, Algorithm research and selection
* **Data Processing**: NumPy, Pandas, Spark, Hadoop
* **Cloud Platforms**: AWS, Azure, Google Cloud
* **MLOps**: Git, CI/CD, MLflow, containerization (Docker), APIs
* **Team Skills**: Team building, leadership, project planning, data storytelling

## Publications and Presentations:

* Poster presentation on “Domain Adaptation of Image Caption Model for Video Descriptions” at AI & Deep Learning Conference (GTC 2018, NVIDIA).