

Yugal Kumar

Gurugram . India . yugal.ahir@gmail.com . +91 9999204178 . [LinkedIn](#) . [yugalkumar.com](#)

AI Technical / Functional Skills:

- * **Programming** - Python, R
- * **Agentic AI & AI Agents** - Developing single- and multi-agent systems with LangChain, LangGraph, CrewAI, AutoGen, implementing tool-calling, memory management and vector-store integration for autonomous task execution.
- * **Microsoft AI & Cloud Ecosystem** – MS Copilot (Agentic Agents, Automation), MS Fabric (Data Agents), Azure Cloud
- * **LLM Ops** - Prompt Engineering, RAG (Retrieval-Augmented Generation), Fine-Tuning, Model Deployment, Chat GPT, GROK, DeepSeek
- * **Vector Databases** - Azure AI Search, Pinecone, FAISS
- * **Machine Learning** - Classification, regression, clustering, feature engineering, model evaluation, and hyper-parameter tuning
- * **Deep Learning** - Designing, training, and deploying neural-network models with TensorFlow, Keras, and transfer-learning workflows (CNNs, RNNs, attention)
- * **Predictive Modelling** - Data Mining, EDA, Sales Forecasting, ARIMA, Stats Models, RFM
- * **Project & Team Leadership** - Managed cross-functional AI/data teams, ensuring delivery of high-value AI projects
- * Mentoring, technical guidance, strategy planning
- * **Certificates** -
 1. **IBM** Certified Badges
 2. **IIM Ahmedabad** Leadership Skill Certified
 3. **Coursera** Data Science Professional certificate
 4. **Udemy** Python for data science and machine learning, Jose Portilla
 5. **Udemy NLP** Natural language processing by Jose Portilla
 6. **Cognitive class.ai** Python
 7. **Management & Strategy Institute** Project Management Essentials
- * **Data validation** - A/B testing, K-fold cross-validation
- * **Power BI** - Data Visualizing & Reporting
- * **Data Storytelling** - Presentations, Visuals
- * **NLP (Natural language processing)**: NLTK, SPACY, Regex, Word2Vec
- * **Chat-Bot** - RASA, WhatsApp integration.
- * **Database** - MySQL, Neo4j Graph Database, Mongo DB, Big Query DB, PostgreSQL

Education:

- | | |
|---|---|
| <ul style="list-style-type: none">* Post Graduation - Master of Computer Application (2012 - 2015) IITM, New Delhi. Affiliated to GURU GOBIND SINGH INDRAPRASTHA UNIVERSITY* High School (Class 12th - Year 2009) ROTARY PUBLIC SCHOOL, Gurugram. Affiliated to CBSE | <ul style="list-style-type: none">* Graduation - Bachelor of Computer Application (2009 - 2012) JIMS, New Delhi. Affiliated to GURU GOBIND SINGH INDRAPRASTHA UNIVERSITY* Secondary School (Class 10th - Year 2007) ROTARY PUBLIC SCHOOL, Gurugram. Affiliated to CBSE |
|---|---|

Work Experience: 10 Years

Lead Data Scientist – MGF Sourcing - mgfsourcing.com (June 2022 - Present)

- **AI-Driven Factory Risk Assessment Agent** | Python, Azure Table, GPT-5 Mini, RAG, MongoDB. Automated factory risk assessments, integrating internal analytics with external intelligence. Performed 55+ compliance checks, generated instant risk scores, and reduced assessment time from **several days to real-time**—saving hundreds of analyst hours annually and accelerating decision-making.
- **LLM-Powered MS SQL Chatbot (DFIX Bot)** | Python, ChatGPT-4, Azure OpenAI, MS Teams. Built a chatbot to convert natural language into SQL, allowing non-technical staff to self-serve data in Teams. Eliminated SQL training needs, **reduced ad-hoc reporting requests by 70%**, and saved significant analyst effort.
- **Fabric Data Agent + Copilot Conversational Analytics** | MS Copilot, MS Fabric, Azure SQL, Azure Cloud, MS Teams
Enabled natural-language access to Azure SQL data by integrating MS Fabric Data Agent with MS Copilot and deploying the solution inside MS Teams. Configured full environment, security, SQL mirroring, and agent orchestration, allowing users to ask English questions and instantly retrieve insights without writing SQL—reducing reporting dependency and accelerating decision-making across teams.
- **GSD Email Ticket Automation System** | MS Copilot, Azure OpenAI, Azure AI Search, Jira API, Outlook
Developed an AI-powered automation system that processes incoming GSD emails, matches the message content to SOPs using Azure AI Search, and uses Copilot + Jira API to auto-route requests and create tickets for the correct team. Eliminated manual triage, improved response speed, and ensured consistent, SOP-aligned ticket creation across the service desk.
- **Garment Image Comparison & Search** | Deep Learning, FAISS, FASTAPI
Developed a high-speed image similarity search, automating garment lookup and improving design traceability. **Reduced manual image matching time by 95%, saving \$50K/month in labor costs.**

- Text-to-Image Search System** | Python, OpenAI CLIP
 Created a CLIP-powered search tool matching text descriptions to 55K+ images. **Cut search time from hours to seconds**, improving designer productivity and saving \$15K/month in operational costs.
- LLM-Powered Document Chatbot (DFIX Docs)** | Python, Azure OpenAI, Azure Cognitive Search
 Built a document chatbot to instantly query 10K+ files and SQL data. **Reduced document search time from Hours to under 1 minute**, eliminating manual research and speeding executive decision-making.
- TRACE AI – Intelligent Document Verification Agent** | Python, NLP, CNN, Azure OCR
 Automated verification of logistics documents with NLP, OCR, and CNN models. **Cut border clearance delays, reduced manual checks, and improved compliance accuracy by over 90%.**
- Sales & Demand Forecasting System** | Python, Power BI, Data Science
 Developed predictive models with 96% accuracy, integrated into Power BI dashboards. **Optimized inventory, reduced stockouts by 40%, and increased revenue via proactive sales strategies.**
- SharePoint Data Extraction & Image Warehouse Integration** | Python, SharePoint API, MongoDB
 Automated SharePoint data and image ingestion into MongoDB. **Removed manual file handling, enabling instant access for AI workflows and saving hours per week for operations teams.**
- Document Data Extraction & Storage Pipeline** | Python, OCR, MongoDB
 Created an OCR-based ingestion pipeline to process PDFs and Word files. **Reduced data preparation time by 80%** and ensured clean, scalable datasets for analytics.
- Web Scraping & Data Visualization** | Python, Power BI, EDA
 Automated large-scale data extraction and dashboard creation. Cut reporting time from days to hours and enabled faster, insight-driven decisions for stakeholders.
- Optimizing Logistic Shipping Schedules (Supply Chain AI)** | Python, ML, Route APIs
 Applied ML to optimize transport routes. **Reduced logistics costs by up to 15%**, improved delivery speed, and maximized resource allocation.
- Social media & Customer Feedback Insights** | Python, MongoDB, EDA
 Automated sentiment and feedback analysis pipelines. Enabled real-time product trend tracking, cutting manual review time by 90% and guiding marketing strategies.

Data Scientist - Ushyaku Software Solutions - ushyaku.com . A Partner of UCS Solutions Dubai UAE www.ucssolutions.com (Jan 2019 – May 2022).

- **Advanced OCR Engine for Invoice & Receipt Processing** | TensorFlow, Google Vision, Flask, Neo4j
Led development of a 97.7%-accurate OCR engine for invoices and receipts. Built deep learning models for classification, object detection, and brand recognition. **Automated manual work for 12 employees, saving ~\$110K/month** and significantly improving processing speed.
- **Leadership in Data Science Project Management** | AI & Data Projects
Directed multiple cross-functional teams through complex AI/data project lifecycles. Provided strategic vision, resource allocation, and performance optimization, ensuring on-time, high-value project delivery.
- **Rasa AI Chatbot for Order Management** | RASA, WhatsApp Integration
Developed a conversational chatbot for order confirmations, rescheduling, tracking, and FAQs. Integrated with WhatsApp for **real-time customer support automation**, reducing manual intervention and enhancing service efficiency.
- **OCR Model Training Interface (RPA Engine)** | Flask, Neo4j, MySQL, jQuery
Built a Python-based backend with a custom RPA interface for OCR model training. Managed a responsive web frontend and integrated Neo4j/MySQL for efficient data mapping—streamlining the model training process and reducing setup time.
- **Real-Time Data Upload & Processing** | Apache Kafka, Python
Designed and deployed a Kafka-based real-time data ingestion system for brokers. Created fault-tolerant producers/consumers, **cutting manual work for 2 employees and saving ~\$5K/month**.
- **Doc Worx: PDF to JSON Parsing System** | Django, AWS, PostgreSQL
Engineered a scalable parser to convert PDFs into JSON using Tabula, Camelot, and custom logic. Solved multi-line table issues and deployed APIs on AWS—automating a previously manual data extraction process.
- **Sales Worx: Data Analysis & Customer Segmentation** | Pandas, Flask, K-Means
Built an analytics platform that boosted sales by 70% with 95.5% accuracy in predictions. Applied RFM modelling and clustering to identify top products and high-value customers—driving targeted, high-ROI campaigns.
- **Customer Analysis for Reason8 Campaigns** | Big Query, Neo4j
Unified multi-channel customer data into Big Query and Neo4j for graph-based analysis. Identified coupon usage trends and segmented customers for precision-targeted marketing campaigns.
- **Python Scripting & Automation** | Python, Regex, File Processing
Developed automation scripts for extracting and analysing log files, Word, PDF, and text documents—reducing repetitive processing time and improving data accessibility.

Software Engineer in Data Science - AuthBridge Research Pvt. Ltd - authbridge.com (June 2016 – Jan 2019).

- **Efficient Data Extraction from ID Cards** | TensorFlow, Object Detection, CNN
Developed a **99% accuracy deep learning model** for extracting data from Aadhaar, PAN, Voter ID, and Passports. Automated classification and OCR workflows using CNN-based pipelines, **reducing manual effort for 4 team members and saving ~\$5.5K/month**, while ensuring secure, scalable identity data processing.
- **RFM-Based Student & Employee Analysis** | Python, RFM Modelling, Data Science
Built RFM-based behavioural analytics models to identify high-risk students, employee attrition likelihood, and top performers. Enabled data-driven retention strategies through actionable segmentation insights, improving organizational performance and engagement.
- **Sales Forecasting for Student Services Bureau** | ARIMA, Time Series Forecasting
Applied ARIMA time-series models to forecast sales demand with high precision. Improved strategic resource allocation and decision-making by enabling proactive responses to projected market trends, boosting operational efficiency and business confidence.
- **Fraud Detection at WorkAtTest.com** | SVM, Decision Trees, Machine Learning
Implemented an ML-powered fraud detection system using SVM and Decision Trees to monitor employee activity. Provided early fraud alerts to compliance and risk teams, strengthening organizational security and protecting against operational losses.
- **Real-Time Data Streaming & Analytics** | PySpark, Apache Kafka, MySQL
Designed an enterprise-grade, real-time analytics pipeline using PySpark and Kafka to process high-velocity data streams. Delivered low-latency insights and improved scalability for critical business decisions, enabling faster, AI-ready data processing across the enterprise.

Software Developer Python- Gamexs (Krieda Solutions) Now Merged With gameloot.in (November 2015 – April 2016)

- **Website Development Project** | Django, Web Development
Developed a full-stack web application using Django, implementing dynamic content management and secure authentication. Designed for **scalability, maintainability, and integration with AI/ML-powered backend services** to support future analytics and automation features.
- **REST API Development** | Flask, Python
Built **RESTful APIs** in Flask for seamless data exchange between applications. Optimized endpoints for **low-latency, AI-ready data delivery**, enabling integration with machine learning and analytics pipelines.

- **Python Scripting for Player Data Matching** | Python, Data Processing
Developed automation scripts to **match and reconcile player data** across multiple sources. Improved data accuracy eliminated manual processing, and prepared clean datasets for analytics and predictive modelling.
- **Certification – CS231N: Convolutional Neural Networks** | Stanford University
Completed Stanford's CS231N course, gaining advanced expertise in CNN architectures, deep learning, and computer vision—applying these techniques to real-world OCR, object detection, and image recognition projects.

Associate Software Developer (Python) - VdoIT technologies vdoitech.com (May 2015 – October 2015)

- **Web Application Development** | Python, Full-Stack
Designed and developed full-stack web applications with Python-based backends and responsive frontends. Focused on scalable architecture and seamless integration with AI/ML-powered data services for analytics and automation capabilities.
- **Website Management & Server Administration** | GoDaddy, BigRock
Managed and optimized web hosting, domains, and server configurations for high availability and security. Implemented performance tuning and monitoring to ensure reliable, enterprise-grade web infrastructure.
- **API Development for Android Integration** | Flask, REST APIs
Built RESTful APIs in Flask to integrate Android applications with backend databases and AI-driven services. Designed endpoints for low-latency, scalable data exchange to support real-time mobile functionality.
- **Team Leadership & Mentoring** | Project Management, Technical Guidance
Led cross-functional teams through AI, data science, and software development projects, providing technical mentorship, project planning, and performance optimization to ensure on-time, high-impact delivery.