

Yugal Kumar

Lead Data Scientist



IBM & Coursera
Data Scientist Certified



IIM - Ahmedabad
Leadership Skills Certified

☎ (+91) 9999 204 178

✉ yugal.ahir@gmail.com

[LinkedIn](#)

[GitHub](#)

📍 Sector 23A, Gurugram, Haryana, India, 122017

Professional Summary:

AI Solutions Architect with expertise in developing custom AI automation solutions, Python AI Projects, AI agents, multi-agent systems, and LLM-based automation. Expert in Python, Azure AI Foundry, Azure OpenAI, and integrating enterprise data with Agentic AI pipelines for processing automation and decision intelligence.

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, planning/reasoning, and vector-store integration for autonomous task execution.
- * **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.
- * **Database** - MySQL, Neo4j Graph Database, Mongo DB, Big Query DB, PostgreSQL.
- * **Project & Team Leadership** - Managed cross-functional AI/data teams, ensuring delivery of high-value AI projects
- * Mentoring, technical guidance, strategy planning
- * **Data Science Certificates** -
 1. **IBM Certified Badges** ([Certificate Link](#))
 2. **Coursera** - Data Science Professional certificate (9 Courses) ([Certificate Link](#))
 3. **Udemy** ([Certificate Link](#)): Python for data science and machine learning, Jose Portilla
 4. **Udemy NLP** ([Certificate Link](#)): Natural language processing by Jose Portilla
 5. **Cognitive class.ai Python** ([Certificate Link](#))
 6. **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.

Education:

- * **Post Graduation - Master of Computer Application (2012 - 2015)** IITM, New Delhi. Affiliated to GURU GOBIND SINGH INDRAPRASTHA UNIVERSITY
- * **Graduation - Bachelor of Computer Application (2009 - 2012)** JIMS, New Delhi. Affiliated to GURU GOBIND SINGH INDRAPRASTHA UNIVERSITY
- * **High School (Class 12th - Year 2009)** ROTARY PUBLIC SCHOOL, Gurugram. Affiliated to CBSE
- * **Secondary School (Class 10th - Year 2007)** ROTARY PUBLIC SCHOOL, Gurugram. Affiliated to CBSE

Work Experience: 10 Years

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

1. **AI-Driven Factory Risk Assessment Agent** | *Python, Azure Table, GPT-4.1 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.
2. **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.
3. **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 \$17K/month in labor costs.**
4. **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.
5. **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 2 days to under 1 minute**, eliminating manual research and speeding executive decision-making.
6. **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%.**
7. **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.**
8. **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.**

9. **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.
10. **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.
11. **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.
12. **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 - www.ushyaku.com . A Partner of UCS Solutions
Dubai UAE www.ucssolutions.com (Jan 2019 – May 2022).***

1. **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 ~\$11K/month** and significantly improving processing speed.
2. **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.
3. **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.
4. **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**.
5. **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**.
6. **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**.

7. **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**.
8. **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**.
9. **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 - www.authbridge.com (June 2016 – Jan 2019).

1. **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.
2. **RFM-Based Student & Employee Analysis** | *Python, RFM Modelling, Data Science*
Built **RFM-based behavioral 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.
3. **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.
4. **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.
5. **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 www.gameloot.in (November 2015 – April 2016)

1. **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.
2. **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.
3. **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.
4. **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 www.vdoitech.com (May 2015 – October 2015)

1. **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.
2. **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**.
3. **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.
4. **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**.