Abdel-Fatah Karam

PHD CANDIDATE — LLM EVALUATION, DOCUMENT QA & RAG

Cairo, Egypt | +20 109 575 5500 | a_karram@pg.cu.edu.eg | abdel-fatah-karam | Google Scholar

RESEARCH INTERESTS

Measurement validity in the evaluation of large language models: whether a reported score can be read as a verdict at all for the kind of answer it summarises. Current work builds a small, reproducible validity audit for document-based question answering and retrieval-augmented generation that decides, per dataset and per answer type, when an evaluation conclusion may be reported, qualified, or withheld — a reporting layer above individual scorers rather than a new benchmark. Related interests in semantic evaluator reliability, oracle construct validity and cross-domain generalisation. Earlier work covers non-parametric and instance-based classification, clustering-driven improvements to classical learners, and convolutional architectures for image detection.

EDUCATION

PhD, Software Engineering 2024 – present
Cairo University, Faculty of Graduate Studies for Statistical Research
Thesis: *Harnessing Large Language Models for Document Retrieval QA: A Cross-Domain Performance Comparison.*
Submitted; defence expected January 2027.
Supervisors: Prof. Mervat Gheith (Computer Science) and Dr. Tarek Ali (Computer Science).

Master, Software Engineering 2019 – 2021
Cairo University, Faculty of Graduate Studies for Statistical Research
Thesis: *An Enhanced Approach for Non-Parametric Machine Learning Classifiers.*
Supervisors: Dr. Ammar Mohammed (Computer Science) and Dr. Abdelmonem Helmy (Information Systems).

Postgraduate Diploma, Software Engineering
Cairo University

Postgraduate Diploma, Computer Science
Cairo University

BSc, Information Technology
Cairo Higher Academy

PUBLICATIONS

Peer-reviewed conference papers

[2] A.-F. Karam, A. Helmy, A. Mohammed. **An approach to enhance KNN based on data clustering using K-medoid.** *2nd International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC), IEEE. DOI: 10.1109/MIUCC55081.2022.9781724*, 2022.

[1] A.-F. Karam, M. Embaby, H. El-Kady, S. Abdel-Hafeez, G. Nabil, A. Mohammed. **Applying convolutional neural networks for image detection.** *International Conference on Smart Applications, Communications and Networking (SmartNets), IEEE. DOI: 10.1109/SmartNets48225.2019.9069750*, 2019.

RESEARCH EXPERIENCE

Doctoral Researcher 2024 – Present

Cairo University, Faculty of Graduate Studies for Statistical Research *LLM Evaluation & RAG*

- Built a reproducible validity audit for document-QA and RAG evaluation that decides, per dataset and answer type, whether a reported number can be read as a verdict, with the admissibility of each metric and of each oracle ceiling settled before any result is seen.
- Ran the audit across several open and domain-specific QA datasets, multiple answer generators and multiple retrieval conditions (sparse, dense, hybrid and diversity-based), on thousands of scored predictions.
- Showed where exact-match scoring is invalid for free-form answers and where a naive oracle ceiling fails for classification, numeric and multi-hop tasks, and how a compatible ceiling changes the conclusion.
- Assessed semantic evaluator reliability with inter-rater statistics across an LLM panel with a blind human overlay, and positioned the procedure against existing RAG evaluation tooling as a complement rather than a replacement.

Master's Researcher

2019 – 2021

Cairo University, Faculty of Graduate Studies for Statistical Research — *Machine Learning*

- Designed KMKNN, a clustering-based enhancement to k-nearest-neighbour classification that confines distance computation to the nearest medoid's cluster rather than the full training set.
- Cut prediction time by at least 25% against standard KNN with no loss of accuracy, validated across 15 UCI benchmark datasets and against published variants including cNK, WAKNN, DKNN, KNNDW and fuzzy-rule methods.

SYSTEMS BUILT — MEETUSVR AI GATEWAY

Head of AI — AI Agents, RAG & Machine Learning

2022 – Present

MeetusVR — *Concurrent engagement*

Accountable for AI strategy and for the enterprise AI platform delivering it: a multi-tenant gateway serving commerce, customer service, finance and analytics for client organisations, deployed through a controlled pipeline across development, UAT and production.

Commercial and product architecture

- **Productised, not bespoke.** Structured as tiered service levels — basic, advanced and premium — so capability maps to price and the platform sells rather than being rebuilt per client.
- **Self-service onboarding.** Organisations are activated from the dashboard, with their catalogue, branches, delivery zones and content indexed automatically. Onboarding cost per client is close to zero, which is what makes a multi-tenant model viable.
- **Vendor independence by design.** Built across multiple model providers rather than committed to one, keeping concentration risk, pricing exposure and migration cost under control as the market moves.
- **Tenant isolation.** Each organisation's knowledge is indexed and served in isolation — a commercial and data-protection requirement before it is a technical one.
- **Running-cost discipline.** Only changed material is re-indexed, and consumption is measured per request, so unit economics stay legible as usage grows. AI programmes fail on running cost as often as on capability.

Capability delivered to client businesses

- **Specialised agents under a single router.** Sales, customer support, company information and orders, with intent classification directing each query, rather than one general assistant answering everything adequately.
- **Grounded in the client's own data.** Retrieval-augmented generation over each organisation's catalogue, branches, delivery zones and published content — the difference between an assistant that sells and one that invents.
- **Answers from the system of record.** Natural-language querying of operational data for the structured questions retrieval cannot serve.
- **Commercial intelligence.** Recommendation built on purchase behaviour, and scheduled competitive monitoring of rival catalogues and pricing.
- **Finance document intelligence.** Question answering over uploaded financial documents, scoped per organisation, user and document.
- **Service escalation.** Complaint and sentiment detection triggering human escalation, so the system knows what it should not attempt to handle.
- **Events and operations.** Event management, order enquiry and logistics support integrated with the client's operational endpoints.
- **Arabic as a first-class language.** Detection and orthographic normalisation so Arabic classifies as reliably as English — a differentiator in this market.
- **Embedded in the product.** Integrated into the VR environment itself rather than bolted on as a web widget.

Governance and operational control

- Authenticated access, structured logging and error capture, documented interfaces and health monitoring — the controls that let an AI system be audited rather than merely demonstrated.
- Per-tenant scheduled retraining, with classifier performance measured on accuracy, F1 and Cohen's κ under stratified validation. Model quality is a governed metric, not an assumption.

APPLIED RESEARCH IMPACT

The dissertation questions came out of building the platform described above, not from reading about it. Running multi-tenant retrieval in production forces the problem the thesis formalises: when a system answers from retrieved documents, across answer forms that no single metric suits, how do you establish whether the answer was right at all? Exact match penalises correct free-form output; an oracle ceiling can be invalid for the very task it claims to bound. Those are not abstractions in that setting — they are the daily question of whether a deployed assistant is working. The same pattern holds on the machine-learning side at Tseppas MG, where CNN quality inspection and demand forecasting run in production across 30+ sites and a model's reported accuracy has to survive contact with the production line.

TECHNICAL SKILLS

LLM evaluation	Metric admissibility, oracle construct validity, answer-type-aware scoring, LLM-as-evaluator design, inter-rater reliability (Cohen's κ, Gwet's AC1, Fleiss' κ), balanced accuracy, clustered confidence intervals, ablation design, reproducibility packaging
Retrieval & RAG	Sparse (BM25), dense bi-encoder, hybrid, MMR and DPP retrieval; RAG pipelines and grounding; RAGAs; embeddings; long-context and multi-hop handling; re-ranking and context construction
Production AI systems	Multi-tenant RAG architecture, token-aware chunking, vector stores (ChromaDB), incremental indexing, LLM agent orchestration and routing, natural-language-to-SQL, session memory, recommender systems, LangChain, Flask, Docker, GitLab CI/CD
Models	GPT-4.1, Gemini 2.5, DeepSeek, Claude; BERT, T5, LLaMA; domain-adapted FinBERT, SciBERT, Legal-BERT
Machine learning	Supervised and unsupervised learning, non-parametric and instance-based methods (KNN), clustering (K-medoids, K-means), cluster evaluation (silhouette, Calinski–Harabasz, elbow, gap statistic), ensemble methods, feature engineering, cross-validation, model evaluation
Deep learning	Convolutional neural networks for image detection and classification, LSTM and recurrent architectures, embedding layers, transformer architectures and self-attention, TensorFlow/Keras
Programming & data	Python, C#/.NET, SQL, LaTeX, Git; Oracle, Microsoft SQL Server, MySQL, ETL pipelines, business intelligence

TEACHING & CORPORATE AI PROGRAMMES

Corporate AI Advisor & Lecturer

2021 – Present

Independent practice — *10+ private companies*

Retained by boards and executive teams to establish where AI creates value in their business and what it would take to realise it. Programmes run from board-level investment framing through to engineering practice, and are designed to leave capability behind.

Advisory mandate

- **Separating opportunity from theatre.** The most valuable service rendered is often telling an executive team which AI proposals are not technically viable, and why, before budget is committed.
- **Investment framing.** Prioritisation by return and feasibility, build-versus-buy, cost and risk exposure, and realistic expectations of what a pilot proves and what it does not.
- **Roadmaps that survive contact.** Use-case discovery, proof-of-concept design, and the pilot-to-production path most organisations never complete.
- **Operating model.** Establishing internal AI teams and centres of excellence, so capability accrues to the company rather than to a vendor.
- **Reducing dependence on generic tooling.** Organisation-specific retrieval and knowledge systems in place of off-the-shelf assistants that cannot see company data.
- **Translation.** Bridging executives, business departments, data teams and software engineers — the failure point in most stalled AI programmes.

Curriculum delivered

1. AI fundamentals	AI concepts and business applications; machine learning versus deep learning versus generative AI; supervised, unsupervised and reinforcement learning; AI maturity and organisational readiness; identifying high-value AI opportunities
2. Machine learning	Classification and regression; clustering and customer segmentation; forecasting and predictive analytics; recommendation systems; feature engineering; model training, testing and evaluation; overfitting, bias, accuracy, precision, recall, F1 and ROC-AUC; model deployment and monitoring
3. Data science	Data collection and preparation; cleaning and transformation; exploratory data analysis; statistical analysis; business intelligence and visualisation; data quality and governance; building data-driven decision-making cultures
4. Generative AI & LLMs	How large language models work; enterprise applications of ChatGPT-style systems; prompt engineering; structured output and tool use; LLM APIs and enterprise integration; model selection and evaluation; cost, latency and quality optimisation
5. Retrieval-augmented generation	RAG architecture; document ingestion and chunking; embeddings; vector databases; semantic and hybrid search; re-ranking; context construction; grounded enterprise assistants; hallucination reduction; RAG evaluation and monitoring
6. AI agents & agentic systems	Agent architecture; planning and reasoning; tool and function calling; memory systems; multi-agent architectures; human-in-the-loop workflows; autonomous business processes; agent orchestration; guardrails and approval mechanisms
7. Enterprise AI architecture	Connecting AI with ERP, CRM, e-commerce and internal systems; APIs and microservices; enterprise knowledge bases; data pipelines; AI gateways; model and vendor selection; cloud versus private deployment; security and access control
8. AI for business functions	Sales and marketing AI; customer-service automation; finance and forecasting; HR analytics and recruitment; supply-chain optimisation; inventory forecasting; e-commerce personalisation; recommendation engines; intelligent document processing
9. AI strategy & transformation	AI use-case discovery; prioritisation by ROI and feasibility; AI roadmaps; proof-of-concept design; pilot-to-production strategy; build-versus-buy decisions; AI operating models; establishing internal AI teams and centres of excellence
10. Responsible AI & governance	Data privacy; AI security; hallucination and reliability; model bias; human oversight; AI governance; risk classification; auditability and monitoring; responsible enterprise adoption

Outcomes for client organisations

The programmes move organisations along a defined progression — **data** → **intelligence** → **prediction** → **automation** → **AI agents** → **business transformation** — and are judged on movement along it.

- Management teams able to identify where AI creates measurable value, and to convert business problems into realistic machine-learning use cases.
- Executives equipped to reason about AI investment, cost, risk and expected return rather than defer to vendors.
- Technical teams fluent in modern architectures — large language models, retrieval-augmented generation and agents.
- Corporate data recognised and governed as a strategic asset, with process-automation opportunities identified against it.
- Predictive analytics introduced for forecasting and decision support, and AI applied to customer experience, personalisation and e-commerce.
- Defined transformation roadmaps, and organisation-specific knowledge systems replacing dependence on generic tools.
- Raised standards of AI security, governance, privacy and responsible deployment.
- Organisations moved from experimentation toward production-ready systems.

PROFESSIONAL EXPERIENCE

Twenty years of director- and board-level accountability for enterprise systems, carried in parallel with graduate research. Currently MIS Director of a food production and retail group, reporting to the board with an eight-figure budget across 30+ sites and 1,500+ users, and Head of AI at MeetusVR. Full detail in the professional CV.

Head of AI — AI Agents, RAG & ML | MeetusVR — 2022 – Present
AI strategy and the multi-tenant enterprise AI platform delivering it

MIS Director | Tseppas MG — 2021 – Present
Food production and retail — board reporting line; machine learning in production across 30+ sites

Chief Information Technology Officer | CONSTEC — 2018 – 2021
Construction — established the technology function, systems and team from nothing

Information Systems & IT Director | Heliopolis Company for Housing and Development — 2015 – 2018
Real estate — publicly listed, state-owned

IT Director / IT Manager | Cairo Cotton Center, Cairo Dyeing Center, T&C, Green Hill — 2007 – 2015
Industrial, construction and real estate

ERP Consultant & Application Support | ECTC; MISR 2000 (Microsoft Gold Partner) — 2001 – 2007
Oracle and SAP implementation

LANGUAGES

Spoken	Arabic (native); English (professional working) — all theses, publications and corporate programmes written and delivered in English

REFERENCES

Prof. Mervat Gheith and Dr. Tarek Ali — PhD supervisors and co-authors on the journal manuscript — and Dr. Ammar Mohammed, MSc supervisor and co-author on both IEEE papers. All at the Faculty of Graduate Studies for Statistical Research, Cairo University. Contact details on request.