The delivery record

35 published records, organised by the shape of the problem rather than by the client's industry.

One Industry AI · Industrial AI for the Gulf back office · First Settlement, New Cairo, Cairo, Egypt · oneindustry.ai. Every record is described by region and sector. No client is named in this document, and no figure here is attached to a named organisation.

How to read this

The library behind this practice holds 85+ documented deliveries. 35 of them are published, and these are those. The rest is delivered work in sectors with no industrial analogue, or work held for private circulation.

Records are grouped by problem shape because that is what determines whether a system will work. A telecom's accounts-payable problem and a contractor's accounts-payable problem are the same shape: a document arrives, a human reconciles it against two others, a system receives the result. The industry differs and the engineering does not.

SHAPE, IN THE BUYER'S WORDS	DOCUMENTS IT COVERS	RECORDS
A human reads it and types it in again	Invoices, contracts, scanned paper, submittals	**6**
The answer exists and nobody can find it	Procedures, standards, specifications	**13**
Too much arrives to check it all, so you sample	Inspections, calls, feedback, camera hours	**7**
The same judgment, made differently every time	Risk scoring, prioritisation, approvals	**9**

Figures are from the practice's own delivery records for the engagement named, measured against the process that preceded it.

A human reads it and types it in again

Invoices, contracts, scanned paper, submittals

A Gulf public health ministry

Healthcare / Government · Gulf · Document extraction

Teams needed to convert paper and scanned records into structured data. Health records on paper cannot be queried, audited, or linked to anything - the information exists but is unavailable to every downstream system that needs it. The scans are also not uniform: form layouts vary, quality varies, and handwriting turns up exactly where it is least convenient, so a fixed-template extractor fails on precisely the documents that matter most.

What was built. We used vision models to segment pages, extract text, preserve each value's location on the page, itemize fields, and prepare the outputs for downstream processing.

MEASURE	BEFORE	AFTER
Handling time per record	18 minutes	**~2.5 minutes**
Extraction accuracy	no prior measure	**96% clean digital forms / 84% poor scans / 61% handwritten entries**
Fields routed to human review with their source region highlighted	no prior measure	**14%**

What it turned on. Preserving where a value was found, not just the value, is what makes a low-confidence field reviewable - the reviewer is shown the source region instead of having to re-read the document. Relevant to any organisation whose paper archive is the reason a digital system still runs on manual entry.

A Gulf telecom operator's finance function

Telecommunications / Finance · Gulf · Contract intelligence

Billing teams needed to validate invoices and partner revenue shares against complex contracts. The terms that determine a settlement are written in prose, negotiated per partner, and amended over time, so the correct figure is not derivable from the billing system alone - someone has to read the contract and apply it. At partner-portfolio scale that stops being verification and becomes sampling, and revenue leakage lives in the unsampled remainder.

What was built. We built a workflow that interpreted contract terms and applied them to customer billing and partner-settlement records, surfacing the exceptions for human review.

MEASURE	BEFORE	AFTER
Partner settlements verified against contract terms	~15% sampled	**100%**
Analyst time per settlement review	40 minutes	**~6 minutes, on flagged exceptions only**

Discrepancy rate surfaced in the first full-coverage cycle	no prior measure	**2.3x the previously sampled rate**

What it turned on. Surfacing exceptions rather than recomputing settlements kept commercial judgment with the billing team while removing the reading burden that made full coverage impossible in the first place. Relevant to any operator whose revenue assurance is bounded by how many contracts a person can read.

A Gulf telecom operator's finance function

Telecommunications / Finance · Gulf · Multimodal invoice validation

AP teams were manually comparing supplier PDFs, iSupplier entries, and Oracle ERP records. Three-way matching is the kind of work that is too rule-bound to be interesting and too variable to script: the same supplier's invoice arrives in a different layout each quarter, and the discrepancies that matter are small ones buried among formatting differences that do not. Because the check was manual it was also quietly sampled under load - which is exactly when errors are most likely.

What was built. We used a multimodal vision model and a managed model platform to extract invoice fields, reconcile all three sources, flag mismatches, and support correction notifications back to the supplier.

MEASURE	BEFORE	AFTER
Invoices reconciled across supplier PDF, iSupplier and ERP	a sample under load	**100%**
Average validation time per invoice	12 minutes	**45 seconds**
Repeat mismatches from the same supplier after correction notices	its own prior level	**down 67%**

What it turned on. Closing the loop back to the supplier is what stopped the same mismatch recurring every month - detection alone would have relocated the work rather than removed it. Relevant to finance operations whose control is technically three-way matching and practically a sample.

Our own build

Cross-industry / Enterprise · Gulf · Arabic NLP / document intelligence

Built and run by this practice in its own environment.

Organisations processing Arabic documents needed reusable capabilities for entity extraction, summarisation, classification, search, and structured information retrieval. In practice every Arabic document project rebuilt the same foundations - normalisation, OCR handling, label sets, evaluation - before reaching anything specific to the customer. The cost of finding out whether a use case was viable therefore approached the cost of delivering it, which is what kept most of these projects from starting.

What was built. We built an accelerator combining Arabic named-entity recognition, document classification, summarisation, OCR-ready preprocessing, semantic embeddings, and configurable output schemas, with notebook experiments and packaged components for rapid evaluation against customer samples before production integration.

MEASURE	BEFORE	AFTER

Manual document-review time	its own prior level	**down 60%**
Classification consistency	its own prior level	**up 30%**
Structured-field extraction coverage	its own prior level	**up 28%**

What it turned on. Coverage rose because OCR, entity recognition, page zones, and validation rules were combined rather than expecting one model to recover every value - and page-level provenance plus review routing kept higher coverage from meaning that more weak predictions were silently accepted as authoritative business data.

Our own build

Financial Services / Sustainability · Gulf · Document intelligence / ESG analytics

Built and run by this practice in its own environment.

Analysts needed comparable environmental, social, and governance information out of long, inconsistently structured reports and scanned documents. Comparability is the hard part. Every issuer reports on its own template, with metrics in tables that differ in units, scope, and reporting period, and some of it exists only as a scan. Most of an analyst's review is therefore spent establishing what a given number actually refers to, before any comparison is possible.

What was built. We implemented a pipeline that ingested PDFs and images, extracted text and metadata, identified ESG concepts and metrics, linked each piece of evidence to its source page, and presented both document-level and portfolio-level comparisons.

MEASURE	BEFORE	AFTER
ESG report-review time	its own prior level	**down 65%**
Extraction completeness	its own prior level	**up 30%**
Documents reviewed per analyst	its own prior level	**up 45%**

What it turned on. The evidence boundary held throughout - raw values and source pages were never overwritten by normalised outputs, and a missing disclosure was never silently converted into a zero or an inferred commitment. That specific failure is what makes most ESG analytics untrustworthy, and avoiding it is what makes the throughput figure meaningful.

Our own build

Professional Services / HR · Gulf · Recruitment document intelligence

Built and run by this practice in its own environment.

Recruiters needed to compare CVs against job descriptions beyond keyword overlap. Keyword matching rewards the candidates who wrote their CV to mirror the posting rather than the ones who can do the job, and it penalizes anyone who describes the same experience in different words. At volume, the screening pass is also where fatigue does most of its damage, because it is the stage that receives the least time per document and the least scrutiny afterwards.

What was built. We built a workflow that extracted CV content, generated semantic representations, evaluated relevance against the stated criteria, and ranked candidates for recruiter review.

MEASURE	BEFORE	AFTER
Time to a ranked shortlist from a 300-CV posting	~9 hours	**25 minutes**

Qualified candidates surfaced that keyword screening had ranked below the cut	its own prior level	**+23%**
Recruiter agreement with the top-20 ranking	no prior measure	**86%**

What it turned on. Ranking for review rather than filtering is the distinction that matters - the recruiter still sees the pool and the reasoning, so the system reorders attention instead of quietly removing people from consideration. Relevant to any high-volume hiring process whose first pass is currently keyword search.

The answer exists and nobody can find it

Procedures, standards, specifications

A Gulf border-crossing authority

Transportation / Border Operations · Gulf · Conversational AI / digital customer service

Travelers needed immediate answers on crossing times, vehicle-insurance validity, and digital services, while service teams needed a scalable alternative to repetitive contacts. These questions are time-sensitive and usually asked when the traveller is already en route, so an answer that arrives at the end of a contact-center queue has little value. Insurance validity in particular cannot be answered correctly from a static FAQ - it requires a live check against the source system.

What was built. We implemented an Arabic-English virtual assistant across WhatsApp, web, and mobile-ready channels, connected to vehicle-insurance and customer-experience services through secured APIs, with human escalation and administrator access to conversation logs, dashboards, content controls, and service reports.

MEASURE	BEFORE	AFTER
Response time for common traveller questions	its own prior level	**down 55%**
Routine demand handled without entering the contact-center queue	no prior measure	**35%**
Digital self-service completion	its own prior level	**up 28%**

What it turned on. Freshness checks and explicit fallbacks were central to the result - the assistant reduced workload only where the source system could support a dependable answer, rather than containing conversations behind stale or speculative responses. Relevant to any service where the customer is mid-journey and a confidently wrong answer costs more than no answer.

A Gulf judicial body

Government / Judiciary · Gulf · Private document intelligence

Teams needed structured data from scanned material without compromising confidentiality. Judicial material is the case where the usual document-AI answer - send it to a hosted extraction service - is simply unavailable, so the real choice had been between manual processing and no processing at all. That constraint is not a preference to be traded off against accuracy; it determines the architecture before any accuracy question can even be asked.

What was built. We built a private vision workflow combining OCR, layout-aware extraction, and itemisation, with all processing kept inside a controlled environment.

MEASURE	BEFORE	AFTER
Handling time per document	21 minutes	**~3 minutes**
Backlog processable	no prior measure	**all of it, where previously none was - no permissible route existed**

Documents, images or extracted values leaving the controlled environment	no prior measure	**none**

What it turned on. The deployment boundary was the design input rather than a detail settled afterwards, which is what made the capability available to material that could never have left the environment. Relevant to courts, regulators, and anyone whose most valuable documents are the ones they cannot send anywhere.

A Gulf national cybersecurity body

Government / Cybersecurity · Gulf · Website AI assistant

Visitors needed simpler access to the organisation's public information and service journeys. People arrive at a national cybersecurity site with a task, not a topic - report something, request something, check a requirement - while the site is necessarily organised by publication type. And a body of this kind cannot tolerate an assistant that improvises: any answer it gives has to be traceable to approved content, which rules out most of what makes a general assistant feel helpful.

What was built. We delivered a configurable grounded assistant that combined public-content retrieval with flows for scheduling, callbacks, and ticketing, so a question could continue into an action.

MEASURE	BEFORE	AFTER
Time to locate the right public information	4-7 minutes	**~11 seconds**
Visitors completing a service action in the same session as their question	27%	**68%**
Answers traceable to approved published content	no prior measure	**all of them**

What it turned on. Carrying the user from an information request into a service flow is what separated this from a search box - the visitor's task, not the answer, was the unit of completion. Relevant to authorities whose public interactions are split between finding something out and then having to start again to act on it.

A Gulf petrochemicals operator

Energy / Petrochemicals · Gulf · Conversational voice AI

Visitors needed conversational access to company, product, facility, location, and contact information. For an industrial operator this traffic is low-value but not low-volume: contractors, suppliers, applicants, and members of the public asking where to go, who to reach, and what a facility does. Each question is trivial in isolation and none of them are what reception or corporate communications exist to do, yet there was no route to the answers other than a person.

What was built. We reshaped the shared voice platform around a Gulf petrochemicals operator - removing the telecom account logic, adding a corporate knowledge tool, and building location and contact journeys into both phone and browser interactions.

MEASURE	BEFORE	AFTER
Reception and switchboard contacts handled without a person	no prior measure	**64%**
Average time to a location, contact or facility answer	2.9 minutes	**31 seconds**
Time to reshape the existing voice platform around a Gulf petrochemicals operator	no prior measure	**~3 weeks**

What it turned on. Reusing a proven platform meant the effort went into what was actually different - the corporate knowledge source and the location and contact journeys - rather than rebuilding turn management, escalation, and closure behaviour that had already been validated elsewhere.

A Gulf postal operator

Logistics / Postal Services · Gulf · Omnichannel service agent

Customers needed dependable support for shipments, deliveries, and logistics services across channels. Almost all of that demand is one question - where is my item - asked repeatedly by the same customer as the answer changes, and asked most often outside working hours. Volume therefore spikes exactly when staffing cannot, and a channel that cannot answer it pushes people into the queue for an update they could have had directly.

What was built. We delivered a grounded agent that joined service information with conversational guidance around tracking and the common delivery inquiries, across the supported channels.

MEASURE	BEFORE	AFTER
Tracking and delivery enquiries resolved without an agent	no prior measure	**79%**
Resolved contacts occurring outside working hours	no prior measure	**46%**
Average time to a shipment status answer	3.8 minutes	**25 seconds**

What it turned on. Moving the high-frequency questions into a consistent conversational flow gave service teams a more structured set of exceptions to work, so human agents handled judgment, complexity, and recovery rather than repeating status. Relevant to any logistics operator whose contact volume is dominated by a single repeated question.

A Gulf social-insurance authority

Government / Social Insurance · Gulf · Legal semantic search

Users struggled to locate relevant retirement-law provisions through exact terms alone. Retirement law is written in statutory language that nobody uses to describe their own situation - someone asking when they can retire early has to guess the formal vocabulary before keyword search can help them. The cost of landing on the wrong provision is not a poor search experience; it is a member forming a wrong expectation about their own pension.

What was built. We combined multilingual embeddings, cosine similarity, structured legal content, and LLM relevance refinement to handle the complex legal language and the wide variation in how members phrase a query.

MEASURE	BEFORE	AFTER
Correct statutory provision in the top three results	62%	**93%**
Everyday-language queries returning a usable provision	34%	**89%**
Retrieved candidates discarded as semantically close but legally wrong	no prior measure	**~1 in 6**

What it turned on. Refining relevance after retrieval, rather than trusting similarity alone, is what kept a semantically close but legally wrong provision from being presented as the answer. Relevant to any body whose entitlements are defined in statute that its members have to interpret for themselves.

A Gulf tax authority

Government / Tax and Customs · Gulf · Arabic NLP / agent assistance

Employees and service agents needed precise answers from extensive Arabic tax guidance, without slow manual searches and without relying on unsupported general-model knowledge. An agent on a live contact has minutes, and tax guidance is long, written in formal Arabic, and unforgiving of approximation. Both available options were bad: search several guides while the taxpayer waits, or defer the question and generate a follow-up contact that would not have existed otherwise.

What was built. We implemented a retrieval-augmented question-answering pipeline over the Gulf tax authority's documents - extracting and chunking source content, retrieving relevant passages with multilingual embeddings, and prompting an Arabic-capable language model and other Arabic-capable models to produce grounded responses inside an agent-assist interface.

MEASURE	BEFORE	AFTER
Agent search time for tax guidance	its own prior level	**down 65%**
First-contact resolution	its own prior level	**up 30%**
Unsupported or off-context answers	its own prior level	**down 38%**

What it turned on. The service became faster not by trusting the model more, but by reducing the evidence an agent had to inspect and making the boundary between sourced guidance and model generation explicit - retrieval relevance, factual support, Arabic language quality, and output format were evaluated separately, and weak-evidence cases withheld a definitive draft rather than producing one.

A Gulf tax authority

Government / Tax and Customs · Gulf · Enterprise generative AI / analytics platform

a Gulf tax authority needed a governed enterprise capability spanning employee support, document processing, risk analysis, compliance, stakeholder engagement, and operational insight across diverse tax and customs data. Each of those needs was being met, where it was met at all, by a separate manual pipeline with its own extraction, its own access rules, and its own review step. Effort therefore scaled linearly with the number of use cases, and nothing built for one workflow could be trusted or reused in the next.

What was built. We delivered an end-to-end generative AI platform spanning data ingestion, document intelligence, retrieval-augmented generation, conversational applications, model deployment, governance, and monitoring - supporting HR Assist, Agent Assist, Executive Assist, summarisation, translation, classification, anomaly detection, sentiment analysis, and content generation.

MEASURE	BEFORE	AFTER
Document-processing time across supported workflows	its own prior level	**down 55%**
Analyst throughput	its own prior level	**up 35%**
Deployment time for new AI use cases	its own prior level	**down 40%**

What it turned on. The leverage was platform-level: every new use case inherited approved connectors, vector stores, model endpoints, prompt patterns, monitoring, and security controls, while use-case-specific evaluation and access rules stopped that reuse becoming a shortcut around tax, customs, or cybersecurity governance.

A Gulf telecom operator

Telecommunications · Gulf · Conversational voice AI

Customers needed natural voice support across account, billing, usage, package, and sales journeys. Voice is unforgiving here in a way text is not: these journeys cross an authentication boundary, involve numbers that are easily misheard, and are usually attempted while the customer is doing something else. A menu-driven line handles none of that well, and demand scaled directly into agent minutes because every routine balance or package question still reached a person.

What was built. We built the service on a real-time speech model and a telephony provider for phone and browser conversation, with separate pre- and post-authentication states, spoken number confirmation, security questions, and business tools wired into the flow.

MEASURE	BEFORE	AFTER
Average time to complete an authenticated account or billing enquiry	6.2 minutes	**94 seconds**
Conversations completed without an agent	no prior measure	**71%**
Spoken account and phone numbers captured correctly	84% first-pass	**98.6% after confirmation**

What it turned on. A voice agent is judged in the spaces between answers - interruptions, silence, misheard digits, changes of intent - so turn management and graceful recovery were treated as core product behaviour rather than polish. Human agents were left with fewer basic requests and a more structured set of exceptions to review.

A Gulf telecom operator's HR function

Telecommunications / HR · Gulf · Employee self-service

Employees needed immediate answers to recurring HR questions within the mobile experience. The questions arrive constantly and repeat, but employees had to reach them through a channel separate from the app they already had open. When the friction of asking exceeds the value of the answer, people stop asking and start guessing - and the same questions resurface later as tickets, or as errors that cost more to correct than the original answer would have.

What was built. We delivered an assistant that drew on approved HR information through secure integration patterns, with mobile delivery providing context-aware support inside the existing experience.

MEASURE	BEFORE	AFTER
Recurring HR questions answered in-app without a ticket	no prior measure	**74%**
Average time to an HR answer	~1.5 working days	**under 60 seconds**
Tickets raised for already-published information	its own prior level	**down 61%**

What it turned on. The third figure measures a cost that never appeared in an HR report: when asking is harder than guessing, people guess, and the guess returns later as a ticket or an error. Putting the answer inside the app employees already had open removed the friction rather than the demand - relevant wherever self-service exists but sits one channel away from where people actually are.

A Gulf tourism and real-estate developer

Tourism / Real Estate Development · Gulf · Conversational AI / employee self-service

Employees needed a simpler way through onboarding, payroll, benefits, leave, training, policies, documents, and HR requests without generating repetitive work for HR teams. Most of these questions are personal - the correct answer depends on the individual's own record - so no policy page could resolve them and each one became an HR ticket. Onboarding compounded the problem: a new joiner has the most questions and the least idea who to ask.

What was built. We implemented an authenticated employee assistant integrated with SAP HCM, SAP SuccessFactors, Active Directory, and Microsoft Teams, supporting onboarding task lists, personalised payroll and benefits answers, leave and timesheet workflows, document upload, ticket creation, and training information.

MEASURE	BEFORE	AFTER
HR response time	its own prior level	**down 60%**
Administrative effort across the supported workflows	its own prior level	**down 35%**
Employee self-service completion	its own prior level	**up 32%**

What it turned on. Identity revalidation, audit logging, and explicit failure handling underwrote the gain - the reduction in HR workload never depended on holding personal data in conversational memory, or on treating an integration timeout as a successful request. Relevant to organisations whose HR ticket volume is dominated by record-specific questions rather than policy ones.

Our own build

Cross-industry / Enterprise · Gulf · Generative AI / retrieval-augmented generation

Built and run by this practice in its own environment.

Enterprise teams needed a reusable way to ask Arabic and English questions over private documents while keeping source control, predictable answers, and deployment flexibility. Those requirements normally conflict: the fastest route to good answers is a hosted model with a large context window, which is precisely what an organisation with private documents and infrastructure constraints cannot adopt. Each customer therefore rebuilt the same pipeline against its own language, cost, latency, and hosting limits.

What was built. We developed a document question-answering pipeline covering extraction, cleaning, chunking, embedding, vector search, prompt assembly, answer generation, and citations, with the model and retrieval components independently replaceable to suit a given customer's infrastructure.

MEASURE	BEFORE	AFTER
Information-retrieval time	its own prior level	**down 70%**
Grounded-answer accuracy	its own prior level	**up 25%**
Configuration time for a new document collection	its own prior level	**down 40%**

What it turned on. Treating preprocessing, metadata filtering, retrieval evaluation, and bounded prompts as independent controls - rather than leaning on a larger context window - is what allowed a weak component to be swapped for a particular language, document type, latency target, or deployment environment without rebuilding the pipeline around it.

Our own build

Enterprise Technology · Gulf · AI workspace

Built and run by this practice in its own environment.

Employees needed one secure place for company knowledge, AI assistants, and tools. Knowledge sat in separate repositories, each with its own search behaviour, and AI tooling had arrived as more separate destinations rather than fewer. Every additional access point made an answer harder to find rather than easier, and none of them could be governed as a single thing - so access control, retention, and auditability were as fragmented as the content.

What was built. We built a single private workspace combining enterprise data management, enhanced search, LLM capabilities, and access to specialised agents inside one private workspace.

MEASURE	BEFORE	AFTER
Average knowledge-discovery time for indexed enterprise content	~12 minutes	**under 90 seconds**
Common internal questions answered or routed to the correct source without a further repository search	no prior measure	**more than 70%**
Separate access points for knowledge, search, and AI tools replaced by a single workspace	no prior measure	**at least 3**

What it turned on. Consolidation was the mechanism - the gain came from removing destinations rather than adding capability, and a single workspace is also the only unit at which access and governance can realistically be applied. Relevant to organisations where each new AI tool has made the knowledge problem slightly worse.

Too much arrives to check it all, so you sample

Inspections, calls, feedback, camera hours

A beverage bottling operation, three high-speed lines

Manufacturing · · Computer Vision (quality inspection)

End-of-line quality relied on manual spot checks at line speeds above 20,000 units/hour, so human sampling caught only a fraction of defects. Defective pallets occasionally reached distributors, and defect spikes were only discovered hours later through downstream complaints.

What was built. We ran a short defect-data audit before proposing cameras, which showed the recurring costly defects were three specific failure modes - underfill, cocked caps, skewed labels - not the twelve in the QA manual. We deployed cameras on the highest-throughput line first, targeting only those three types, flagging rather than auto-rejecting while operators validated every flag for weeks before enabling automated diversion.

MEASURE	BEFORE	AFTER
Defect escape rate to distribution	its own prior level	**down 84%**
Detection latency	hours	**under 2 minutes, enabling same-shift correction**
Customer returns tied to quality	its own prior level	**down 61% over the following quarter**

> *“Every vendor wanted to sell us a suite catching everything - but three defects were 90% of our real cost, and solving those first is what made the business case obvious.”*

Plant Quality Manager

A Gulf energy operator

Energy / Industrial Security · Gulf · Autonomous robotics / security analytics

the Gulf energy operator's industrial facilities needed continuous security patrols, environmental anomaly detection, and personnel authorisation checks across routes that were repetitive and, in places, higher-risk for the people walking them. The real exposure sat in the gaps between manual rounds: an anomaly that appeared after one inspection was typically discovered at the next, and the observations that were made arrived as narrative reports rather than the time- and location-linked records a control room could act on.

What was built. We delivered an autonomous surveillance robot with self-charging and route-based patrols, combining environmental anomaly recognition, person detection, face identification, and ID-card authorisation with map-based incident reporting and a web interface for remote driving, zoom, and pan.

MEASURE	BEFORE	AFTER
Effective patrol coverage	its own prior level	**up 45%**

Incident-detection time	its own prior level	**down 30%**
Security staff exposure to repetitive routes	its own prior level	**down 25%**

What it turned on. The safety benefit held because autonomy was deliberately conservative - weak face matches required human confirmation, loss of connectivity triggered a local safe behaviour, and blocked or unexpected routes were handled by remote control rather than forcing the robot to continue on its own. Relevant to operators of large industrial or energy sites where patrols are staffed continuously and the coverage gap between rounds is the actual risk.

A Gulf fuel-retail operator

Energy / Retail · Gulf · Sentiment analytics

Product teams needed to understand mobile-store feedback without reading every review manually. App-store reviews arrive continuously and unstructured, mixing one-off complaints with the early signal of a real regression. Read manually, they get read when someone has time - which is rarely the week a release actually broke something, so the feedback loop closed too late to influence the release that caused the problem.

What was built. We built a service that collected App Store and Play Store feedback, classified sentiment, grouped related issues, and presented the trends through a dashboard.

MEASURE	BEFORE	AFTER
Lag from a review appearing to a grouped, scored issue reaching the product team	~3 weeks	**under 24 hours**
Review coverage	a read sample	**100% of App Store and Play Store feedback**
Differently-worded reviews correctly consolidated to one issue	no prior measure	**87%**

What it turned on. Grouping issues rather than scoring sentiment alone is what made the output actionable - a product team needs to know which specific thing twenty people are describing in twenty different ways, not that satisfaction moved. Relevant to any team whose release feedback currently arrives too late to act on.

A Gulf industrial-zone operator

Logistics / Industrial Zones · Gulf · Video analytics

Operations teams needed actionable monitoring from the Al-Wakra Logistics Park video streams. The cameras were already recording, but nothing in the footage was searchable - establishing what had happened meant a person watching it. Events were therefore found after the fact if they were found at all, and the practical value of the whole camera estate was capped by however many review hours the team could spare.

What was built. We delivered a centralised computer-vision service that analysed operational footage, detected the defined events, and fed the results into review and response workflows.

MEASURE	BEFORE	AFTER
Detection accuracy on the defined operational events	no prior measure	**89% under representative conditions**

Time to locate a specific event in the footage	~15 minutes	**under 30 seconds**
Camera hours under continuous analysis	no prior measure	**100%, previously bounded by review capacity**

What it turned on. Separating recognition, OCR, localisation, and contextual enrichment made quality measurable component by component, so the weakest part could be improved without replacing the whole solution, and low-confidence detections were kept for validation rather than counted as confirmed events. Relevant to any operator holding large volumes of camera footage nobody has time to watch.

A Gulf insurer

Insurance · Gulf · Arabic call transcription

Call-center teams needed searchable Arabic transcripts across dialects. Recorded calls that cannot be searched are closer to a liability than an asset - they must be retained and produced on request, but nothing can be learned from them. Arabic sharpens the problem: callers speak in dialect while most systems are built for Modern Standard Arabic, so accuracy varies with who happens to be calling, which is the worst possible property for a compliance record.

What was built. We delivered speech-to-text processing that converted conversations into structured text suitable for search and downstream analysis.

MEASURE	BEFORE	AFTER
Word error rate on Gulf-dialect calls	31%	**12.7%**
Spread between best- and worst-performing dialect groups	19 points	**5**
Call archive searchable	no prior measure	**all of it, previously none**

What it turned on. Handling dialect variation is what turned the archive from stored audio into a queryable record - uneven accuracy across callers would have made any aggregate drawn from it misleading. Relevant to any regulated operation retaining Arabic voice records it currently cannot read.

A Gulf telecom operator

Telecommunications · Gulf · Call intelligence

Supervisors needed scalable summaries, quality checks, compliance monitoring, and coaching insight. Contact-center quality assurance is a sampling exercise - a supervisor listens to a handful of calls per agent per month, which is too few to be fair to the agent and too few to be usable as evidence. Compliance monitoring built on that sample is monitoring in name only, and coaching built on it is largely anecdote.

What was built. We combined transcription-derived summaries, quality assessment, inappropriate-language detection, compliance flags, and training suggestions into a single service over the call estate.

MEASURE	BEFORE	AFTER
Calls receiving a quality review	~2% sampled	**100%**

Supervisor time per reviewed call	11 minutes	**~90 seconds**
Compliance-phrase detection recall on the audited set	no prior measure	**94%**

What it turned on. Moving from a sample to full coverage changes what the output is for: quality scores become defensible to the agent as well as to the auditor, and coaching can point at a pattern rather than at one call someone happened to hear. Relevant to any operation whose quality process is bounded by supervisor listening hours.

Our own build

Cross-industry / Enterprise Operations · Gulf · Process mining / operational analytics

Built and run by this practice in its own environment.

Operations teams needed an evidence-based view of how processes actually moved through their systems - where cases stalled, and which variations drove delay or rework. Every organisation has a documented process, and it is reliably not the one being executed. Improvement effort therefore gets spent optimising the ideal path while the high-volume variant that carries the real cost stays invisible, because nobody has assembled the event data that would show it.

What was built. We built ingestion and preparation for case, activity, and timestamp data, discovered the actual process variants, compared them against the target flows, calculated cycle and waiting times, and highlighted rework and nonconforming paths.

MEASURE	BEFORE	AFTER
Cycle time in the targeted processes	its own prior level	**down 30%**
Bottleneck-resolution speed	its own prior level	**up 20%**
Manual process-analysis effort	its own prior level	**down 65%**

What it turned on. The result depended on validating the log before trusting the map - batch timestamps, broken identifiers, and incomplete cases were corrected or disclosed, so an apparent bottleneck that was only an artifact of how a system records events never attracted investment. Relevant wherever process improvement is being planned against a diagram rather than evidence.

The same judgment, made differently every time

Risk scoring, prioritisation, approvals

A five-site outpatient hospital network

Healthcare · · Core data science / forecasting & decision support

Clinics ran a flat 8% no-show assumption and overbooked uniformly, while real no-show rates swung from 4% to 22% by specialty. The result was patients waiting up to 90 minutes in some clinics while specialists sat idle in others.

What was built. Before building any prediction, we audited the scheduling workflow and found the "no-show" field itself was unreliable - late cancellations, walk-ins, and true no-shows were logged inconsistently across sites. We fixed the data definitions first, then built a per-appointment risk score translated into simple scheduler decision rules, and piloted at one high-variance site for six weeks before expanding.

MEASURE	BEFORE	AFTER
Network no-show rate	14.6%	**9.1%**
Patient wait time in piloted specialties	its own prior level	**down 31%**
Specialist chair utilisation	its own prior level	**+11 points, recovering ~240 slots per month per site**

> *“The win wasn't a fancier algorithm - it was that we stopped treating every specialty the same, and our schedulers trusted the tool because they could see the reasoning.”*

Director of Ambulatory Operations

A mid-market online fashion and homeware retailer, ~14,000 SKUs

Retail / E-Commerce · · Core data science (forecasting + decision optimisation)

Buying and markdown decisions ran on a single blended sales average and merchant intuition, so fast sellers stocked out mid-season while slow movers piled up and cleared at deep, margin-destroying markdowns. Markdowns were set reactively and uniformly, applied late once inventory was already a problem.

What was built. We began with a data and process audit rather than a model, which surfaced that promotions weren't consistently logged and returns weren't netted out - distorting the demand history any forecast would learn from. We cleaned the demand signal first, then built SKU-level forecasts with markdown-timing guidance, prioritising the top-revenue and slowest-moving tiers and piloting on two categories across a full eight-week selling window.

MEASURE	BEFORE	AFTER
Forecast accuracy (WAPE)	~42% error	**19% on piloted categories**

End-of-season clearance inventory	its own prior level	**down 27%**
Gross margin on piloted categories	its own prior level	**+3.8 points**

"Our own sales history was lying to us because of how we logged promotions and returns - fixing that was unglamorous, and it's where most of the accuracy gain came from."

Head of Merchandising

A mid-sized regional property & casualty insurer, ~1.2M active policies

Financial Services / Insurance · · NLP (document processing / classification)

Claims arrived as unstructured attachments - PDFs, phone-photo scans, handwritten forms - into a shared inbox, where 14 clerks read, classified, and keyed each one by hand. First-touch to adjuster assignment averaged 4.3 business days, and roughly 9% of claims were misrouted at least once.

What was built. We began with a two-week audit rather than a model, sampling 1,800 submissions to map where time actually went. Two findings reshaped the scope: 60% of volume came from four claim types, and most misrouting traced to a data-entry step, not a reading problem. We redesigned the routing workflow first, then built NLP extraction against the four high-volume types, running it in shadow mode until accuracy held above 94% on a live holdout.

MEASURE	BEFORE	AFTER
First-touch to adjuster assignment	4.3 days	**7 hours on 72% of volume**
Classification accuracy	no prior measure	**96.8%, versus a 91% human baseline**
Misrouting	9%	**1.4%**

"We expected an AI tool. What we got first was an honest map of why our intake was slow - and fixing that before the model went in is why the numbers held."

Head of Claims Operations

A national economic-development agency running an SME grant program

Government / Public Sector · · NLP (document understanding + intent/eligibility routing)

Case officers read several thousand long-form applications in full to check basic eligibility before any substantive review, even though many failed on straightforward checkable criteria. Decision timelines stretched past three months, much of it spent clearing applications that were ineligible for simple reasons.

What was built. We audited a full prior cycle before designing anything and found eligibility rested on a small set of hard, deterministic criteria buried in unstructured text. We scoped the system to eligibility screening only - substantive scoring stayed entirely human - and ran it in parallel with manual screening for a full cycle to prove it never wrongly rejected an eligible applicant before it influenced any real decision.

MEASURE	BEFORE	AFTER
Officer time spent on eligibility screening	its own prior level	**down 68%**

Time-to-eligibility-decision	3+ months	**under 6 working days**
	no prior measure	**Zero eligible applicants wrongly screened out across the validation cycle**

"The condition was simple - it must never reject someone who actually qualified, and they proved that before we let it touch a live decision."

Director of Programme Delivery

Our own build

Financial Services / Payments · Gulf · Machine learning / anomaly detection

Built and run by this practice in its own environment.

Payment teams needed to identify unusual transactions in highly imbalanced data without overwhelming investigators with false alerts. Imbalance is what makes this deceptive: a model that flags nothing is over 99% accurate, so headline accuracy is not merely uninformative but actively misleading. The binding constraint is also not detection but investigator capacity - an alert nobody has time to open is indistinguishable from an alert that was never raised.

What was built. We built an anomaly-detection service over credit-card transaction data using preprocessing, feature scaling, isolation-based models, evaluation against labeled events, and configurable risk thresholds feeding review queues.

MEASURE	BEFORE	AFTER
High-risk transaction detection	its own prior level	**up 25%**
False-positive alerts	its own prior level	**down 20%**
Initial screening time	its own prior level	**down 55%**

What it turned on. The service kept its own boundary explicit. Feature anonymisation is a privacy control, and it constrains how fully a score can be explained - so the output drives prioritisation and threshold policy, while acting adversely on a customer still requires interpretable customer, merchant, device, and transaction evidence.

Our own build

Retail / Supply Chain · Gulf · Optimisation / inventory analytics

Built and run by this practice in its own environment.

Supply-chain teams needed to place limited inventory across locations while balancing expected demand, service levels, holding cost, and transfer constraints. Those constraints are only solvable together. Correcting one at a time - closing a service gap here, then a transfer cost there - produces a sequence of locally reasonable moves that add up to a worse plan, and planners were doing exactly that across thousands of product-location decisions every cycle.

What was built. We built a pipeline that estimated demand by product and location, calculated allocation priorities, applied stock and logistics constraints, and generated recommended transfers and replenishment quantities alongside scenario dashboards.

MEASURE	BEFORE	AFTER
Inventory-allocation cost	its own prior level	**down 20%**
Product availability	its own prior level	**up 15%**
Manual allocation planning time	its own prior level	**down 55%**

What it turned on. Infeasibility reporting and logged overrides kept it operationally credible - the system stated plainly when the requested service levels could not all be met, and let local knowledge modify a recommendation without concealing the tradeoff being made.

Our own build

Energy / Industrial Operations · Gulf · Predictive analytics / asset maintenance

Built and run by this practice in its own environment.

Energy operators needed to reduce unplanned equipment failures, unnecessary scheduled maintenance, and the cost of making maintenance decisions without timely condition insight. Calendar-based maintenance is wrong in both directions simultaneously - it services healthy equipment and misses deteriorating equipment - but the usual alternative fails on alert fatigue, because startup transients, shutdowns, and faulty instrumentation generate the same warning as a genuine emerging fault.

What was built. We built a pipeline that collected IoT readings and equipment logs, calculated health indicators, detected abnormal patterns, predicted failure risk, and generated early warnings with recommended maintenance actions.

MEASURE	BEFORE	AFTER
Unplanned downtime	its own prior level	**down 30%**
Maintenance expenditure	its own prior level	**down 20%**
Equipment availability	its own prior level	**up 15%**

What it turned on. Separating operating regimes and sensor faults from genuine equipment anomalies is what made the signal actionable - startup, shutdown, and bad instrumentation stopped producing failure warnings, which is the specific reason technicians could trust an alert enough to schedule work against it.

Our own build

Cross-industry / Enterprise · Gulf · Responsible AI / LLM guardrails

Built and run by this practice in its own environment.

Enterprise question-answering systems needed enforceable boundaries - for unsafe topics, unsupported answers, sensitive data, conversation flow, and required response formats. In most deployments those boundaries lived entirely in prompt wording, which meant they could not be audited, could not be tested independently of the model, and quietly degraded whenever the prompt, the model, or the retrieved content changed. Organisations were left unable to say which control had actually stopped a bad answer, or whether one existed at all.

What was built. We built and compared implementations across a model-based safety classifier, a conversational-flow control framework, and an output-validation framework, combining input and output classification, topic controls, predefined dialog flows, structured-output validation, corrective

prompting, and source-grounding rules into a single accelerator.

MEASURE	BEFORE	AFTER
Unsafe or off-policy responses	its own prior level	**down 45%**
Required-format compliance	its own prior level	**up 30%**
Manual review of low-risk conversations	its own prior level	**down 35%**

What it turned on. The operational win was not maximum restriction but risk-tiered control - the full review chain was reserved for sensitive topics and actions, ordinary knowledge questions kept a fast path, and every intervention stayed attributable to a logged policy or validator. Relevant to any regulated organisation that has to explain to an auditor why a generative system refused, escalated, or answered.

Our own build

Retail / Manufacturing · Gulf · Time-series forecasting

Built and run by this practice in its own environment.

Planning teams needed more reliable forecasts across products and locations, to cut stockouts, excess inventory, and spreadsheet-driven planning cycles. A single model across a whole catalogue is wrong in both directions at once - over-smoothing intermittent items while under-reacting on fast movers. The most damaging error is also the least visible: a period with zero sales because stock was unavailable looks identical in the data to genuine zero demand, so the forecast learns to expect the stockout it caused.

What was built. We built a forecasting pipeline over historical demand, pricing, calendar effects, product attributes, and available external drivers, comparing model families per product series, generating confidence ranges, and publishing forecasts into the planning workflow.

MEASURE	BEFORE	AFTER
Forecast error	its own prior level	**down 25%**
Stockout incidence	its own prior level	**down 18%**
Recurring planning effort	its own prior level	**down 50%**

What it turned on. Distinguishing genuine zero demand from unavailable inventory is the specific correction behind the stockout figure. Visible uncertainty and retained planner overrides did the rest - the service improved replenishment decisions without pretending that sparse histories, promotions, and new products can always support a precise point forecast.