EARTHSCAN

WHITEPAPER 19 PAGES / NOVEMBER 2025

# Building Sovereign Subsurface-AI Capability: From an R&D Engagement to In-House Continuous AI

Most operators commission a subsurface-AI programme expecting to receive a model. What durably changes an asset is not the model — it is the capability to keep building models after the vendor leaves. This whitepaper reframes a three-phase, roughly twenty-month formation-evaluation engagement with a mid-sized Middle East carbonate operator we partnered with as a capability story rather than a modelling one. We trace the R&D innovation funnel that produced the algorithms — a NaN remover, unsupervised sinusoid pickers, DBSCAN and hierarchical clustering, a self-supervised segmenter, GAN and KNN imputation, and finally a detection-transformer lineage (DETR through a domain-specialised SUB-DETR) — and show how that funnel was industrialised into a productized Continuous-AI platform (automated fracture, vug, and well-to-well tools running on an on-premise MLOps control plane). Crucially, the programme trained 55 young professionals, 15 of them local nationals from regional universities, so the judgement to run, debug, and extend the platform stayed in-country. We argue that in-house capability is an industrialisation gate with its own acceptance test — the operator climbs the commercial phases themselves — and that a programme which builds the model but not the team has shipped a dependency, not a capability.

**Tarry Singh**
Founder & CEO

**Tannistha Maiti**
Senior AI Researcher

SUBSURFACE-AI / CAPABILITY-BUILDING / CONTINUOUS-WORKFLOW / MLOPS / POC-TO-PRODUCTION / DATA-SOVEREIGNTY

CONTENTS

A subsurface-AI programme has two possible products, and operators routinely buy the wrong one. The first product is a model: a trained network that beats a baseline on a held-out set, wrapped in a demo and presented to a steering committee. The second product is a capability: an in-house team and a platform that can keep producing models — retraining on new wells, extending to new features, and operating the whole apparatus without the people who built it. The first product depreciates the moment the consultants leave. The second compounds. This whitepaper is about how to build the second, drawn from a programme where we did.

## The thesis: capability is an industrialisation gate, not a handoff

Over roughly twenty months and three phases, working with a mid-sized Middle East carbonate operator we partnered with, we built a formation-evaluation system on borehole high-resolution borehole image logs. The phases were not arbitrary; each was a distinct machine-learning posture. Phase 1 was unsupervised — could we find sinusoids at all without labels? Phase 2 was supervised — could a detection transformer learn to pick beddings and fractures with dip and azimuth from a geoscientist's ground truth? Phase 3 was correlation — could the picks from one well be tied to the next across the field? That ladder is real engineering progression, but it is not the part of the programme that determined the operator's eventual return.

The part that determined the return was whether, at the end of the three phases, the operator owned a *capability* or owned a *checkpoint*. Those are different acquisitions. A checkpoint is a file. A capability is a productized platform, a documented pipeline, an operating posture the operator chose deliberately, and — above all — a trained team that can run it. Industrialisation is the gate between the two, and most programmes never define it as a gate at all. They treat the final modelling result as the finish line, hand over a repository, and discover a year later that the asset team is back to buying interpretation from a service company because nobody internal can drive the system.

> **THE TWO PRODUCTS**
>
> A subsurface-AI engagement ships one of two things. A **model** is a depreciating asset: it is as good as it will ever be on the day it is delivered, and it decays from there. A **capability** is an appreciating one: an in-house team plus a Continuous-AI

> platform that produces the next model, and the one after that. The entire argument of this whitepaper is that the second is the one worth commissioning — and that it has to be engineered in from Phase 0, not bolted on at the end.

This whitepaper is written for the CTO, the VP of technology, the digital-transformation lead, and the R&D or capability owner who has to decide what their organisation actually acquires from a multi-year AI engagement. It walks four things in order: the R&D innovation funnel that produced the algorithms; the maturity model that turned that funnel into products; the human capability built alongside it; and the operating-posture decision that hands the dial to the operator.

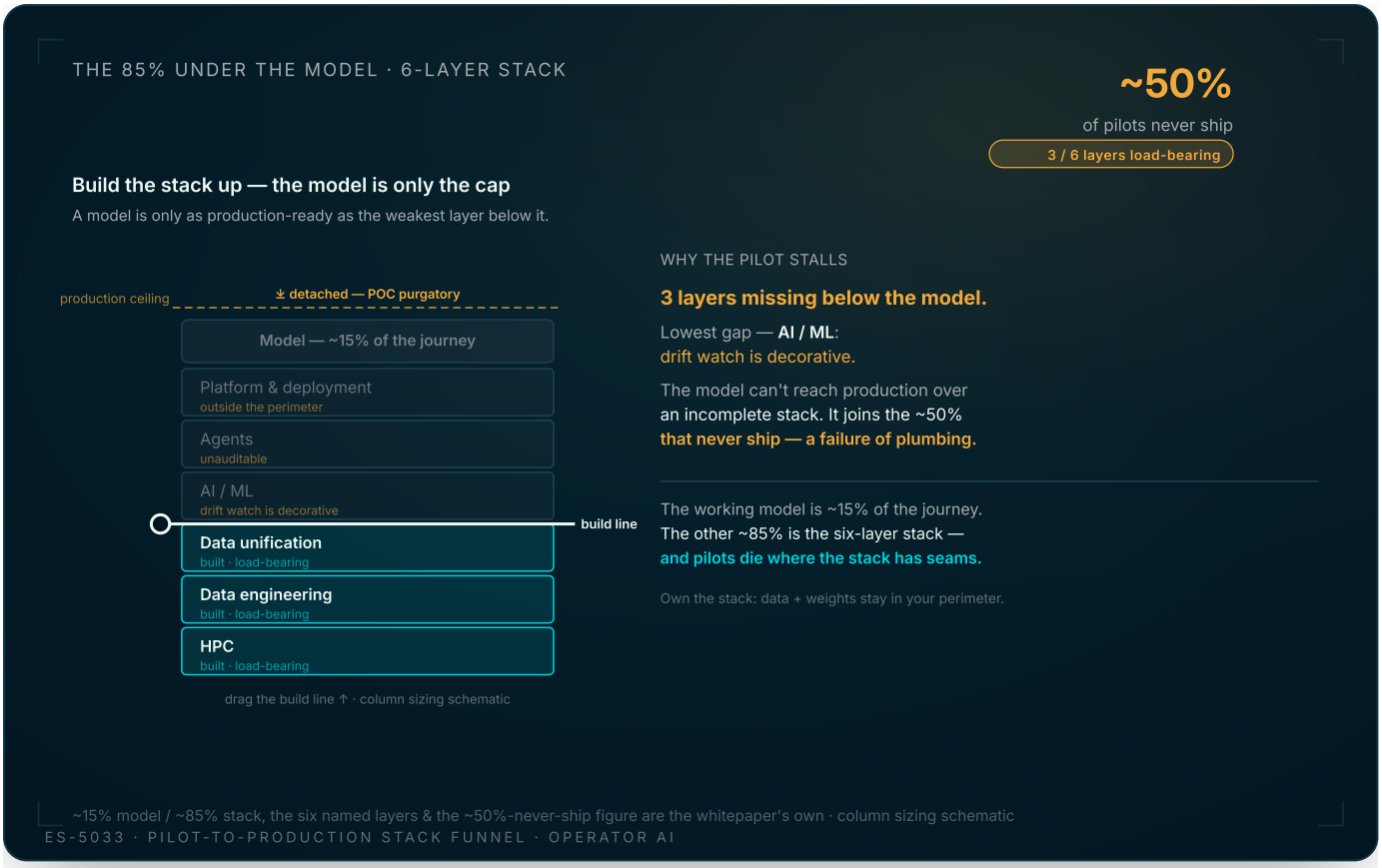
## The innovation funnel: where the algorithms came from

You cannot industrialise a capability you cannot describe. The first discipline is to make the R&D legible — to show, as an explicit funnel, every technique that was tried, why it failed or graduated, and what the surviving lineage is. We ran this engagement against a documented innovation funnel spanning three engineering streams: data management, classical machine learning, and deep learning. The funnel is not a marketing diagram; it is the provenance of the capability, and an operator inheriting the platform needs to be able to read it.

The deep-learning lineage is the spine of the story, and it is worth walking as engineering rather than as a list of acronyms. The earliest work was a **NaN remover** and a normalisation layer — unglamorous data engineering, but the gate everything else passes through, because the binary wireline log files arrive with dead pixels, abnormal static value ranges, and tool-dependent dynamic ranges that defeat naive normalisation. We then built **unsupervised sinusoid pickers**, because the first question was whether the geometry of a bedding plane or fracture — a sinusoid in the unwrapped borehole image — was even recoverable without labels. We tried **DBSCAN** over a three-dimensional point representation of the image and swept more than ten thousand parameter combinations; it was unstable. We tried **hierarchical clustering**; it was computationally ruinous and ran out of memory even on a large machine. We tried a **self-supervised CNN segmenter** (SCOPS-style) at two-, three-, and five-class settings. Each of these was a real attempt with a recorded verdict, and each verdict narrowed the search.

In parallel, the data-imputation stream tested **1-D interpolation**, **KNN imputation**, and **GAN-based imputation** to fill gaps in the image where the tool lost contact with the borehole wall — and produced a counter-intuitive, recorded result that non-imputed data outperformed imputed data for the supervised task, which is exactly the kind of finding that only survives if the funnel is documented. The classical-CV stream built a **path-opening** algorithm with automated threshold tuning for vug detection, a morphological technique that needs no labels and that became one of the shipped products.

The lineage converged on the deep-learning answer: a **detection transformer**. Borehole feature picking is a set-prediction problem — an image patch contains an unknown number of sinusoids, each with a class, a depth, a dip, and an azimuth — and that is precisely the problem DETR was designed for. We took DETR through internal revisions (V1, V1.5) and then into a domain-specialised variant, **SUB-DETR**, tuned for the single-channel, small-data, geometry-regression reality of borehole images: a light ResNet-10 backbone trained from scratch to resist overfitting, a combined focal-and-L1 loss, augmentation applied only to the sinusoid-bearing patches, and well angle incorporated into the geometry. Sitting over all of it was a **data-management dashboard** that itself went through twelve numbered versions — the software-engineering layer that made the funnel operable rather than a folder of notebooks.




*Pilots don't stall because the model is weak. The working model is only ~15% of the journey; the other ~85% is a six-layer engineering stack (HPC → Data engineering → Data unification → AI/ML → Agents → Platform/deployment), and a project ships only when every layer below the model is built to production grade. Drag the build line up the load-bearing column: with all six built the model reaches the production ceiling; with any gap below it the model detaches into POC purgatory — the ~50% that never ship. The ~15%/~85% split, the six layers and the ~50% figure are the whitepaper's own; the equal-sixths column sizing is schematic.*

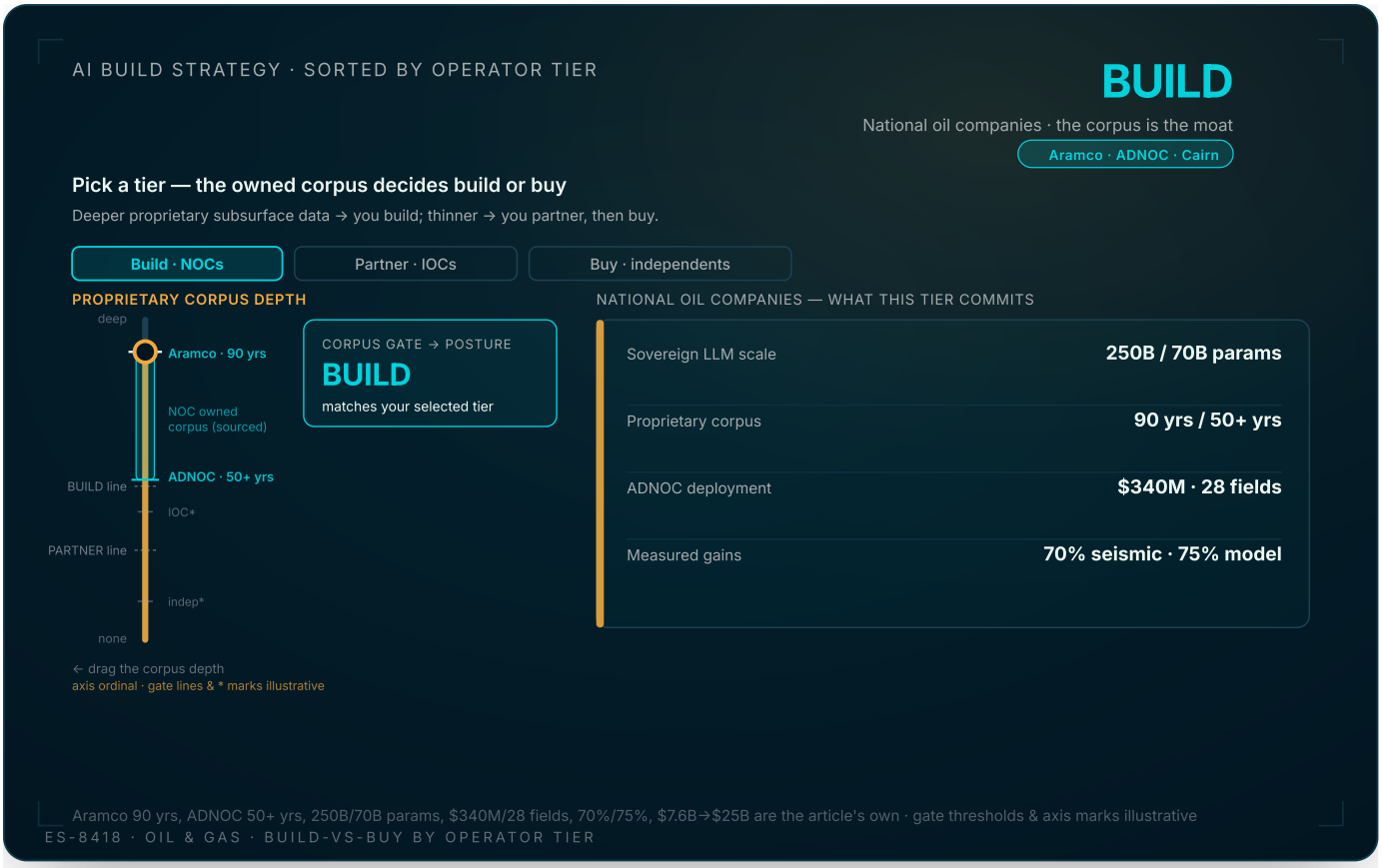
The funnel above is the right mental model for why this matters to a capability owner. The model — the SUB-DETR that picks the fractures — is the visible 15 percent. The other 85 percent is the stack beneath it: the HPC, the data engineering, the data unification, the AI/ML training discipline, the agentic workflow layer, and the platform that serves it inside the operator's perimeter. A programme that hands over only the model hands over the cap of a column whose load-bearing layers were never built on the operator's side. The innovation funnel is how you make those layers inspectable — and an inspectable stack is a transferable one.

## The maturity model: from R&D to productized Continuous AI

A funnel of algorithms is not yet a capability. The second discipline is to industrialise the survivors — to take the techniques that graduated and turn them into products with versions, owners, and service levels. We managed this against an explicit **AI Product Maturity Cycle** that the steering committee tracked phase by phase: the engagement was framed throughout as an R&D project maturing toward commercial phases, with the commercial Phase 4 and Phase 5 deliberately held as the operator's decision to make at the gate rather than the vendor's to assume.

Three algorithms from the funnel matured into named products. **AutoFrac** — automated bedding-and-fracture detection — was the productized SUB-DETR, delivering interpretation roughly five times faster than the manual workflow. **AutoVug** — automated vug detection and quantification — was the productized path-opening lineage, also around five times faster, and it does something no manual workflow does: it produces an unbiased, repeatable vug-dimension curve where the existing practice had no systematic dimensioning method at all. **Well-to-Well (W2W)** correlation tied the per-well picks into a field-scale structural picture using 2-D kriging to extend bedding density beyond the borehole; in the productized form it targeted around 95 percent target-location precision and was associated with a roughly 60 percent productivity gain and a roughly 75 percent improvement in interpretation accuracy. These are not three models. They are three products, each handed over as a complete unit — versioned dataset, frozen model, architecture, output format, and runbooks.

The dataset engineering underneath productization deserves a sentence of its own, because it is where small-data discipline lives. The core sinusoid dataset grew from roughly 900 image-and-ground-truth pairs to over 55,000 — a 65-fold expansion driven by overlapping-patch generation and augmentation — and the well count rose from 8 to 11 over the phases, an increase that alone improved depth, dip, and azimuth error by roughly 0.007 mean absolute error. Both facts are productization facts, not research footnotes: a Continuous-AI platform exists precisely so the operator can keep adding wells and keep the curve moving, and the augmentation pipeline that 65×'d the dataset is part of what was handed over so they could.



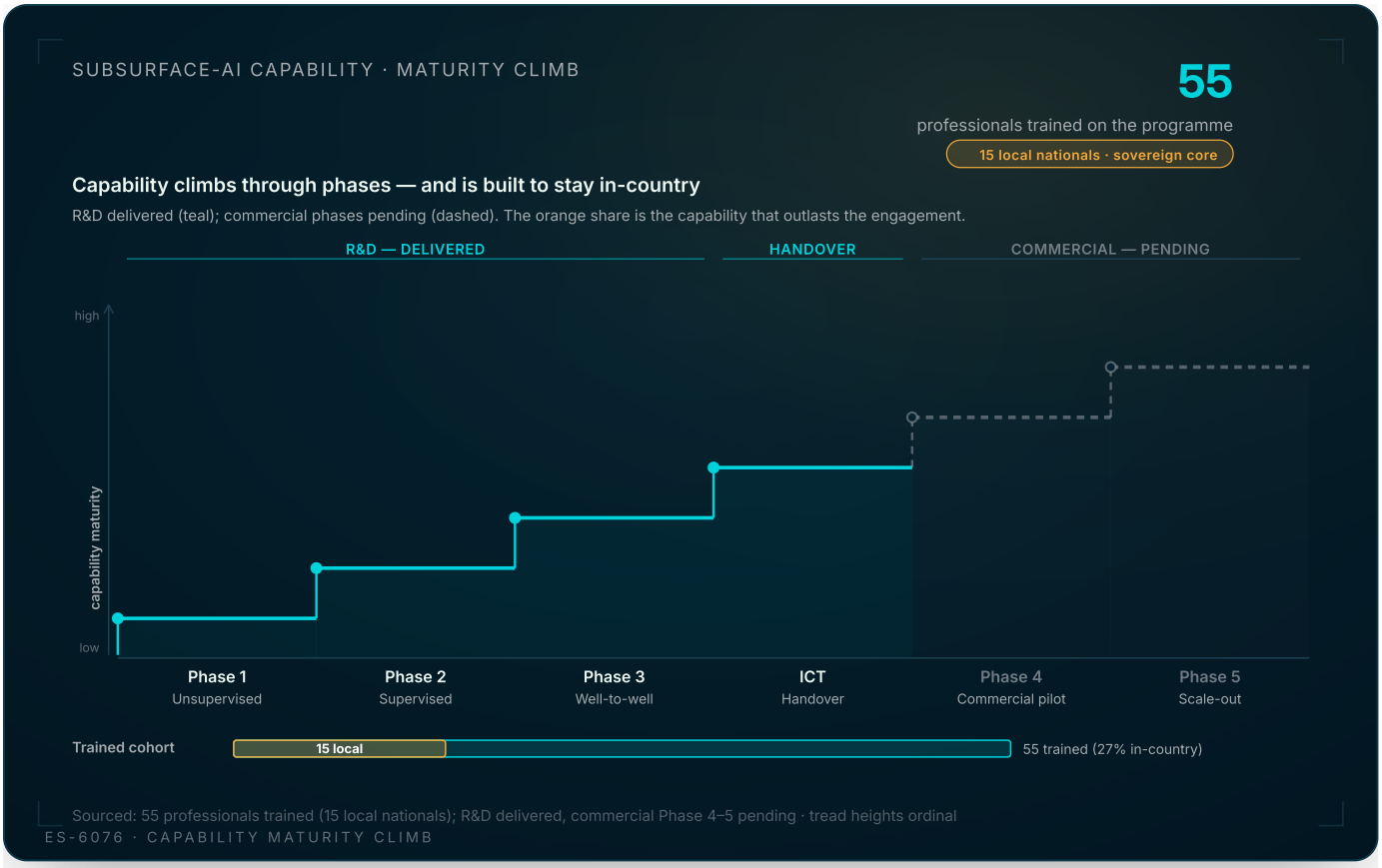

*In 2026 the AI build-vs-buy split in oil & gas is sorted by operator tier, and the deciding variable is the depth of the proprietary subsurface corpus an operator owns. Pick a tier — NOCs (Build), Western IOCs (Partner), mid-tier independents (Buy) — and the panel reconfigures to that tier's posture, named operators and the article's own commitments. The orange ladder is the single argument: the deeper the owned corpus (the sourced NOC band runs from ADNOC's 50+ years to Aramco's 90 years), the further toward BUILD a tier sits. Drag the corpus-depth marker — or step tiers with the chips / arrow keys — and the recommended posture snaps to the band the depth lands in. Named operators, the NOC corpus depths, model sizes ($340M / 28 fields, 250B / 70B params), the 70% / 75% gains and the $7.6B→$25B market are the article's own; the corpus-depth axis, the gate thresholds and the IOC / independent marker positions are illustrative.*

The maturity model also forced an honest infrastructure decision, and the build-versus-buy gate above is the shape of it. The platform ran on an on-premise MLOps control plane — a custom stack (the operator's own ICT estate, with DGX-class compute, an experiment tracker, a versioned data server, and a custom MLOps layer) deliberately kept inside the operator's perimeter rather than rented from a cloud that would also be ingesting confidential subsurface data. For a capability owner, the gate is not "cloud or on-prem" in the abstract; it is whether the data, the weights, and the retraining loop live where the operator's sovereignty and IP protection require them to. In a confidential-subsurface context, that answer pointed on-premise, and the maturity model made that a designed choice rather than a default.

## The capability: 55 trained, 15 in-country

Here is the discipline almost every programme skips, and the one this whitepaper exists to argue for. Software hands over cleanly. The judgement to operate it does not. A versioned dataset, a tracked experiment history, and a reproducible checkpoint are teachable artefacts — but only if someone is taught. The programme deliberately built local capability alongside the platform, training a cohort of **55 young professionals**, **15 of them local nationals** drawn from regional universities, so that the know-how to run, debug, and extend the system lived in the region rather than departing with the delivery team.

This is not a corporate-social-responsibility line item. It is the acceptance test for the whole capability claim. The maturity model puts commercial Phase 4 and Phase 5 in the operator's hands — but a phase the operator cannot staff is a phase that does not happen. The 15 in-country nationals are the sovereign core: the people who can take AutoFrac, point it at next quarter's wells, retrain it, and read the experiment tracker to know whether the retrain helped. The other 40 are the regional bench around them. Train the model and not the team and you have built a dependency on the vendor; train the team and the model and you have built a capability the operator owns outright.




*The programme as a maturity staircase. The teal treads are delivered — the R&D phases (unsupervised → supervised → well-to-well) and the ICT handover that follows; the dashed treads are the commercial phases still pending at engagement close. The argument is in the orange: the climb continues after the vendor leaves because the capability was transferred in-country. 15 of the 55 professionals trained on the programme are local nationals — the sovereign core that owns the commercial steps. The trained-cohort split and the delivered/pending phase boundary are the engagement's own; the tread heights are an ordinal maturity ladder, not a measured score.*

The staircase above is the argument made visual. The delivered treads are the R&D phases and the ICT handover that followed; the dashed treads are the commercial phases pending at engagement close. The point is in the orange marker: the climb continues after the vendor leaves, because the capability was transferred in-country. A handover that moves files but not judgement produces shelfware with a maintenance contract. A handover that moves judgement — embodied in 55 trained people, 15 of them local — produces an organisation that climbs the remaining steps on its own.

There is a second-order return here that capability owners undervalue. The same engineering that makes the platform operable by a trained team is the engineering that makes the operator independent of external service providers for routine interpretation. The unbiased dip-and-azimuth dataset that the supervised pipeline produces in quick

turnaround is the dataset the asset would otherwise purchase from a service company per well, on the service company's schedule. Capability, in other words, is not only resilience — it is the line item that moves a recurring external cost inside the fence.

## The operating-posture decision: handing over the dial

The final discipline is to refuse the false binary of "the operator runs everything" versus "the vendor runs everything," and instead hand over a dial. Capability transfer is not a switch thrown on the last day; it is a posture the operator should be able to set and re-set as their internal bench matures. We costed three honest scenarios with their genuine trade-offs.

The first is **self-operated**: the operator's own team retrains independently on their own on-premise hardware, on a cadence measured in days. This is maximum sovereignty and the destination the 15 in-country nationals are trained toward. The second is the **current managed path**: retraining inside the existing arrangement, where a retrain completes in minutes to hours because the platform and the people running it are already warm. This is the pragmatic middle, and the one a capability typically occupies in its first year. The third is **off-premise managed**: retraining handled externally on a two-to-three-week cadence, the lightest-touch option, appropriate when internal capacity is genuinely constrained. The discipline is laying out all three with real numbers attached, so the operator chooses their position on the dial against their actual bench strength — rather than discovering, post-handover, that the only posture they were equipped for is the one the vendor defaulted them into.

> **THE CAPABILITY ACCEPTANCE TEST**
>
> The honest test of whether a subsurface-AI programme built a capability or just a model: a quarter after the delivery team has gone, can the operator's own people take a new batch of wells, retrain a shipped tool, read the experiment tracker to confirm the retrain improved on the last, and promote the new checkpoint to the asset — without a call to the vendor? If yes, the programme industrialised a capability. If no, it delivered a model and a slow-motion dependency, and the commercial phases on the maturity model will never be climbed.

## What good looks like

For a CTO or capability owner scoping, running, or rescuing a multi-year subsurface-AI engagement, the questions that decide the outcome are not about model architecture. They are about whether the programme was engineered to leave a capability behind:

- Is the R&D legible as a documented innovation funnel — every technique tried, every verdict recorded — so the surviving lineage is inspectable and therefore transferable, not a folder of one person's notebooks?
- Did the graduated algorithms become *products* — versioned, owned, service-levelled units (dataset, model, architecture, output, runbooks) — rather than checkpoints with good metrics?
- Does the maturity model put the commercial phases in the operator's hands as a deliberate gate, with the on-premise-versus-cloud sovereignty decision designed in rather than defaulted?
- Were people trained alongside the platform — a real in-country cohort with the judgement to run, debug, and extend it — so the commercial phases can actually be staffed?
- Was the operating posture handed over as a dial with three costed positions, so the operator sets capability transfer against their own bench strength rather than inheriting the vendor's default?

If the answers are yes, the operator owns a capability that compounds. If they are no, the operator owns a model that depreciates — and a countdown to the next service-company invoice.

This argument generalises across the operators we have worked with in the Middle East and the United States: the funnel, the maturity gate, the people, and the dial are not specific to one field. What is specific to a single engagement are the numbers. Every quantified result in this whitepaper — the 65-fold dataset growth, the 8-to-11-well expansion, the roughly five-times interpretation speed-up, and the cohort of 55 trained with 15 in-country — traces to the Middle East carbonate programme described here, and to that programme alone. The capability pattern is the transferable asset; no single number inside it is.

### WHAT THIS WHITEPAPER ARGUES

1. A subsurface-AI engagement ships one of two products: a model (depreciating) or a capability (compounding). Commission the second — and engineer it in from Phase 0, not at the end.
2. Make the R&D legible as a documented innovation funnel — NaN remover, unsupervised pickers, DBSCAN/clustering, self-supervised segmenter, GAN/KNN imputation, DETR through a domain-specialised SUB-DETR — so the surviving lineage is inspectable and transferable.
3. Industrialise the survivors into versioned products (AutoFrac, AutoVug, Well-to-Well ~5x faster) on an on-premise Continuous-AI / MLOps platform inside the operator's perimeter, with the sovereignty decision designed in.
4. Train people, not just models: a cohort of 55 — 15 of them local nationals — was the sovereign core that lets the operator staff the commercial phases the vendor deliberately left in their hands.
5. Hand over a dial, not a switch: three costed operating postures (self-operated days, current managed minutes-to-hours, off-prem managed 2-3 weeks) let the operator set capability transfer against their own bench strength.

## References


International Energy Agency, 2025 International Energy Agency. Energy and AI Special Report (2025). Identifies missing internal expertise — not model quality — as the dominant barrier to AI adoption across the energy sector. https://www.iea.org/reports/energy-and-ai

McKinsey & Company, 2025 McKinsey & Company. The State of AI (2025). Workflow redesign and in-house capability identified as the strongest correlates of AI value capture, over model performance alone. https://www.mckinsey.com/

Carion et al., 2020 N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, S. Zagoruyko. End-to-End Object Detection with Transformers (DETR). ECCV 2020. The set-prediction architecture at the head of the detection-transformer lineage industrialised in this programme. https://arxiv.org/abs/2005.12872

Sculley et al., 2015 D. Sculley et al. Hidden Technical Debt in Machine Learning Systems. NeurIPS 2015. The canonical argument that the trained model is a small fraction of a production ML system — the basis for the 15%/85% stack framing. https://papers.nips.cc/paper/2015/hash/86df7dcfd896fcaf2674f757a2463eba-Abstract.html

EARTHSCAN

## Building Sovereign Subsurface-AI Capability: From an R&D Engagement to In-House Continuous AI

Published November 2025. Generated 2026-07-22.
https://www.earthscan.io/whitepapers/building-in-house-subsurface-ai-capability-continuous-ai