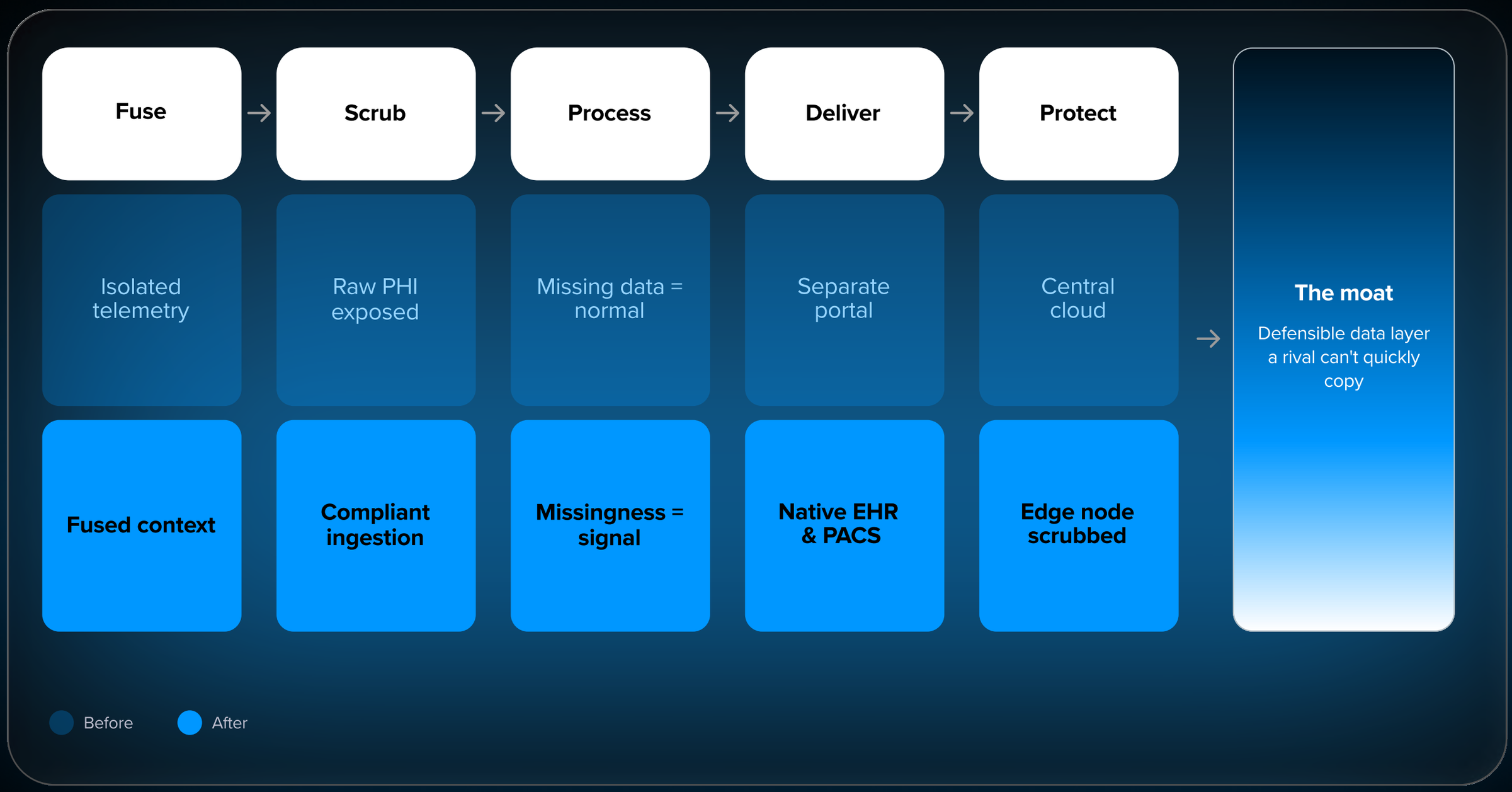
Star



From Hardware Vendor to Indispensable Clinical Partner.

How a multimodal data pipeline transforms device telemetry into a clinical AI asset no competitor can quickly copy.

Where does the data layer come from?

The workflow analysis identified where care breaks down and what the digital component should do. The harder problem starts now: sourcing the data that makes that component work.

Medical device OEMs default to device telemetry because they own it. But telemetry shows what the hardware did, not what happened to the patient — and it can't support the high-value roles the analysis pointed to: monitoring, decision support, coordination. Clinicians never decide from a single stream.

They read device signal against imaging, lab trends, and clinical notes, together. The pipeline has to do the same. Five technical problems sit between a well-defined digital component and a working one.

Most MedTech OEMs are sitting on valuable device data and don't realise it's only half the picture. The moment you fuse that telemetry with what's happening clinically — imaging, labs, care notes — you stop competing on hardware price and start building something a competitor genuinely can't replicate overnight.

Manuel Vera
Product Manager, Star

The Pipeline

Five stages. One defensible data layer.

HOSPITAL PREMISES (EDGE NODE / LOCAL FIREWALL)

- **Continuous Telemetry** (MQTT / Logs)
- **PACS Imaging** (DICOM /DICOMweb)
- **EHR System** (HL7 FHIR APIs)
- **Clinical Narratives** (Unstructured notes)

01 | INGEST & FUSE — Multimodal Alignment

- **Time-Keyed Indexer** (Temporal sync)
- **Latent-Space Projector** (Joint representation)

02 | SCRUB & SECURE — Boundary De-identification

- **OCR Pixel-Scanner** (Redacts burned-in PHI)
- **BioBERT NER Model** (Redacts text quasi-IDs)

05 | LEARN & PROTECT [Edge] — Edge

- **Edge Utility Audit** (Resolution & biomarkers)
- **Deterministic Temporal Shifter** (Relative delta conversion)

HOSPITAL FIREWALL (ZERO-TRUST BOUNDARY)

EDGE-SIDE PROCESSING

All four edge stages run on hospital premises, inside the local firewall.

Only de-identified, temporally shifted, anonymized payloads are permitted to cross into the secure cloud pipeline.

This satisfies hospital IT, legal, and compliance review without requiring raw PHI to ever leave the network.

Anonymized Payloads Only

SECURE OEM CLOUD PIPELINE

03 | PROCESS & IMPUTE

Missingness Classifier (Data gaps parser)

- Technical dropouts → **MCAR / MAR Path** — Routine Imputation - Muted
  - **Model-based imputation** (Interpolation / Mean / Median)
  - Background contribution
- Clinically informative → **MNAR Path** — Active Clinical Moat - High Contrast
  - **Clinical Signal Extractor** (The gap is the feature)
  - Actionable alert → **Targeted Patient Outreach** (Engagement trigger)
  - Primary, weighted input

Resilient Feature Vector (Built on the clinical signal)

Deployed model update

04 | TRANSLATE & DELIVER — Invisible Assistant

- **Interoperabllitv encoder** (HL7 FHIR/DICOM SR Mappings) →
- **Native Clinical Workspace** (EHR pre-pop/PACS worksheets)

05 | LEARN & PROTECT [Cloud] — Centralized Model Retraining

Proprietary Anonymized Lake → (Scheduled batch) **Centralized Retraining Job** (Deterministic, scheduled) → (Versioned artifact) **Model Registry** (Versioning & rollback)

STEP 1

Ingest & Fuse: Build multimodal clinical context

Device telemetry tells you what the hardware did — not what happened to the patient. To escape commodity status, OEMs must programmatically fuse device signals with imaging, lab trends, and clinical notes, mirroring how clinicians actually reason.

How it works

Ingest continuous hardware telemetry (sensor readings, operational logs, usage metrics) as the foundational data stream.

Connect to hospital PACS to pull imaging studies (CT, MRI, X-ray, wound photography, OCT) aligned to the same patient and time window.

Extract structured lab results and unstructured clinical progress notes from the EHR via HL7 FHIR APIs.

Run a joint representational model that learns cross-modal interactions — e.g., a temperature spike in sensor data paired with a tissue-color change in imaging and a nurse's note on exudate.

Output a fused, context-aware risk score or therapeutic recommendation that no single data stream could generate alone.

Before

- Device outputs raw telemetry in isolation — basic usage logs, physical measurements, battery status.
- Competitive OEMs can replicate the same data stream; the device remains a commodity.
- Clinical teams must manually correlate device readings with their own imaging and notes.
- Algorithm alerts lack clinical context, driving high false-positive rates and alert fatigue.

After

- Device telemetry is automatically fused with imaging, lab values, and clinical notes into a single context model.
- The fused multimodal dataset is proprietary and structurally difficult for competitors to replicate.
- The AI surfaces synthesized, cross-modal risk signals directly — as a clinician would reason.
- Predictions are contextually grounded, reducing noise and increasing clinical confidence.

STEP 2

Scrub & Secure: Automated compliance at the ingestion boundary

Hospital IT, legal, and compliance teams will reject any integration that exposes raw patient data to a third-party cloud. The ingestion boundary must act as an automated, zero-trust compliance shield — scrubbing Protected Health Information (PHI) from both text and images before a single byte leaves the hospital.

How it works

Route all DICOM binary image files through an Optical Character Recognition (OCR) pixel-scanning layer that detects and blacks out 'burned-in' patient identifiers embedded directly in the image pixel matrix — not just the metadata headers.

Simultaneously route all unstructured clinical text (EHR notes, radiology dictations, nursing entries) through a domain-specific Named Entity Recognition (NER) model (e.g., BioBERT) to detect and redact direct and quasi-identifiers.

Preserve critical clinical, diagnostic, and anatomical context — the NER model redacts identity, not medical signal.

Output a Redacted Text JSON payload and a Redacted Pixel Array that are safe for secure cloud ingestion.

Log all scrubbing events with a verifiable audit trail for HIPAA, GDPR, and PDPA compliance documentation.

Before

- De-identification relies on standard DICOM header stripping — patient name and DOB tags removed.
- Unstructured clinical text is passed through unchanged or manually reviewed, creating bottlenecks.
- Hospital IT identifies the gap during security review and blocks the integration.

After

- OCR-driven pixel masking detects and redacts burned-in text within the image frame itself, beyond the metadata layer.
- BioBERT NER redacts indirect quasi-identifiers (rare occupations, specific locations, familial details) from clinical notes.
- Automated audit logs satisfy compliance officers and accelerate hospital IT security approvals.

STEP 3

Process & Impute: Handle incomplete real-world data and informative missingness

Real-world clinical data is never pristine. Patients forget to sync devices, batteries die, labs are ordered irregularly, and EHR charting is incomplete. A resilient multimodal pipeline must be engineered with advanced imputation strategies that dynamically handle these gaps — but more importantly, it must distinguish between random missing data and Informative Missingness (Missing Not at Random — MNAR), where the absence of a data point is clinical context in itself.

How it works

Recognise that not all missing data carries the same meaning.
The critical category is MNAR: data that is absent for a clinical reason, not a logistical one.

Flag MNAR events as active clinical signals rather than gaps to be filled. A missing serum lactate means a clinician chose not to order it. A missing device sync may mean a patient stopped using their inhaler entirely.

Treat the absence as the input. Route MNAR signals directly into the risk model as weighted features — the gap itself informs the output.

Trigger targeted human outreach for MNAR patterns rather than automated imputation or silent suppression.

Continuously validate signal logic against ground-truth outcomes to prevent systematic risk underestimation in the patients who need attention most.

Before

- Missing data is handled uniformly using MCAR or MAR assumptions — gaps are filled with historical averages or flagged as data quality errors.

After

- MNAR is identified as a distinct clinical signal category — the absence is treated as a meaningful input, not a data gap to correct.

STEP 4

Translate & Deliver: Seamless workflow integration into EHR and PACS

The most technically accurate multimodal SaMD is commercially worthless if it requires clinicians to leave their primary system. Physicians, radiologists, and nurses have zero tolerance for portal fatigue. The pipeline must deliver its outputs as an invisible assistant — surfacing insights inside the tools clinicians already use, in formats they already understand.

How it works

- Convert model outputs into standardized healthcare interoperability schemas: HL7 FHIR DiagnosticReport and Observation resources for EHR delivery; DICOM Structured Reporting (SR) files for PACS delivery.
- Write structured findings and draft narratives directly back into the clinician's native workspace.
- Design outputs as editable suggestions, not blocking alerts.
  The clinician reviews, modifies, and signs off with a single action.
- Eliminate separate login portals, manual data transcription, and copy-paste workflows entirely.
- Track acceptance, modification, and override rates per output type
  to continuously tune alert thresholds and reduce non-actionable interruptions.

Before

- SaMD outputs are delivered via a separate web portal requiring unique credentials and a manual login.
- Clinicians must transcribe or copy-paste findings into the EHR or PACS dictation template by hand.
- Between 90% and 96% of clinical decision support alerts are overridden or ignored outright.

After

- AI findings are written directly back to the EHR as FHIR resources and to PACS as DICOM SR files — no separate portal.
- The clinician opens their normal dictation template and finds the AI's draft findings pre-populated and ready to review.
- Clinician adoption increases because the tool fits around their existing process rather than disrupting it.

STEP 5

Learn & Protect: Hybrid governance architecture for continuous model improvement

Fusing device telemetry with hospital clinical data creates an immediate data governance deadlock. OEMs guard their device IP; health systems and regulators guard patient records under GDPR, HIPAA, and PDPA. Centralising raw patient data in a vendor cloud is rejected by enterprise health systems. Full decentralisation is operationally and financially unworkable. The resolution is a Hybrid Governance Architecture.

How it works

Deploy a lightweight, automated De-identification Edge Node behind each hospital's firewall. This node performs all PHI scrubbing locally — EHR text, clinical notes, and DICOM image pixel arrays — before any data leaves the hospital network.

Apply deterministic temporal shifting at the edge: absolute timestamps are converted to relative delta offsets (e.g., 'Day 0, Hour 4') anchored to a randomized, patient-specific date generated inside the firewall. This breaks linkage to external event registries while preserving cross-modal temporal dependencies for AI training.

Stream only fully anonymized, structured data payloads to the centralized OEM cloud pipeline for high-compute model fusion, training, and validation.

Run an automated edge utility audit before each payload export: verify that image resolution, signal frequencies, and key biomarker values remain intact and uncorrupted. If utility falls below threshold, trigger a local alert rather than polluting the central data lake.

Maintain a single, centralized cloud data lake for cost-effective, deterministic model training at scale — without any raw PHI ever crossing the hospital boundary.

Before

- Raw patient records are sent to a central OEM cloud repository — immediately rejected by enterprise hospital legal and IT teams.
- A fully decentralised architecture is attempted across dozens of isolated hospital environments, creating operational and financial unsustainability.
- Legacy DICOM header-clearing scripts miss burned-in pixel PHI — raw patient data escapes the hospital firewall.
- Over-aggressive scrubbing strips temporal offsets and biomarker signals, leaving the cloud data lake clean but analytically useless.

After

- A lightweight edge node performs all PHI scrubbing locally behind the hospital firewall — only anonymized payloads leave the network.
- The centralized OEM cloud pipeline receives clean, structured, clinically rich data from all sites simultaneously — scalable and cost-effective.
- OCR pixel masking at the edge node catches burned-in identifiers; deterministic temporal shifting prevents re-identification via timestamp linkage.
- Automated utility auditing ensures that each export retains the clinical features the AI pipeline requires — governing quality, not just privacy.

Pitfalls:

- The DICOM Leakage Trap: Relying strictly on header-clearing scripts while legacy modalities burn patient names and MRNs directly into the visual pixel matrix, leading to immediate PHI exposure and regulatory shutdown.

- The Re-Identification Trap: Failing to protect continuous telemetry signal timestamps from linkage attacks, allowing bad actors to cross-reference millisecond-exact device logs with external registries to re-identify individuals.

- The Utility Degradation Failure: Implementing over-aggressive edge scrubbing scripts that blindly destroy relative time offsets or key unstructured biomarker signals, leaving the centralized data lake clean but mathematically useless.

How multimodal data pipeline turns your digital component into the defensible moat

Hardware telemetry alone is a commodity. Any competitor with a comparable device can replicate the data stream — and purchasing departments will substitute on price. The defensible moat is not the device and not the signal it generates. It is the proprietary multimodal data layer built underneath it: a fused dataset that binds device telemetry to clinical imaging, lab trends, and care narratives in a way that took years of hospital relationships, compliance infrastructure, and clinical validation to construct. That is what a competitor cannot quickly copy — and what drives clinician preference and procurement retention back toward your hardware.

What makes this genuinely difficult and therefore valuable is that the moat sits across four disciplines that rarely live in one team: clinical data modelling, FHIR and DICOM interoperability, regulatory and data governance, and clinician workflow design.
The OEMs pulling ahead are not winning on hardware specifications. They are winning because they have brought those four capabilities together and embedded context-aware intelligence directly and invisibly into the clinician's daily workflow. That is the asset. That is what compounds.

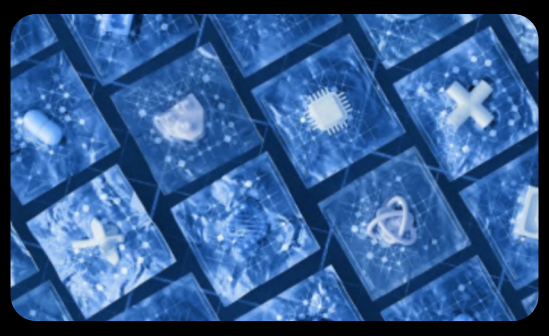
About Star

Star is a global consultancy delivering better health outcomes with compliant MedTech solutions.

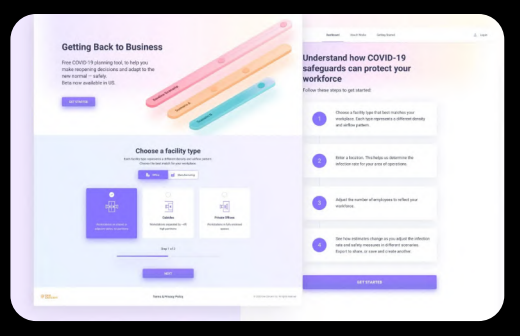
Innovating with purpose, building with expertise

End-to-end healthcare solutions development

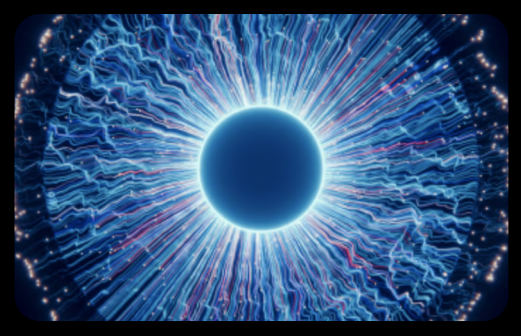
Digital Therapeutics and Companion Apps



Interoperability and Patient-Centered Intelligence



Scalable HealthTech Platforms



AI-powered Regulatory and AI compliance

Our clients

Phillips Medisize a molex company · xerox · Magic Leap · ZEISS · ondosis · actegy health · stryker

Medicom · Allianz · gmendel · Clarify · Constant Therapy · skin analytics · Roche

one concern · Abbott · DESIGN INstitute for HEALTH · BRAINLAB · Closed Loop Medicine · EGYM

Capital Rx · vocera · novo nordisk · CONDUENT · presidiohealth · JOHNSON

star.global —— healthtech@star.global

Delivering impact with cross-functional expertise

Workflow Design

UX & UI

System integrations and Interoperability

Machine Learning and AI

Patient Experience and Care Journey Strategy

Regulatory & Compliance

Compliant software development

Data Engineering, Management and Analytics

Deployment, Operations and Maintenance

Let’s build the future of healthcare together.

Connect with us to explore how we can turn your vision into impact.

Contact us

Mahesh Naphade
Global Head of Healthcare and Life Sciences

Manuel Vera
Product Manager

Pavel Kyrylchenko
Senior Product Manager

Antonina Burlachenko
Head of Quality and Regulatory Consulting

Yanick Gaudet
Interoperability Solution Architect

star.global —— healthtech@star.global