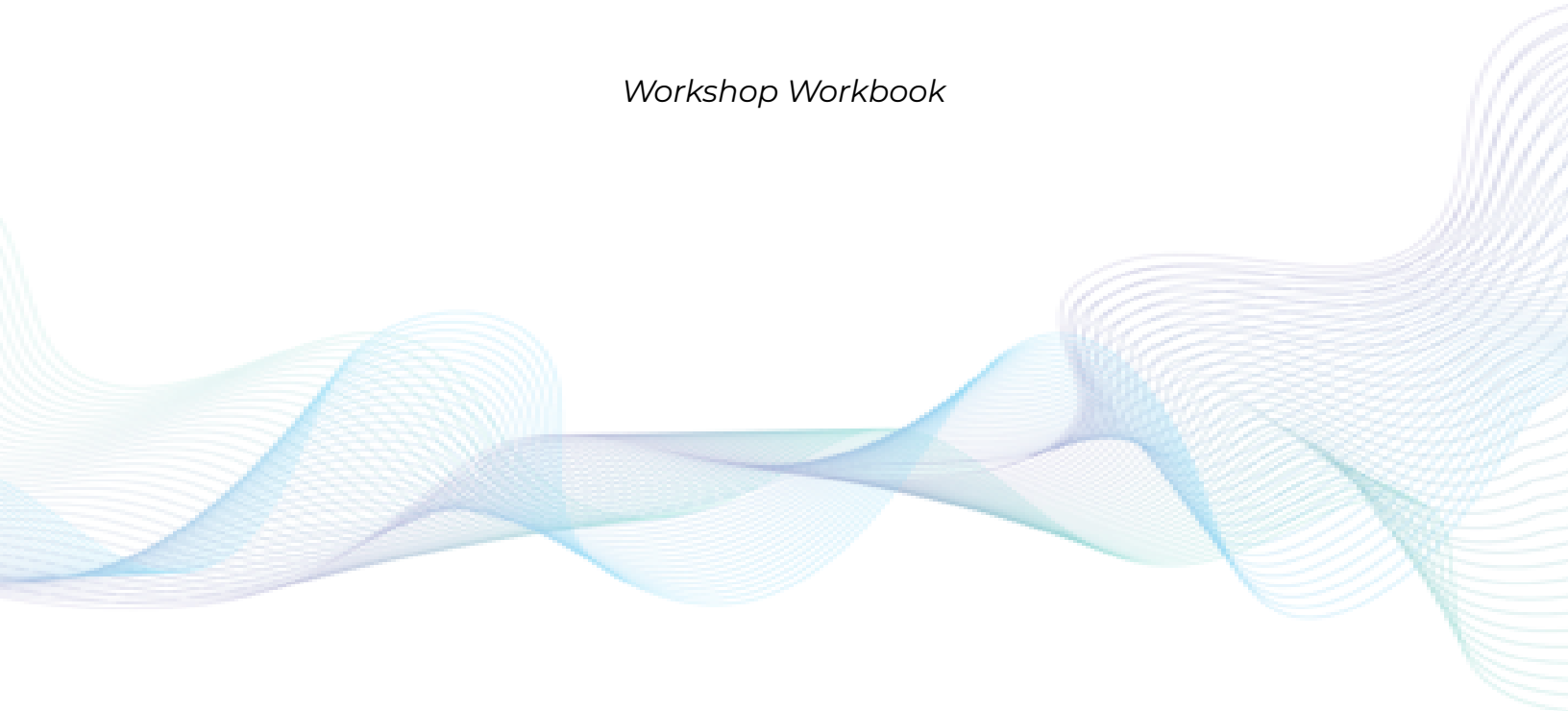




# Building a Modern AI Data Practice

*Workshop Workbook*



## **Contents:**

---

Exercise 1 – AI Initiatives Retrospective & Data Failure Patterns.....	2
Exercise 2 – Data Quality Evaluation: Clean, Consented, Actionable.....	3
Exercise 3 – Data Orchestration Swimlanes Flow Map.....	4
Exercise 4 – AI Data Metric Inventory & Quality Analysis.....	5
Exercise 5 – Modern AI Data Practice Outline.....	7
Exercise 6 – Customer Data Orchestration Strategist Overview.....	8
Exercise 7 – Bonus Round: Building a Tealium-Powered AI Data Orchestration Map.....	10

# Exercise 1 – Data Initiatives Retrospective & Failure Patterns

**Goal:** Make AI failures concrete and trace them back to data and orchestration.

List 2–3 past or current AI / advanced analytics initiatives.

#	Initiative Name	Business KPI / goal	Actual Outcome (hit/miss/stalled/cancelled)
1			
2			
3			

For **one initiative**, what happened; Initiative selected: \_\_\_\_\_

**Impacts observed:**

- Timelines, scope, or expectations slipped
- Manual workarounds were needed
- Governance / privacy / legal challenges impacted the project
- Executive confidence / ownership or sponsorship issues

Primary contributors, check all that apply and add examples.

- Data quality & coverage** (missing, late, inconsistent, no clean source)
  - Example: \_\_\_\_\_
- Identity & joining** (fragmented IDs, duplicates, hard joins)
  - Example: \_\_\_\_\_
- Consent & governance** (unclear if data could be used, late “no”)
  - Example: \_\_\_\_\_
- Activation & integration** (scores never used, latency too high, no connectors)
  - Example: \_\_\_\_\_
- Ownership & operating model** (no one owned process end-to-end)
  - Example: \_\_\_\_\_

**What patterns do you see across initiatives?**

- Pattern 1: \_\_\_\_\_
- Pattern 2: \_\_\_\_\_
- Pattern 3: \_\_\_\_\_

**If we had a Data for AI CoE, it would have helped by...**

- Standardized: \_\_\_\_\_
- Made visible earlier: \_\_\_\_\_
- Owned decisions about: \_\_\_\_\_
- Other: \_\_\_\_\_
- Other: \_\_\_\_\_

## Exercise 2 – Data Quality Evaluation: Clean, Consented, Actionable

**Goal:** See where your core sources are strong or weak on **Clean, Consented, Actionable**.

Step 1: List your main data feeds that matter for AI.

#	Data source / feed (e.g., web, mobile, CRM, POS)	System(s)	Primary owner / team
1			
2			
3			
4			
5			

Step 2: Rate the quality of those items on the scale: **1 = very weak, 5 = strong** (or use R/A/G).

Data source / feed	Main AI use cases	Clean (1–5)	Consented (1–5)	Actionable (1–5)	Notes / examples

- **Clean:** completeness, accuracy, consistent schema, clear IDs.
- **Consented:** purpose and lawful basis for AI uses clearly recorded and enforced.
- **Actionable:** reliably reaches AI and activation systems with acceptable latency.

Step 3: Circle or mark **2–3 high-value, low-score sources**.

For each, capture:

Data source / feed	Biggest AI risk if left as-is	90-day improvement ideas (data quality / consent / pipelines)

### Exercise 3 – Data Orchestration Swimlanes Flow Map

**Goal:** Map a real AI use case across **Connect** → **Govern** → **Transform** → **Activate**, including teams, tools, and gaps.

Step 1: Select One Priority AI Use Case

- Use case name: \_\_\_\_\_
- Target KPI / outcome: \_\_\_\_\_

Step 2: Use the space below (or a separate sheet/whiteboard) to draw four horizontal lanes and note systems, teams, and flows.

Use Case Name	Connect	Govern	Transform	Activate
	<ul style="list-style-type: none"> <li>• Sources (web, app, APIs, offline)</li> <li>• IDs captured (visitor, account, device, etc.)</li> </ul>	<ul style="list-style-type: none"> <li>• Where consent is collected &amp; enforced</li> <li>• Policies/checkpoints</li> </ul>	<ul style="list-style-type: none"> <li>• Key features / attributes needed for this use case</li> <li>• Where they're computed (Tealium, CDW, AI platform, etc.)</li> <li>• Quality checks in place</li> </ul>	<ul style="list-style-type: none"> <li>• Channels using AI outcomes (onsite, email, ads, support, etc.)</li> <li>• How predictions are consumed (segments, scores, flags)</li> <li>• Feedback signals (conversions, churn, complaints, etc.):</li> </ul>
Systems Used				
Types of Data Involved				
Teams / Owner				
KPI's				
Orchestration Vulnerabilities				
Improvement Ideas				

# Exercise 4 – AI Data Metric Inventory & Quality Analysis

**Goal:** Identify the metrics that show whether your data is truly **AI-ready** and where maturity gaps are.

**Step 1:** Pick one AI use case (same or different from earlier):

- AI use case: \_\_\_\_\_

**Step 2:** Define metrics and quality by layer

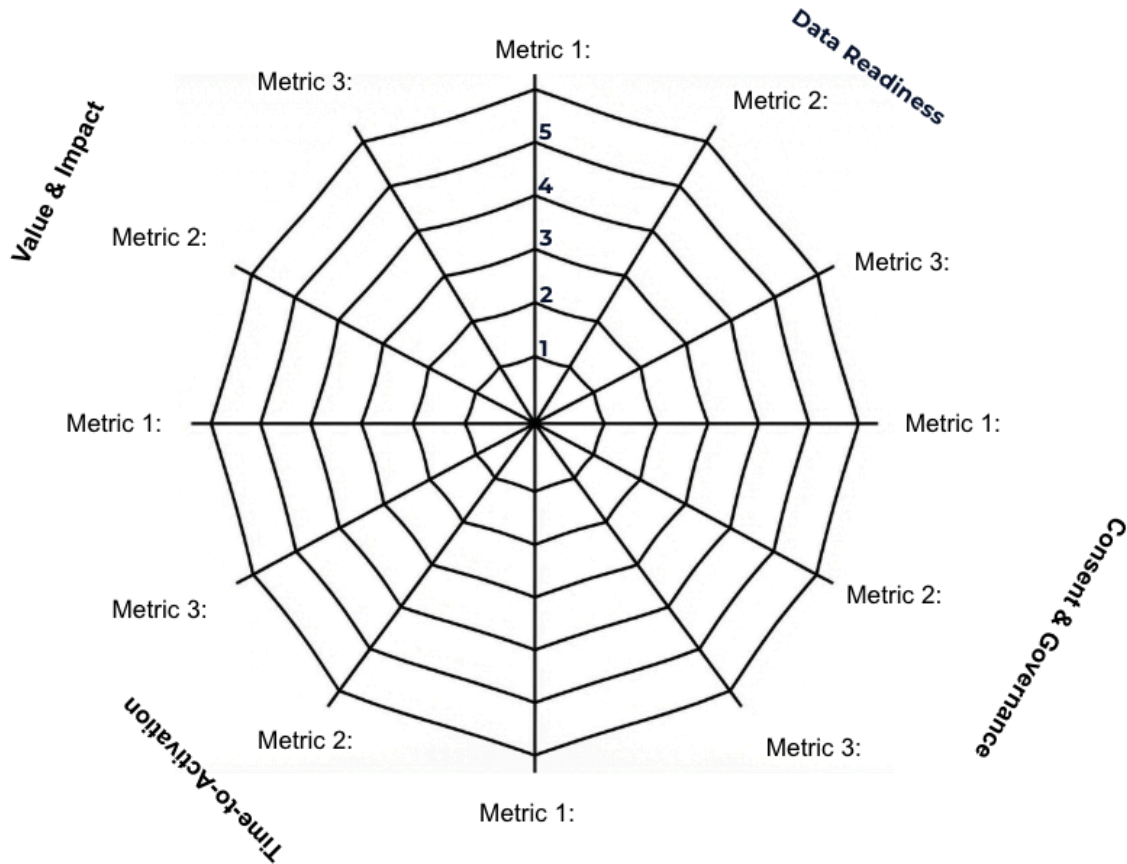
For each layer, list **2–3 desired metrics**, **measurement maturity (1–5)** [1 = not measured, 3 = measured ad hoc / siloed, 5 = measured, trusted, in regular reviews], **performance status** (Green / Amber / Red)

	Metric Desired	Maturity (1–5)	Status (G/A/R)	Notes (owner, sources, usage)
<b>Layer 1 – Data Readiness</b> (coverage, quality, latency, identity)				
Metric 1:				
Metric 2:				
Metric 3:				
<b>Layer 2 – Consent &amp; Governance</b> (coverage, mismatch rate, incidents, auditability)				
Metric 1:				
Metric 2:				
Metric 3:				
<b>Layer 3 – Time-to-Activation</b> (idea → data → model → first activation)				
Metric 1:				
Metric 2:				
Metric 3:				
<b>Layer 4 – Value &amp; Impact</b> (revenue, churn, cost, risk, CX, model utilization)				
Metric 1:				
Metric 2:				
Metric 3:				

**Step 3: Visualize Your Data Quality & Identify the Weaknesses - Draw Your Readiness Radar**

On a the spider chart below:

- Plot one point per layer using **average maturity** for that layer (1–5).
- Connect the points to visualize your **AI Data Readiness shape**.



**Step 4: Identify The Weakpoint**

- **Which Layer is Weakest?** \_\_\_\_\_
- **Why is it Weak?** \_\_\_\_\_

**Step 5: Identify Top Metric Improvement Actions**

Pick **1–2 metrics** to improve in the next 90 days and fill in the action table below.

Metric	Why it matters for AI readiness/risk:	Data source (Tealium / CDW / AI / destination):	Metric owner (team/role):	Top three actions to take right away:

## Exercise 5 – Modern AI Data Practice Outline

**Goal:** Turn the idea of a **Data for a Modern AI Data Practice** into an initial charter and near-term plan.

**Step 1:** Define the Need and Initial Success

Complete the prompts:

- Having a Modern AI Data Practice Is Important Because...:

---



---

- In 18–24 months, success looks like...(for data, AI, governance, business):

---



---

### Step 2: Outline Components of a Practice Charter

Purpose– This Practice Exists To:	
Scope – What We Focus On in Each Step (Connect, Govern, Transform, Activate):	
Connect:	
Govern:	
Transform:	
Activate:	
Initial AI Use Cases:	
Use case 1:	
Use case 2:	
Use case 3:	
Core Team (by role/team, not just names):	
Data/architecture:	
AI/data science/ML:	
Privacy/governance/risk:	
Business/prod/market:	
Stakeholders:	
Success Metrics (3–5 KPIs for the CoE)	
Data readiness:	
Consent & govern:	
Time-to-activation:	
Value & impact KPI:	
Operating Cadence	
Intake & priority forum(s):	
Review/steering cadence:	
Reporting rhythm:	

## Reference Measures for AI Data Practices

**1. Data Readiness:** Show whether inputs to AI are complete, consistent, identifiable, and timely for the specific AI use cases in scope—not just “lots of data.”

- **Coverage of required AI fields**

- Metric: % of model-critical fields populated per AI use case and per source.
  - Example: For churn model features, `completed_onboarding`, `last_login_ts`, and `plan_type` must be > 95% populated on all training and scoring records.
- Variants:
  - % of events/records dropping due to missing required fields
  - % of profiles w/ min feature bundles available (e.g., ≥ 10 of 12 required attributes)

- **Schema conformance & consistency**

- Metric:  $\text{Schema conformance rate} = 1 - (\text{invalid\_events} / \text{total\_events})$  for the model's upstream feeds.
  - Track by **event type** and **source** (web, app, POS, batch, etc.).
- What “good” looks like:
  - Critical AI events (e.g., `purchase`, `session_end`) have > 99% conformance; non-critical events can tolerate lower.

- **Identity resolution quality**

- Metrics:
  - **Identity coverage:** % of events linked to a persistent person/account ID vs. anonymous IDs.
  - **Match rate across systems:** % of warehouse/customer records that can be joined to Tealium profiles / first-party IDs.
  - **ID fragmentation:** average # of IDs per person; % of profiles suspected duplicates.
- Why it matters: Directly affects sample bias, double-counting, and the ability to attribute outcomes to AI decisions.

- **Ingestion & processing latency**

- Metrics:
  - **Event time → AI-available latency** (p50/p95) per use case (real-time vs batch).
  - **Model output freshness** (time from feature update → score updated in activation).
- Examples:
  - Real-time personalization: p95 < 500 ms from click to feature/segment update.
  - Daily risk scoring: batch completes by 07:00 local time with < 1% failures.

- **Source reliability & data health**

- Metrics:
  - Ingestion failure rate per connector/source.
  - % of days with full delivery for all model-critical feeds (SLO-style).

### How to use this:

- Set per-use-case SLOs (e.g., “churn model must have ≥ 98% feature completeness and p95 latency < 5 minutes”).
- Review these in a Data for AI readiness dashboard before green-lighting new use cases or model changes.

**2. Consent & Governance:** Ensure models only use data they're legally, ethically, and purpose-aligned to use—and that you can prove it under audit.

### Key metric families

- **Consent coverage**

- Metric: % of records used in AI training/inference with explicit, valid consent (or other lawful basis) recorded for the specific AI purpose.
- Slices: by region, channel, data domain (behavioral, transactional, CRM).

- **Consent mismatch rate**

- Metric: % of events/records where actual data use (e.g., “AI personalization”) is not permitted by stored consent purpose(s).
- Examples:
  - Data collected for “service improvement” being pulled into an “advertising / profiling” model without proper consent.
- Target: Drive toward near-zero mismatches for production models; any spike is a Sev-1 governance issue.

- **Policy enforcement success**

- Metric: % of policy checks passed vs. failed at the orchestration layer (e.g., Event Specs, consent filters, masking rules).
- High blocked event counts can indicate either:
  - A healthy control catching mis-configured feeds, or
  - A recurring design problem (e.g., teams consistently sending disallowed PII).

- **Restricted-data incidents & near misses**

- Metric: Count and severity of incidents where:
  - Restricted attributes were used in training/serving.
  - Non-compliant datasets reached an AI platform.
- CoE maturity signal: You eventually track near misses picked up by pre-production checks, not just post-incident reviews.

- **Audit readiness & lineage completeness**

- Metrics:
  - % of production models with documented data lineage (sources, consent basis, purposes, transformations).
  - % of AI training datasets with an approved “AI Data Use / Governance Checklist” on file.
- Goal: For any high-impact model, you can answer quickly: “Which data, from where, under which consent, with which policies, trained this?”

- **Governance cycle time**

- Metric: Median time from AI data request → governance decision (approve/modify/reject).
- Why it matters: Measures whether governance is an accelerator or bottleneck for AI.

### How to use this:

- Make consent mismatch rate & audit readiness non-negotiable for models going live.
- Publish a quarterly governance scorecard: coverage, mismatches, incidents, and time-to-decision for AI initiatives.

**3. Time-to-Activation:** Measure how quickly the organization can move from idea → data → model → real actions in channels, not just “model in a notebook.”

### Key metric families

- **End-to-end time-to-activation**

- Metric: Median days from idea logged → first production activation per AI use case type.
- Use: Compare across business units; identify bottlenecks (data engineering, consent review, MLOps, channel integration).

- **Stage-level cycle times**

- Metrics (median/p75):
  - Idea → Data Ready
  - Data Ready → Model Live
  - Model Live → First Activation
- Example insight: If Data Ready → Model Live is fast but Model Live → Activation is slow, the constraint is channel orchestration / ownership, not data science.

- **Activation coverage**

- Metric: % of production models that are actually activated in ≥ 1 channel and % in ≥ 2 channels.
- Why it matters: Highlights models that are “deployed” but unused.

- **Change lead time**

- Metric: Time from new signal/feature identified → available to models and channels.
- Important for fast-moving domains (fraud, pricing, marketing optimization).

- **Failure / rollback rate at activation**

- Metric: % of AI activations rolled back or disabled within 90 days due to data, governance, or performance issues.
- High rates point back to weaknesses in Data Readiness or Consent & Governance stages.

### How to use this:

- Set explicit **time-to-activation targets** for priority AI use cases (e.g., “< 90 days from idea to first activation”).
- Use these metrics in CoE intake and retrospectives to show how patterns and shared components (identity, consent, orchestration) are reducing time-to-value over time.

**4. Value & Impact:** Demonstrate that AI is changing outcomes, not just generating scores—tying CoE work to revenue, cost, risk, and CX.

### Key metric families

- **Incremental revenue / conversion**

- Metrics:
  - Lift in conversion rate, average order value, or upsell vs. a control group or pre-AI base.
  - **Incremental revenue attributable to AI = (AI group revenue – control group revenue) – (baseline trend).**
- Examples:
  - “Next-best-offer increased campaign revenue by 8% vs business-as-usual over 6 weeks.”

- **Churn & retention**

- Metrics:
  - Change in churn rate, renewal rate, or retention curves where AI is applied vs. not.
  - **Net retention uplift** for AI-driven interventions (e.g., risk / outreach).

- **Cost savings & efficiency**

- Metrics:
  - Reduction in manual workload (e.g., fewer manual reviews per 1,000 transactions).
  - Lower cost-to-serve (e.g., deflected calls, more self-service completions).
  - Infra / data pipeline cost avoided through better minimization and governance.
- Example: Fraud model reduce false positives → fewer escalations → hours saved per month.

- **Risk avoidance**

- Metrics:
  - Fewer governance incidents related to AI (consent, privacy, bias).
  - Regulatory findings or customer complaints avoided; where possible, estimate financial exposure avoided.
  - Early-warning indicators from AI data metrics (e.g., consent mismatch spikes).

- **Customer experience (CX) improvements**

- Metrics:
  - Changes in NPS/CSAT for journeys influenced by AI (e.g., personalization, support).
  - First-contact resolution, handle time, task completion rate for AI-supported flows.
  - Qualitative indicators: complaint volume, opt-out rates, or “this feels creepy” feedback.

- **Model utilization & sustainability**

- Metrics:
  - Model utilization rate: **% of models actively used in decision flows vs. “shelfware”.**
  - Average lifetime of models in production before abandonment.
- Why it matters: Shows whether the org can sustain AI, not just launch pilots.

### How to use this:

- Require every AI use case to have: 1) 1–2 primary business KPIs (revenue, churn, cost, risk, CX) and 2) at least one activation metric (e.g., **% of predictions actually acted on**).
- Roll into an executive AI Data & Impact scorecard—top-line view of readiness + realized value for the AI portfolio.

## Exercise 6 – Defining a Data Orchestration Strategist

**Goal:** Define a version of the **Data Orchestration Strategist** role that fits your organization and set up a basic success scorecard.

### Step 1: Outline the Role Description

Role title (in your org terms):				
Purpose-“This role exists to...”				
Scope & Mandate – main spheres				
Data strategy & roadmap:				
Use-case design & translation:				
Business case & ROI:				
Governance & compliance integration:				
Capability & org development:				
Top Responsibilities & Deliverables (3–5)				
Item 1				
Item 2				
Item 3				
Item 4				
Item 5				
Org Placement & Stakeholders				
Where role sits to start?				
Where sit long term?				
Top stakeholders?				
12-Month Priority Outcomes				
Strategy & roadmap:				
Wins on top use cases:				
Improved readiness:				
Embedded governance:				
Organizational standing:				
Role Success Scorecard				
KPI	Type (Functional / Initiative / Governance)	How measured	Data source	Review cadence

# Data Strategist – Role Outline for Consideration

## 1. Role Purpose

Owns the end-to-end strategy for how customer data is collected, governed, and activated to drive revenue growth, reduce cost and risk, and enable AI-ready, data-driven experiences across the organization.

## 2. Role Scope

Primary mandate

- Turn fragmented customer data, underused platforms, and disconnected teams into a cohesive, high-impact customer data capability.
- Bridge marketing/business, data/engineering, and governance/compliance so that data work maps directly to measurable business outcomes.

Core spheres

- Data strategy & roadmap
- Use-case design & translation
- Business case & ROI
- Governance & compliance integration
- Capability & org development

## 3. Top Responsibilities & Deliverables

### A. Customer Data Strategy & Roadmap

Responsibilities

- Lead assessment of current customer data landscape and maturity.
- Define target-state data flows, unified profile strategy, and activation patterns.
- Maintain a rolling 18-month roadmap for customer-data initiatives.

Key deliverables

- Current-state data landscape & maturity assessment (sources, owners, usage, gaps).
- Written customer data strategy (principles, target architecture, priority use cases).
- 90 / 6 / 18-month roadmap with sequenced initiatives and success criteria.

### B. Translation of Business Needs into Data Requirements

Responsibilities

- Facilitate requirements between marketing/product and data/engineering.
- Convert business goals into concrete data/tech specs (events, IDs, attributes, latency).
- Ensure every major initiative has clear, feasible data requirements up front.

Key deliverables

- Standard use-case brief template shared across teams.
- Data & identity specs for top 5–10 high-value use cases (per year).
- Prioritized use-case backlog with effort vs. impact estimates.

## C. Business Case, Metrics, and ROI

### Responsibilities

- Build and maintain business cases for customer-data investments.
- Define success metrics and baselines for each initiative.
- Report realized value and learnings back to leadership.

### Key deliverables

- Business case library (3–5 flagship use cases with modeled impact).
- Standard metrics framework (speed, precision, scale, risk) for data initiatives.
- Quarterly impact report tying data work to revenue, cost, and risk outcomes.

## D. Governance, Compliance, and Risk Integration

### Responsibilities

- Partner with legal, security, and privacy teams to embed requirements into data flows.
- Define ownership, RACI, and guardrails for collection, storage, activation, and retention.
- Ensure new use cases adhere to agreed governance patterns.

### Key deliverables

- Customer data governance playbook (roles, policies, decision rights, escalation).
- Set of standard patterns for consent, identity, and activation (e.g., reference architectures).
- Audit-ready evidence pack for use cases (data lineage, consent handling, access control).

## E. Capability & Org Development

### Responsibilities

- Define competencies for the Customer Data Strategist and related roles.
- Identify and coach internal talent who can evolve into strategist roles.
- Shape operating model between CMO/CDO/IT and line-of-business teams.

### Key deliverables

- Role & competency model for strategists and adjacent functions.
- Development plans for at least 1–3 internal candidates (rotations, mentoring, exposure).
- Documented operating model (org placement, engagement model, forums, cadences).

## 4. 12-Month Priority Outcomes

By the end of year 1, a high-performing Customer Data Strategist will have:

1. Strategy & roadmap in place
  - Published and socialized customer data strategy and 18-month roadmap.
  - At least one pilot use case designed, delivered, and measured end-to-end.
2. Two to three measurable wins
  - 1–3 initiatives showing clear uplift (e.g., +X% conversion / CTR, reduced time-to-launch, reduced media waste) attributed to improved data strategy.
3. Improved readiness
  - Additional “Yes” answers on readiness checklist compared to baseline.
4. Embedded governance
  - Standard consent/identity/activation patterns adopted by all *new* major use cases.
  - No material compliance “surprises” tied to customer data.
5. Standing in the org
  - Recognized by C-level sponsors as the go-to owner for customer data strategy and as a key voice in AI and personalization decisions.

## 5. Success Metrics (Scorecard)

### Functional KPIs

- Goal-achievement uplift
  - Movement toward higher initiative success rates for data-driven projects (e.g., closer to the ~89% benchmark with strong data strategy).
- Time-to-value
  - Average reduction in time from idea → live experiment for prioritized use cases (target reductions of 3–10x on at least one major initiative).
- Platform & data utilization
  - % of major campaigns using the unified profile and governed data attributes.
  - Increase in feature utilization for key platforms (CDP, warehouse, analytics).

### Initiative-Level KPIs (per flagship use case)

- Speed: Time from approved brief → first live activation.
- Precision: Uplift in engagement/conversion, reduced wasted media or mistargeted contacts.
- Scale: Number of markets/brands/channels using strategist-defined data flows.
- Risk/quality: Reduction in data incidents, rework, or compliance escalations.

### Governance & Readiness KPIs

- Readiness score
  - Baseline vs. current score on the 10-question readiness checklist.
- Policy adherence
  - % of new initiatives following defined consent/identity/governance patterns.
- Audit & incident trend
  - Fewer critical findings or customer-data-related incidents year over year.

## 6. Key Stakeholders

- Executive sponsors: CMO, CDO, CIO (or equivalent)
  - Primary partners:
    - Marketing & CX leadership, Marketing Ops
    - Data engineering, analytics/BI, architecture
    - Product management / digital experience
    - Privacy, legal, security, risk & compliance

## 7. Core Competencies (Profile Snapshot)

- *Technical literacy*: SQL & basic scripting; CDP/warehouse, identity resolution, data architectures.
- *Business acumen*: Commercial orientation; can build ROI cases, connect data work to P&L and KPIs.
- *Bridge & influence*: Strong facilitator, communicator, and change-agent across technical & business teams.
- *Strategic thinking*: Systems thinker able to prioritize high-value use cases and sequence delivery.
- *Data storytelling*: Comfortable presenting complex data and trade-offs in simple, exec-ready narratives.

# Exercise 7 – Bonus Round: Building a Tealium-Powered AI Data Orchestration Map

**Goal:** Map how Tealium (or an equivalent orchestration layer) connects your **sources, governance, transformations, AI platforms, and activation endpoints**.

If you don't use Tealium yet, use this as a **desired-state map**.

## 7.1 Inputs & Outputs

**Key input sources (today or planned):**

- Web / app (iQ / Collect / SDKs):
- Server-side / APIs / backend:
- CRM / marketing platforms:
- Offline files / batch imports:
- Data warehouse / lakehouse:

**Key activation endpoints:**

- ESP / marketing automation:
- Ad platforms:
- Onsite / in-app personalization:
- Contact center / support tools:
- Product or in-product experiences:
- Analytics / BI tools:

## 7.2 Connect–Govern–Transform–Activate Grid

Use the matrix to place systems, controls, and transformations.

Row \ Column	<b>CONNECT</b> (collection & routing)	<b>GOVERN</b> (consent & policies)	<b>TRANSFORM</b> (features & profiles)	<b>ACTIVATE</b> (channels & tools)
<b>Data sources</b>				
<b>Governance controls</b>				
<b>Transforms / features</b>				
<b>Activation platforms</b>				

Ideas for what to note in cells:

- **Products / systems** (e.g., Tealium Collect, EventStream, AudienceStream, CDW, AI platform, ESP, ad tools).
- **Policies / checks** (consent filters, masking, specs, approvals).
- **Data flows** (real-time vs batch, model scores in/out).

### 7.3 Gaps, Risks, and Quick Wins

Mark with symbols:

- ★ **Quick win**
- △ **Risk / gap**

For example:

- A source not connected or standardized.
- No consent enforcement before sending data to AI.
- AI scores not flowing back into orchestration for activation.

**List your top 3 orchestration opportunities:**

- 1.
- 2.
- 3.

**Which one would you most like to address in the next 90 days, and how could Tealium (or an equivalent layer) help?**

---

Final Reflection

- One insight you're taking back to your team about **data for AI**:
- One conversation you will schedule in the next month as a result of this workshop (with whom, about what):

Use this workbook as a **living artifact** — revisit each section quarterly as your AI data maturity and CoE evolve.