

GET THE FUTURE  
YOU WANT

# Data Classification and Synthetic Data

Capgemini's Response to FirmX

August 2025

Capgemini





## Notice to the recipient of this proposal ("Collins")

*The information contained herein, as well as any information shared by Capgemini in furtherance of this proposal, are the proprietary and confidential information ("Confidential Information") of Capgemini, and their release would offer substantial benefit to competitors offering similar services. The Confidential Information includes descriptions of methodologies and concepts derived through substantial research and development efforts undertaken by Capgemini. Recipient shall use this proposal and/or any Confidential Information solely for the purpose of evaluation for contract award. If Capgemini and Recipient are not otherwise parties to a non-disclosure agreement which protects Capgemini's Confidential Information herein, Recipient agrees to maintain in confidence and not to disclose the Confidential Information except to its employees or agents on a "need to know basis." Your acceptance of the submission of this proposal and evaluation thereof indicates your agreement to these terms. It is the position of Capgemini that this proposal and/or Confidential Information are not considered subject to release under the Freedom of Information Act, if applicable.*

*Copyright 2025 Capgemini America, Inc. All rights reserved. No part of this proposal may be reproduced by any means or transmitted without the prior written permission of Capgemini except internally by you for the purpose of evaluating this proposal. All copies of this proposal (or any portion hereof) should be returned to Capgemini or, at Capgemini's option, destroyed at the end of the proposal evaluation period if Capgemini is not selected.*

*Neither submission by Capgemini nor your acceptance of this proposal, in whole or in part, constitutes acceptance by Capgemini of any contractual terms contained in your Request for Proposal or similar document ("RFP"), if any, and shall not form a binding agreement between the parties, other than with respect to confidentiality as set forth herein. Such an agreement shall only exist upon the execution of a mutually acceptable contract by both parties ("definitive agreement"), which will contain the exclusive agreement between the parties regarding the scope and terms of Capgemini's engagement. Failure by Capgemini to comment upon any contractual terms in the RFP shall not constitute acceptance thereof. The RFP, and this proposal, will be superseded by the definitive agreement. Any proposed duties or obligations referenced herein should be understood as being subject to commercially reasonable efforts. Capgemini makes no representations, warranties, or assurances to you, except as may be mutually agreed in the definitive agreement. The terms of this Notice shall control over the terms in any RFP in the event of a conflict or inconsistency.*

*If the services involve third party components (including software and/or cloud services), open source software, or artificial intelligence enabled tools or materials, the parties will work together to address the terms applicable to such items. Client's access and use of such items shall be governed by the applicable third-party terms and conditions.*

*The submission of this proposal is expressly subject to and conditioned upon the enforceability of the terms of this Notice. In the event that any of the terms of this Notice are held not to be enforceable, then, at Capgemini's option, Capgemini may withdraw this proposal without liability of any nature.*

*The term "Capgemini" appearing in this proposal may refer to Capgemini America, Inc., or to one or more of its global affiliates. However, this proposal is being submitted only by Capgemini America, Inc., which is solely responsible for its contents, and Capgemini America, Inc. shall be the contracting entity if its proposal is selected.*

**Capgemini America, Inc.**



# Agenda

- 01 Executive Summary**
- 02 Program Objectives & Our understanding of scope**
- 03 POC Solution Approach**
- 04 Implementation Approach**
- 05 Commercials**
- 06 Acceptance Criteria**
- 07 Case studies**
- 08 Capgemini capabilities**



# 01

## Executive Summary





# Executive Summary

Dear Collins team,

We are excited about the possibilities presented by this opportunity to be chosen as a long-term strategic partner for FirmX.

This proposal outlines a strategic initiative for FirmX to enhance the classification and handling of International Traffic in Arms Regulations (ITAR)-controlled items, Controlled Unclassified Information (CUI), and Personally Identifiable Information (PII) while leveraging generative AI for synthetic data generation. The solution aims to ensure compliance with stringent regulatory requirements, mitigate risks of unauthorized access, and accelerate secure innovation in defence technology development. The defence industry faces increasing complexity in managing ITAR-controlled technical data, CUI, and PII amidst evolving regulatory landscapes and cybersecurity threats.

Our solution integrates advanced classification frameworks with generative AI to address compliance and data challenges and plans to utilize AI-driven metadata tagging and natural language processing (NLP) to streamline classification, ensuring compliance with DDTG guidelines and reducing human error and will implement generative AI models to create synthetic datasets mimicking ITAR, CUI, and PII characteristics without containing sensitive information.

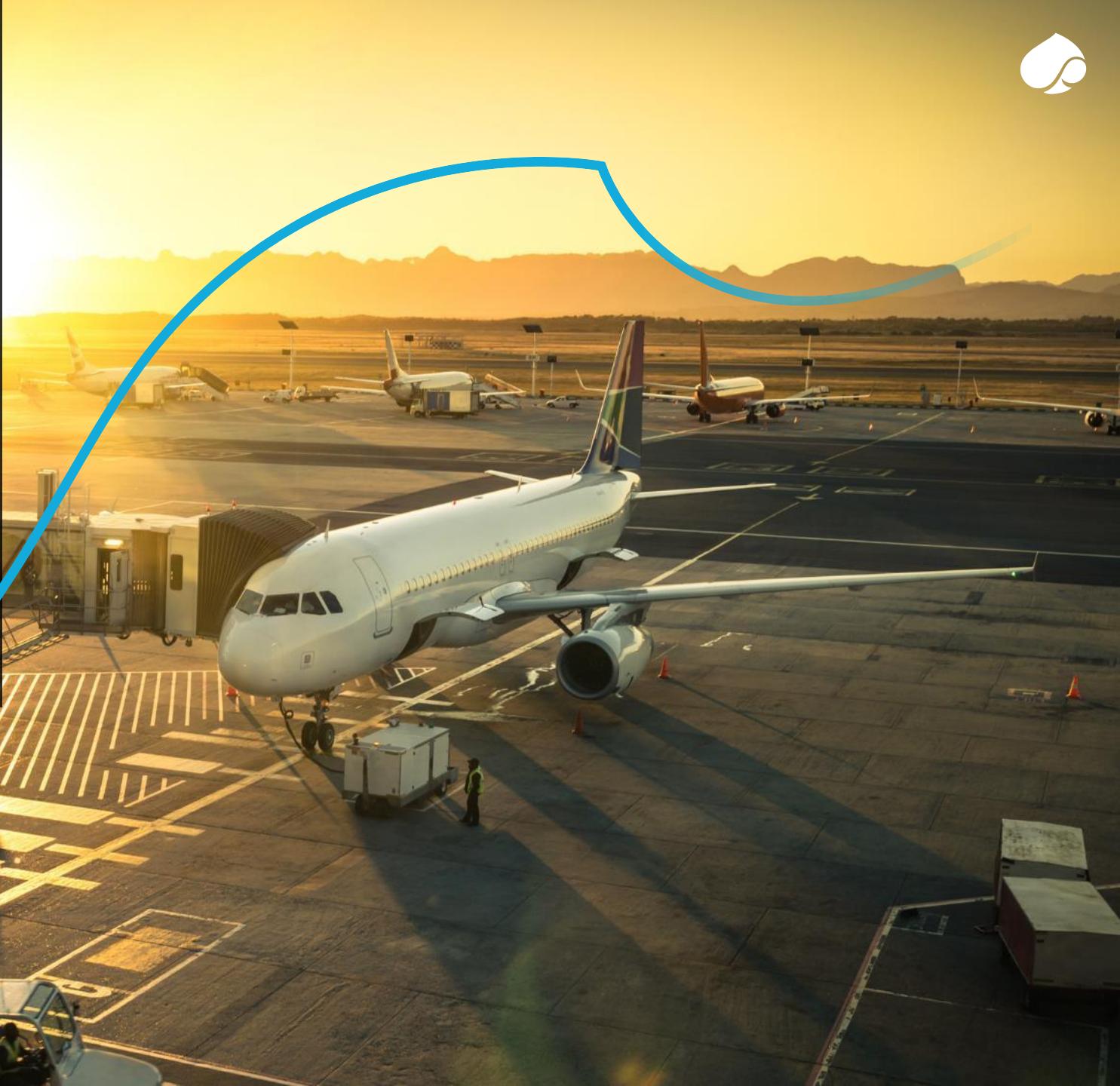
## Key Benefits

- **Compliance Assurance:** Automated classification reduces risks of ITAR violations, CUI mishandling, and PII breaches, ensuring adherence to federal regulations.
- **Innovation Acceleration:** Synthetic data enables secure collaboration with global partners and accelerates AI-driven R&D without compromising sensitive information.



# 01

## Program Objective & Scope





# Our Understanding

Collins operating in the aerospace and defense sector, is focused on modernizing its data and analytics platform to ensure compliance with export control regulations (ITAR, SCOMET, EAR) and enhance data accessibility, scalability, and governance. The current manual classification process is slow, resource-dependent, and non-scalable, necessitating an automated solution.

## OBJECTIVES

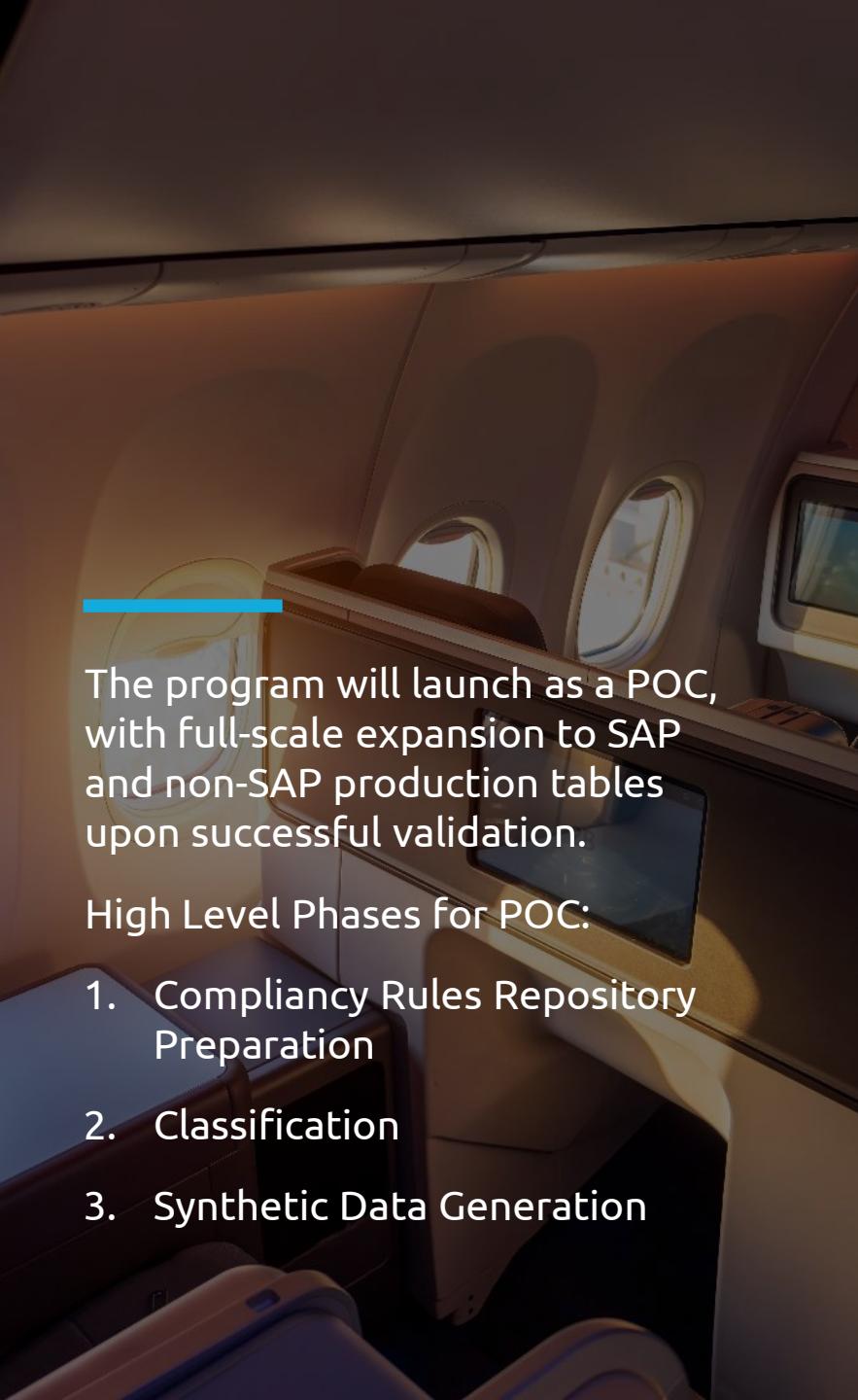
- **Automate classification** to reduce manual effort and improve scalability.
- **Enable secure development, testing, and collaboration** using synthetic data.
- **Ensure compliance** with ITAR, PII, while maintaining data fidelity.
- **Deliver Synthetic Data Generation** of 50 SAP tables for MM/PP/SD tables.

## KEY GOALS ARE TO:

- Automate **granular data classification** for ITAR and PII compliance.
- Generate **realistic, privacy-compliant synthetic datasets** with full referential integrity.
- Simulate one or two cases for **complete end-to-end SAP business workflows** in MM, SD, and PP domains.
- Support **classification and generation for 50 SAP tables** (38 master, 12 transactional).
- Enable **SME validation feedback loops** and allow **custom user-defined inputs**.
- Operate **securely in Collins environment** without external data sharing or sandbox dependency.

## Expected timeline:

- August 25th - SoW – 1 week
- Team Onboarding : 4 weeks lead time
- September 22nd – POC Kick off



The program will launch as a POC, with full-scale expansion to SAP and non-SAP production tables upon successful validation.

## High Level Phases for POC:

1. Compliancy Rules Repository Preparation
2. Classification
3. Synthetic Data Generation



# High Level scope

## Scope of Activities

- Define POC architecture for data classification, and synthetic data generation approach.
- Detail migration plan for data extraction, classification, and synthetic data creation for 50 SAP tables across 3 domains: **MM** (Material Management), **SD** (Sales & Distribution), **PP** (Production Planning) Align with compliance requirements (ITAR, and PII)

## Execution Scope with Testing / Validations

### Project Initiation

- **Team Onboarding:** Onboard project team members and define roles and responsibilities.
- **Environment Setup:** Configure Capgemini environment, Git repositories, IDE and required AI/GenAI libraries setup for collaboration.
- **Deliverables:** Onboarding completion for team members, environment setup, and access configuration for team.

### Discovery & Preparation

- **Data Extraction and Business Glossary Preparation:** Extract metadata and prepare business glossary including Table name, Column Names, column datatype, Significance, Referential Integrity details and Data Owner with Functional SME input.
- **Business Glossary Review:** Validate business glossary with Defense Data SMEs
- **Classification Rules Repository:** Define and review classification rules for ITAR and PII compliance with client defense SMEs and Signoff.
- **SAP Table Data and Compliance Analysis:** Understand SAP table data complexity and referential integrity.
- **Deliverables:** Extracted metadata, validated business glossary, classification rules repository, and compliance analysis report.

### Data Preparation and Classification

- **Sample Synthetic Data Generation:** Generate sample synthetic data for 50 SAP tables and validate with client Defense Data SMEs.
- **Generate 5K Data for Model Training:** Create 5K synthetic records for model training from the sample data.
- **Classification Engine for Table/Column Classification:** Develop classification engine using GenAI and compliance rules.
- **Data Labeling and Vectorization:** Label data using GenAI and vectorize compliance rules for classification.
- **Row-Level Classification:** Implement row-level classification with explainability, by testing both LLM and AI/NLP approaches and select the best fit.

- **Deliverables:** Sample synthetic data, 5K training records, classification engine, labeled data, and row-level classification results, Signoff from client Defense Data SME.

### UI Development and Validation

- **Streamlit UI Development:** Develop a simple UI for SME validation.
- **SME Validation via UI:** Enable SMEs to download and validate files through the UI.
- **Deliverables:** Functional Streamlit UI, validated classification results, and SME feedback.

### Feedback Loop and Retraining

- **Feedback Loop and Retraining:** Incorporate SME feedback and retrain the model.
- **Deliverables:** Retrained model and updated classification rules.

### Synthetic Data Generation

- **Sample Data Creation:** Create sample data with proper data distribution using automated scripts to generate Synthetic Data.
- **Synthetic Data Generation:** Generate 30-35K synthetic records with referential integrity.
- **Deliverables:** Automated script for data distribution, 30-35K synthetic records.

### Testing and Validation

- **Testing and Validation:** Validate synthetic data and classification as per compliance rules.
- **Deliverables:** Test reports and compliance validation results.

Final Presentation and Handover

NO additional implied scope is assumed beyond this scope page.



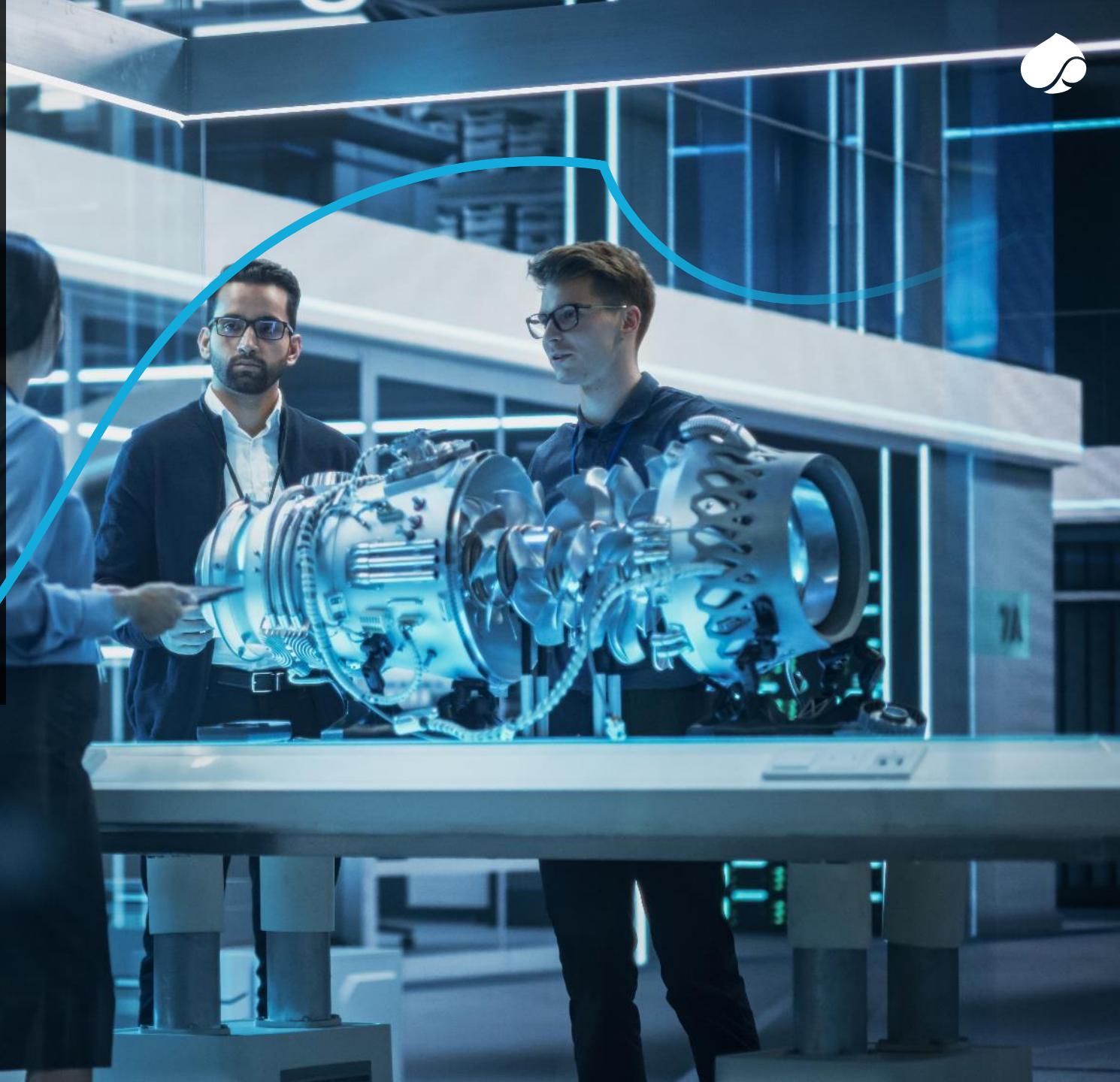
# Out of scope

- Classification of non-SAP systems
- Fully automated classification without human review
- Real-time classification of new data
- Integration to IMMUTA and CLARA.
- Full-scale production deployment for all SAP and non-SAP tables , Tables outside MM (Materials Management), SD (Sales & Distribution), and PP (Production Planning).
- Authorization and authentication for POC
- Real-time integration with non-SAP systems (e.g., CRM, ERP).
- Custom ABAP developments outside the scope of metadata extraction
- Data masking implementation for production scenarios.
- Integration with all upstream/downstream systems beyond POC scope.
- Deployment on public cloud platforms (e.g., AWS, Azure). Only on-premise/internal server deployment is included.
- Full-scale end-user training programs.
- Hypercare and UAT
- Long-term maintenance and support.
- Custom development for non-standard business logic.
- Establishment of a full data governance framework.
- Performance optimization for large-scale data beyond POC dataset.
- Procurement and management of third-party software licenses.
- Full-scale security and compliance audits.
- Advanced analytics and AI model enhancements beyond POC requirements.
- Classification of unstructured data (e.g., PDFs, images) outside SAP tables.

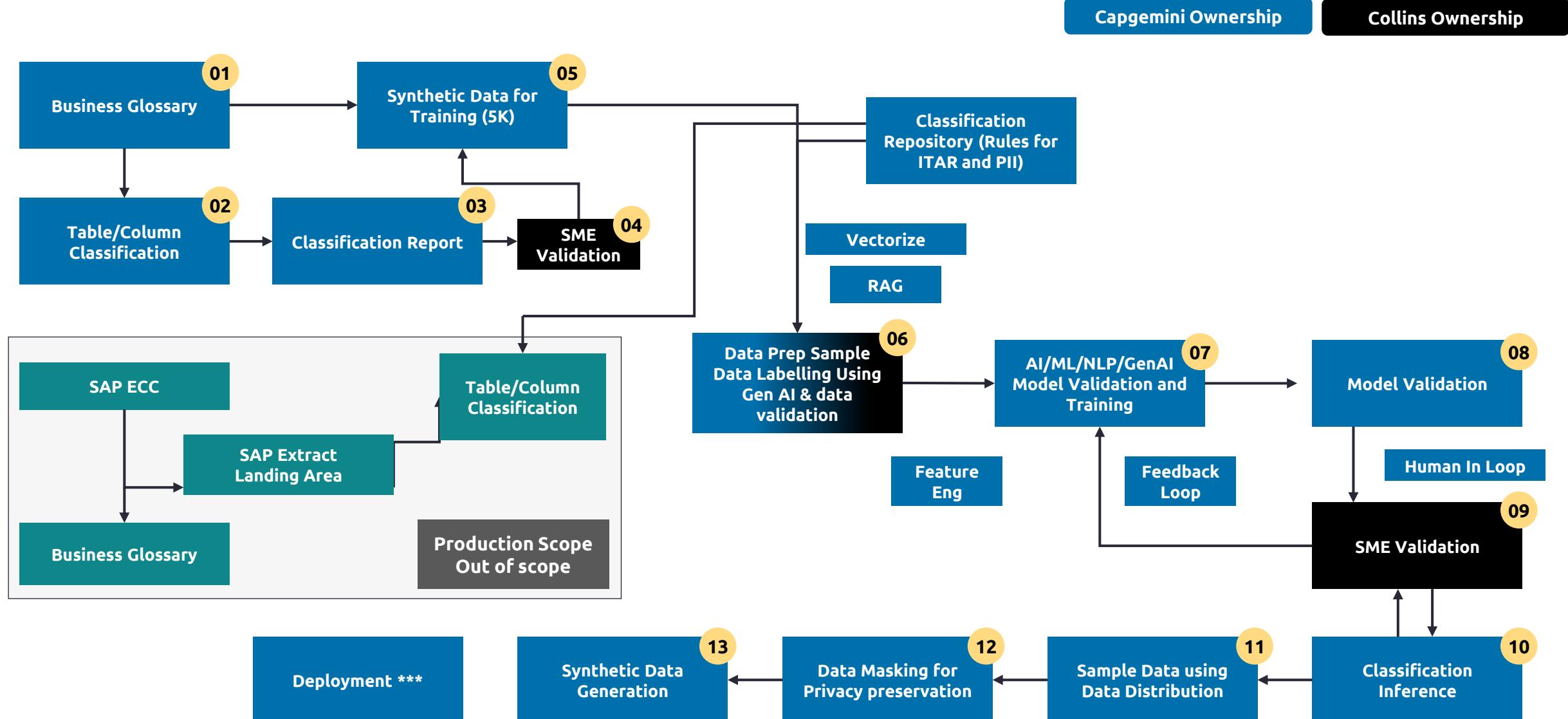


# 03

## POC Solution Approach



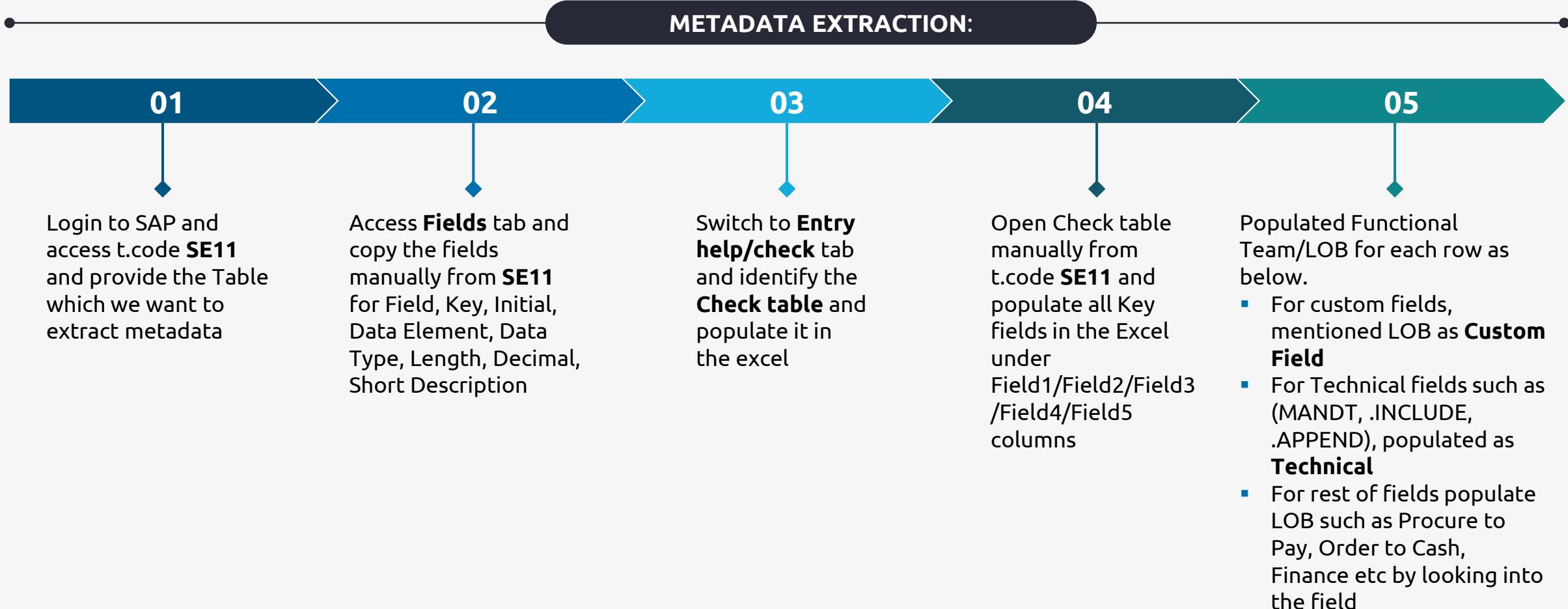
# Overall Solution Approach





# Our Approach To Enable Metadata Extraction

SAP System Version: SAP ECC 6.0 EHP 6





# 04

## Implementation Approach



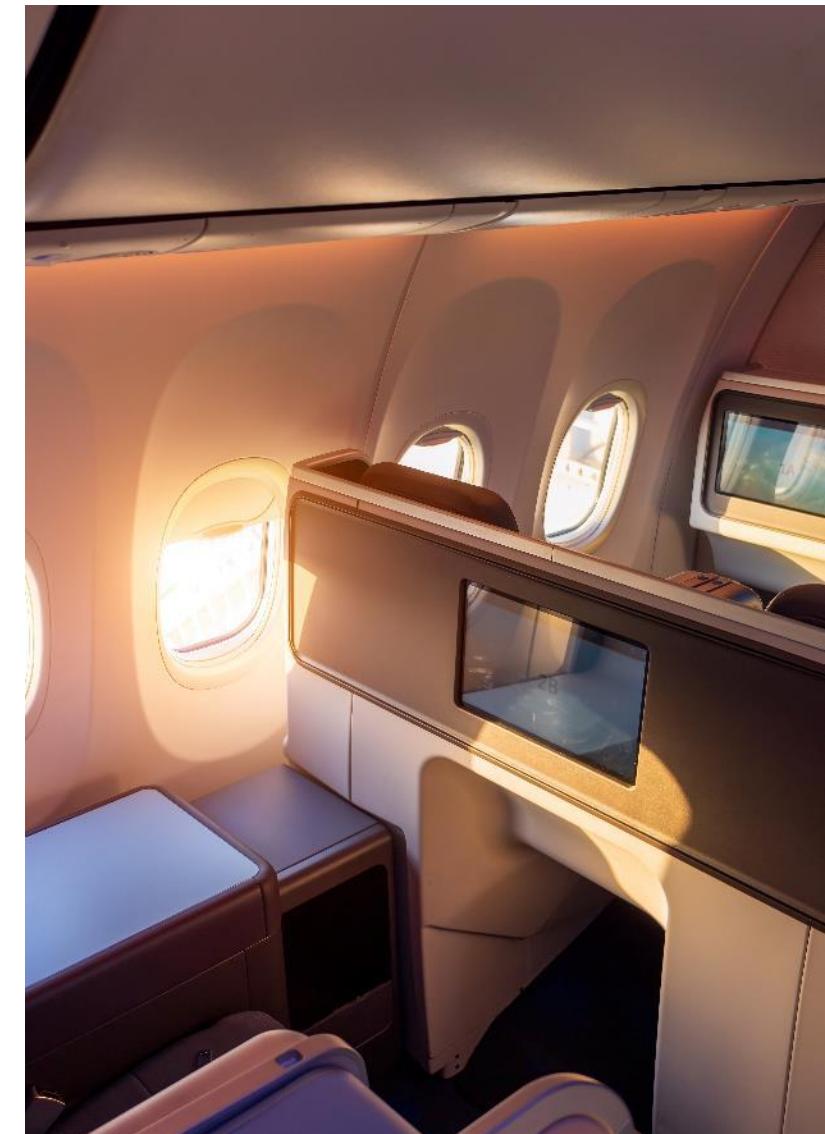
# Implementation Plan

	Capgemini Ownership									Collins Ownership	
	Key Activities	W0	W1	W2	W3	W4	W5	W6	W7	W8	W9
Kick - Off	<ul style="list-style-type: none"> <li>Program Governance &amp; Functional Expertise</li> <li>Team Onboarding</li> <li>Capgemini Environment Setup and Availability with git Configuration</li> </ul>										
Discovery & Preparation	<ul style="list-style-type: none"> <li>Data Extraction and Business Glossary Preparation Including Functional SME</li> <li>Classification Rules Repository for ITAR and PII etc compliancy</li> <li>SAP Table Data and Compliancy complexity understanding w.r.t. Referential integrity</li> <li>Sample Synthetic Data Generation for 50 SAP Tables and SME Validation and for model training</li> <li>Classification Engine for Table/Column Classification using GenAI + Compliancy Repository Rules</li> <li>Data Labelling using GenAI and Vectorise the Compliancy Rules for classification</li> <li>Classification Engine - Row Level Classification for 50 Tables along with Explainability (LLM or AI/NLP)</li> <li>Streamlit UI - Simple</li> <li>SME Validation via UI (Download file for validation)</li> <li>Feedback Loop and retraining</li> <li>Sample data creation with proper Data Distribution (Via Automated script)</li> <li>Data Masking (As per POC scope only)</li> <li>Synthetic Data Generation Implementation with referential Integrity and from Sample Data with Referential Integrity- 50K Records</li> <li>Testing and Validation as per compliancy rules</li> <li>Forward Planning Dialogue with Collins team</li> </ul>										



# Activities & deliverables

Phase	Key Activities	Deliverables
Project Initiation	Team Onboarding	Onboarding documentation and access setup
	Capgemini Environment Setup and Git Configuration	Configured VM and AI GenAI Libraries, LLM and Git repository
Discovery	Data Extraction and Business Glossary Preparation (Including Functional SME)	Extracted metadata and draft business glossary
	Business Glossary Review by Functional SME	Validated business glossary
	Classification Rules Repository for ITAR and PII, Compliance and Collins Compliance SME Review	Documented classification rules and Client SME review and Signoff.
	SAP Table Data and Compliance Complexity Analysis (Referential Integrity)	Analysis and understanding on data complexity and referential integrity
	Sample Synthetic Data Generation for 50 SAP Tables and SME Validation	Sample synthetic datasets and SME validation report
Development	Generate 5K Data for 50 Tables for Model Training	5K synthetic records for model training
	Classification Engine for Table/Column Classification (GenAI + Compliance Rules)	Developed classification engine for table/column tagging
	Data Labeling Using GenAI and Vectorizing Compliance Rules	Labeled data and vectorized compliance rules
	Classification Engine for Row-Level Classification with Explainability (LLM vs. AI/NLP Comparison)	Row-level classification engine with explainability report and approach recommendation
UI Development	Streamlit UI Development	Simple Functional Streamlit UI for SME validation
Validation	SME Validation via UI (Download File for Validation)	Validated classification results and SME feedback





# Scope considerations and Dependencies (1/2)

Areas	Scope considerations	Areas	Dependencies
Data	This POC does not involve actual data; synthetic data will follow classification.	SME Support	Collins will provide SMEs for business glossary, classification, and synthetic data validation.
Client Collaboration	Data owners are available for validation & timely SME availability from Collins for validation and feedback.	Feedback Timelines	Collins will provide timely feedback during data validation and classification phases.
Environment & Tools	All work will occur in Capgemini's own unclassified environments. Capgemini will setup on-prem VM environment & Git to support the POC Scope. Production Environment is considered out of scope for this POC.	Compliance Approval	Collins compliance team approves classification rules and synthetic data generation methods.
Compliance	Collins compliance team will validate classification rules for ITAR and PII.	Timeline Adherence	Collin's SME feedback timelines are adhered to, to avoid delays in subsequent phases.
Technical Feasibility	GenAI models and classification engines are performing as expected with the provided data; however, any deviation may impact the project timeline.	Sign-Off	Collins sign-off is critical for the data validation plan, any deviation may impact the project timeline.
UI Usability	Streamlit UI will meet basic usability requirements for Collin's SME validation.	Stakeholder Buy-In	Collins communicates POC objectives and secures stakeholder buy-in, any deviation may impact the project timeline.
Scope Consideration	POC scope is limited to 50 SAP tables, excluding full-scale production deployment.	Post-POC Ownership	Collins defines ownership and next steps for full-scale Production deployment post-POC



# Scope considerations and Dependencies (2/2)

- FirmX is responsible for identifying the Export Control Categorization Number (ECCN) for all materials that Capgemini works on.
- No data worked on will be transferred to any 3rd party including Capgemini.
- **We will provide on best effort basis sample organization structure across the 3 domains of SAP (MM, SD, and PP) to enable Collins in identifying the key roles responsible for data validation.**
- **Sample Organizational Structure for Data Validation wherein the roles will be structured by functional domains to ensure clear ownership and accountability across the entire data lifecycle for instance as under:**
  - Chief Data Officer (CDO) / Program Sponsor
  - Data Governance Council
    - Domain Data Leads
    - Data Owners
    - Data Stewards
- **Basic UI will be delivered solely for demonstration purposes. SME validation will be conducted through exported CSV files. A fully enhanced UI/UX design is expected to be implemented during the production phase.**
- Production Data must always be resident in FirmX systems, should never be transferred to Capgemini systems and should never be transferred by e-mail, even within FirmX systems. Production data will not be used by Capgemini teams. All production data must be stored on FirmX drive / network and must not be brought into Capgemini networks or storage devices.
- Capgemini assumes that generated synthetic data will not contain actual ITAR-controlled technical data or CUI but will only mimic statistical properties, structures, or patterns of the original data without revealing sensitive information. If synthetic data inadvertently becomes ITAR-restricted (e.g., due to fidelity levels), FirmX will handle reclassification and any export control implications.
- Capgemini does not require direct access to classified networks or systems; **all work will occur in Capgemini's own unclassified environments.**
- This POC qualifies as "fundamental research", exempting it from certain export controls unless otherwise specified by statute or regulation reported by Collins.
- Capgemini assumes the project scope is limited to synthetic data generation and classification algorithms, excluding deployment, integration into operational systems, or handling of live ITAR/ PII beyond prototype testing.
- Capgemini will not be responsible for third party tools CUI/ITAR compliance.



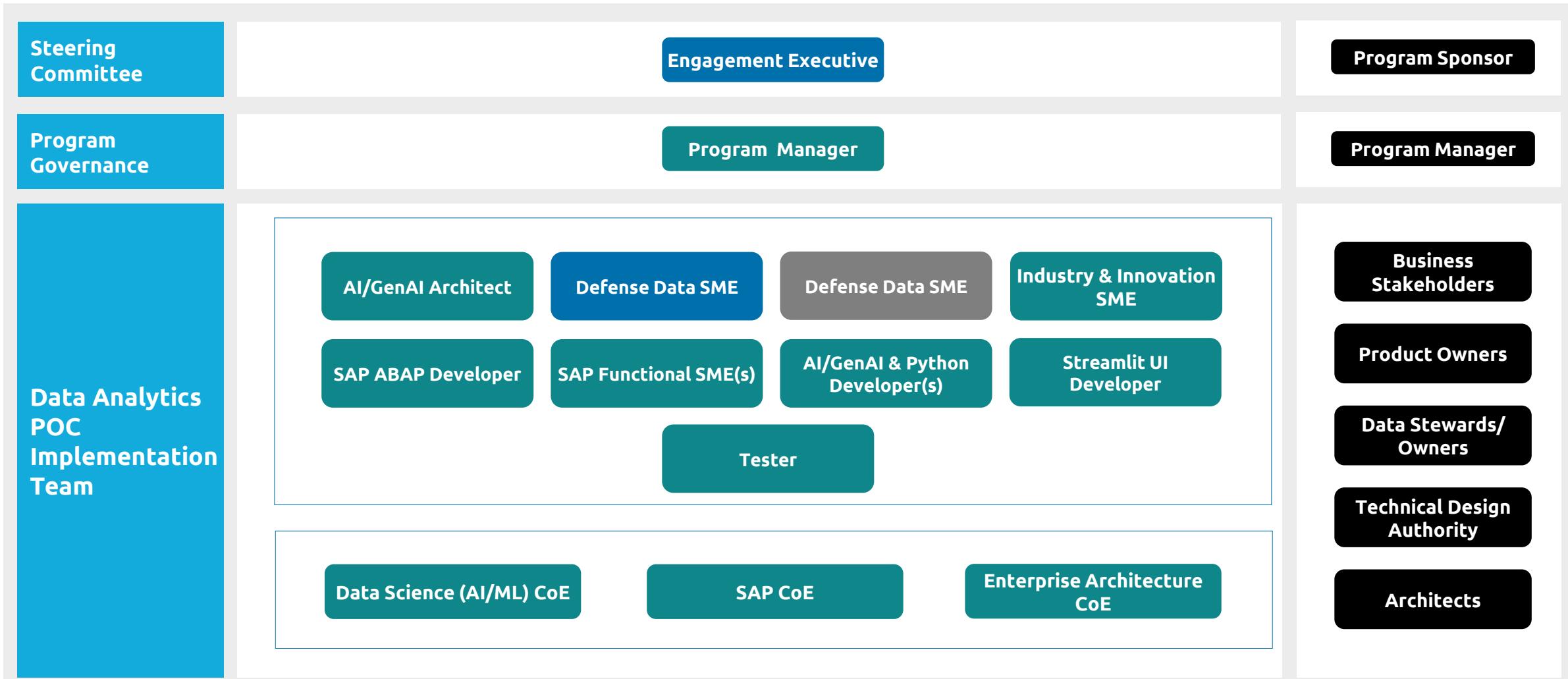
# Team Structure

Capgemini US

Capgemini India

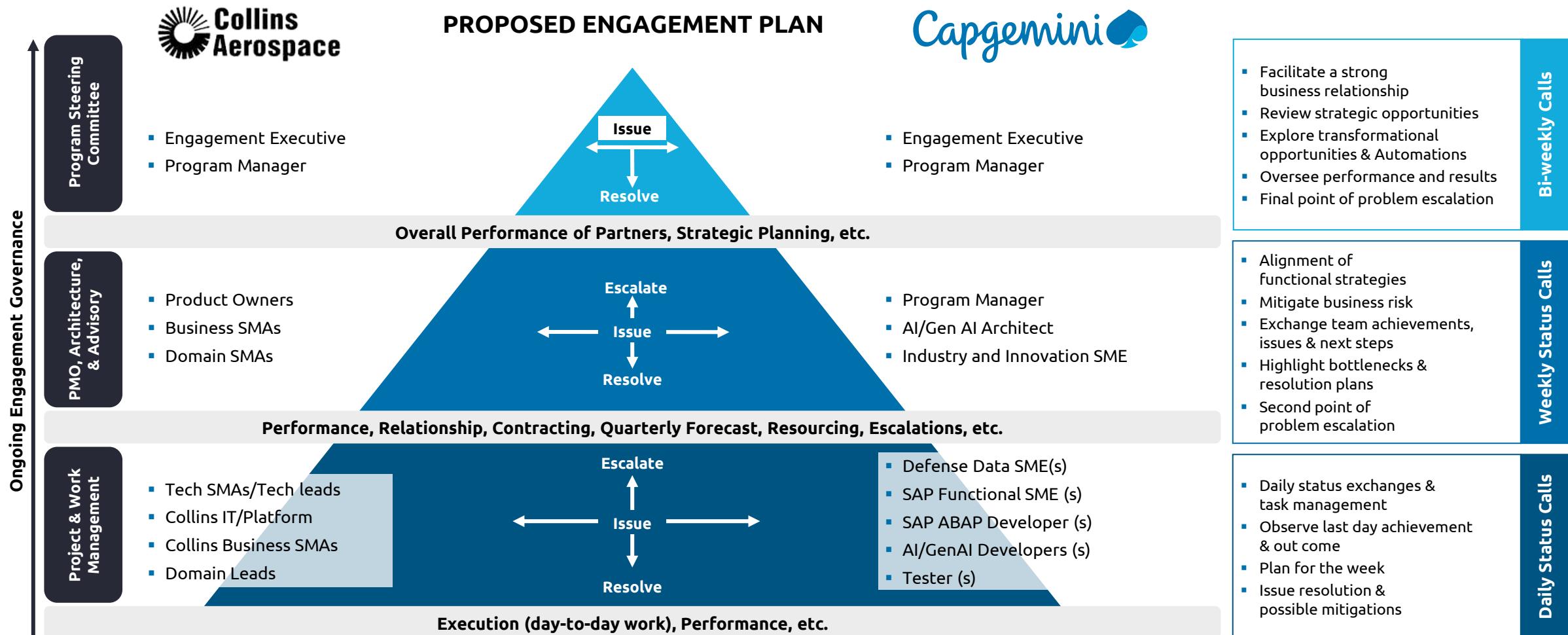
Capgemini Germany

FirmX





# We propose a collaborative, transparent governance model to achieve faster, and sustainable outcomes



*Strong, Transparent Governance is critical to measure and communicate program success and challenges to stakeholders*



# Risks & Mitigations

Sr. no.	Key Risks	Impact	Mitigation Plan
1	Incomplete metadata or unclear column descriptions	High	Engage SAP functional SMEs early
2	LLM misclassification due to ambiguous fields	High	Use few-shot prompting and confidence scoring
3	Delays in manual validation	Medium	Allocate SMEs per domain and plan phased reviews
4	Export Control tagging ambiguity	High	Request access to ITAR questionnaire and examples
5	Incomplete or inaccurate business glossary	High	Conduct multiple SME review cycles and validation sessions for the business glossary.
6	Delay in SME validation and feedback	High	Define clear timelines and responsibilities for SMEs; schedule regular follow-ups.
7	Non-compliance with ITAR and PII	High	Engage compliance SMEs early; do the compliance repository reviews during the POC.
8	Inadequate synthetic data realism	Medium	Use advanced GenAI techniques and validate synthetic data with SMEs iteratively.
9	Model performance issues in classification engine	Medium	Test multiple approaches (LLM, AI/NLP, Rule based) and select best fit; incorporate feedback loops for continuous improvement.
10	Resistance to change from end-users	Medium	Conduct change management workshops; highlight benefits and ease of use of the new system.
11	Data privacy concerns with synthetic data	High	For the POC, synthetic data will be used to generate SAP's 50 tables—avoiding production data—while robust masking aligns with the overall solution; in production, all anonymization will be validated by legal and compliance teams for full adherence.
12	Resource constraints from Collins	High	Secure commitment for dedicated SMEs and resources upfront for query and Data validation.

# RACI Matrix: Defining Accountability Across Project Phases

		R: Responsible	C: Consulted	A: Accountable	I: Informed
Phase	Activity	Collins	Capgemini		
Project Initiation	Define project scope and objectives	I		R	
	Provide SMEs for business glossary and Data validation	R		C	
	Set up project governance and timelines	I		R, A	
Environment Setup	Provide access to SAP systems and data	I, C		R, A	
	Configure Capgemini environment and Git	I		R, A	
	Procure and set up required tools/licenses	C		R, A	
Data Extraction & Preparation	Extract SAP metadata	C		R, A	
	Review and validate business glossary	R, A		C, I	
	Signoff business glossary	R, A		C, I	
Classification Rules & Compliance	Define and validate classification rules	R, A		C, I	
	Ensure compliance with ITAR and PII and provide Sign Off	R, A		C, I	
Synthetic Data Generation	Generate synthetic data for 50 SAP tables	C, I		R, A	
	Validate synthetic data Classification and integrity as per Public Domain Rules and Sign Off	R, A		C, I	
Classification Engine Development	Develop classification engine (AL/ML and GenAI + Rules)	C, I		R, A	
	Test and validate classification accuracy	R		C	
UI Development (Streamlit)	Develop and deploy Streamlit UI on Capgemini VM	C, I		R, A	
SME Validation & Feedback	Conduct SME validation via CSV download via UI	R, A		C, I	
	Incorporate feedback and retrain models	C, I		R, A	
Testing & Validation	Conduct Unit Testing	C, I		R, A	
	Validate compliance and data quality	R, A		C, I	
Post-POC Transition	Define ownership for full-scale deployment	R, A		C, I	
	Forward Planning Dialogue	R, A		R, A	



# 05

## Commercials



# Commercials



Scope	Total Cost(USD)
<ul style="list-style-type: none"> <li>Define POC architecture for data classification, and synthetic data generation approach.</li> <li>Detail migration plan for data extraction, classification, and synthetic data creation for 50 SAP tables across 3 domains: <b>MM</b> (Material Management), <b>SD</b> (Sales &amp; Distribution), <b>PP</b> (Production Planning) Align with compliance requirements (ITAR, and PII)</li> </ul>	\$ 206,000

## Key Highlights:

- This POC is for 50 SAP tables list for SD, MM, PP provided by Collins
- The above Cost excludes schedule slippage due to the delays attributed to changes linked to items called out in Risks Sections / Dependencies sections or scope additions which are beyond the control of Capgemini or Collins. Additional Cost will be discussed and will be mutually agreed by both Capgemini & Collins.
- Does not include any Taxes or levies
- There is \$ value added for:
  - VM Cost is 70 USD in 3 users/ month i.e., 630 USD
  - GIT Cost is 17 USD/ user/ month i.e., 510 USD
- SOW to be signed between the America or India Entity to be mutually discussed and agreed
- Proposal validity is till 30<sup>th</sup> September 2025.

# 06

## Acceptance Criteria





# Acceptance Criteria

No	Acceptance Criteria
1	Metadata for 50 MM/SD/PP tables is extracted successfully with >95% completeness..
2	Fields are classified with =>90% accuracy for ITAR and PII labels. (based on public domain knowledge of domain expertise) . Final Validation of the classification will be done by Collins US team and share the success measurement.
3	A simple Streamlit interface will allow SMEs to download classification results as a CSV file, review and update it with feedback, and upload the revised file to support model retraining.
4	Synthetic datasets maintain referential integrity and fidelity to real life like data, validated by SMEs. There will be at least 2 Material lifecycle within sample data.
5	Classifications include confidence scores or explanations (AI/NLP or rules or LLM-based).
6	Timely SME feedback within 7 Business days is captured and used to refine the model within 2 iterations.
7	Solution is deployed on dedicated VM on premise using Streamlit App.
8	Solution complies with ITAR & PII which are known in the public domain
9	Business glossary, data owners, and classification rules are documented and approved by SMEs within 3 business days.

*No additional accepted criteria is implied or derived.*





# 07

## Case Studies



# GenAI - Synthetic Data to Train Models for Identifying Hazardous Obstacles on Railway Tracks



INTELLIGENT INDUSTRY

## CLIENT CONTEXT:

- Customer is a major player in the rail industry and a leading provider of transport services.

## CHALLENGES/PAIN POINTS:

- Inability to have representative images in occurrence or in sufficient number, especially in situations where operators are crossing railways or when specific risky situations occurs

## GAIN POINTS:

- Improve the performance of detection systems that helps taking decision for the trains for avoiding obstacles

## APPROACH & SOLUTION:

- Based on GenAI, built an image generator to create diversity of phenomena, ultimately improving the robustness and reliability of detection models.
- Our team of Data Scientists has designed and developed a system for patching / masking images to extract a particular phenomenon. We have also designed a system that allows these patches to be pasted onto other new images, making them realistic so as not to bias the final detection. Thanks to this system, a new representative dataset can be generated, safely and with automatic annotations.

**Improved Detection**  
with new realistic images

**Enhance Decision Making**  
to avoid obstacles

**Reduce Risk**  
with a Stronger Detection Mechanism

BENEFITS DELIVERED

# GenAI Synthetic Data Creation for Training AI Models on GPRD-protected Medical Data



## CLIENT CONTEXT:

- Capgemini helped a Swedish Welfare Agentur protect patient data to comply with the General Data Protection Regulation.

## CHALLENGES/PAIN POINTS:

- Limited data access: Due to privacy and compliance regulations, particularly GDPR, access to patient data is restricted, hindering the amount of data available for AI applications
- Data scarcity: Scarcity of data impedes the development and performance of AI models in healthcare, leading to suboptimal outcomes

## GAIN POINTS:

- Accelerate testing and quality assurance (QA) using advanced deep learning techniques
- Ensure GDPR compliance
- Offer valuable insights into patient situations

## APPROACH & SOLUTION:

- Capgemini developed a solution, utilizing Generative Adversarial Networks, AI, and analytics alongside Capgemini's Artificial Data Amplifier to create and validate synthetic datasets
- Combination of deep learning methods, a sample of the real data is fed into the model
- Output of the model is a generated synthetic datasets, very similar to the original data in terms of statistical similarity and distribution

BENEFITS DELIVERED

**Accelerates**  
QA with ample production data

**Generates**  
medical data for patient insights

**Ensures GDPR Compliance**  
via data synthesis



# 08

## Capgemini Capabilities



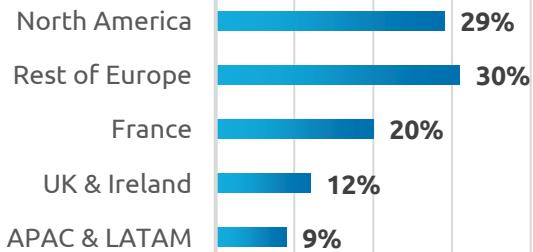
# Capgemini – A Leader for Leaders ...

Capgemini group is one of the world's largest management and IT consulting companies

## GROUP REVENUE

Group Revenue for FY2024:

**\$23, 903 Mn**



## ONE TEAM

Present in more than

**50 countries**

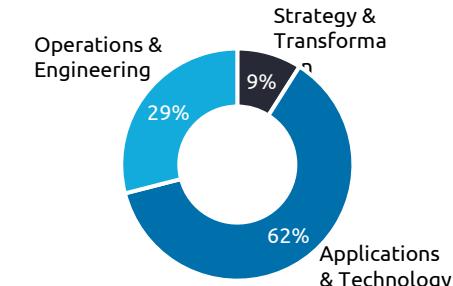
and

**160+ nationalities**

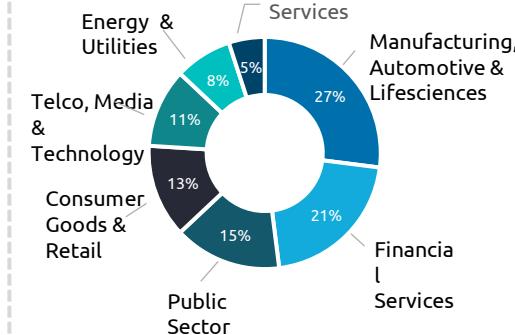


**340,000** people worldwide working together as **one team**

## OUR SERVICES



## SECTOR FOCUS



## BUSINESS SUB-BRANDS

**Capgemini invent**



**frog** Part of Capgemini Invent

**Capgemini engineering**

**sogeti**  
Part of Capgemini

[See other sub-brands](#)

## THE BRAND PROMISE TO ITS PEOPLE

**GET THE FUTURE YOU WANT**



## THOUGHT LEADERSHIP AND INNOVATION

[Capgemini Research Institute](#)



Ranked #1 consecutively for the 5<sup>th</sup> time for the quality of its research.

**Capgemini**  
APPLIED INNOVATION EXCHANGE

**Capgemini Ventures**

**Business Ventures**

Leveraging a global [AIE network](#) & framework for action to help clients test innovation and turn it into business value.

Investing and partnering with startups to co-innovate on the market.

Partnering with technology and industry leaders on new joint ventures and companies that provide cutting-edge solutions and approaches to solve today's biggest challenges.

**We are Architects of Positive Futures**

Diversity & Inclusion | Digital Inclusion | Environmental Sustainability

# Capgemini's Insights & Data Footprint



**3500 + client**  
engagements  
across 50  
countries

**RAISE:** Development & deployment solution to industrialize custom Gen AI projects

**Robust portfolio:** Data engineering, AI Analytics & Data Science, Generative AI

**Industry know-how** and world-class solutions across all major sectors

## Industries We Serve

- Consumer Products & Retail
- Manufacturing & Life Sciences
- Financial Services
- Energy Transition & Utilities
- Telecom, Media & Technology
- Automotive
- Public Sector and Services

## Top Analyst Recognition



▪ **NelsonHall's NEAT Evaluation** in AI & Analytics area of Core Banking Services 2025



▪ **HFS Horizons – Generative Enterprise Services 2025**



▪ **AVASANT – Applied AI Services 2024-2025 RadarView**



▪ **IDC Market Scape: Worldwide Data Modernization Services 2024 Vendor Assessment**



▪ **ISG Provider Lens: Advanced Analytics & AI Services 2024**



▪ **Everest Group: Data and Analytics (D&A) Services PEAK Matrix® Assessment 2024**



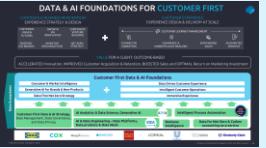
▪ **Forrester Wave™: AI Service Providers 2024**



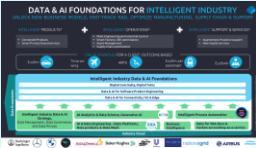
# AI Analytics & Data Science Core Offer

## Delivering Trusted Outcomes At Scale

### OUR PLAYING FIELDS



**Customer  
First**



**Intelligent  
Industry**



**Enterprise  
Management**

### GENERATIVE AI OFFERS



**Generative  
AI Strategy**



**Gen AI for  
SDLC**

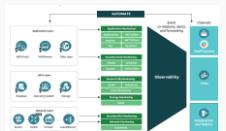


**Gen AI for  
CX**



**Gen AI for  
Enterprise**

### FLAGSHIP ASSETS/SOLUTIONS



**AIOps:** It brings a unified, open, and dynamic **suite of intelligent automation tools leveraging AI/Gen AI**, serving businesses as a continuously evolving source of innovation and value.



**AI GlassBox:** Enables comprehensive framework for **Model Life Cycle Management**. AI Glass Box performs 4CP framework for any set of use cases



**RAISE:** An accelerator for **developing and deploying in production your Custom Generative AI projects**, at scale, in a cost-controlled and trusted manner

### COMPETENCY HIGHLIGHTS

**200+**

**Reusable Industry Solutions** in the AI & Analytics Marketplace

**22**

**Applied Innovation Exchange (AIE)**, helping clients test innovation and turn it into business value

**30+**

**Solution Accelerators** for Data Science, Machine Learning, AI, and Gen AI.

**8+**

**Strategic Partners** to deliver Generative AI Solutions

**22**

**Applied Innovation Exchange (AIE)**, helping clients test innovation and turn it into business value

### OUR STRENGTH

- 6000+** Analytics & Data Science Professionals across the Globe
- "Leader"** – The Forrester Wave: AI Services, Q2 2024
- "Leader and a Star Performer"** – Everest Group's AI and Gen AI Services PEAK Matrix® Assessment 2024

### STRATEGIC PARTNERSHIPS



## About Capgemini

Capgemini is a global business and technology transformation partner, helping organizations to accelerate their dual transition to a digital and sustainable world, while creating tangible impact for enterprises and society. It is a responsible and diverse group of 340,000 team members in more than 50 countries. With its strong over 55-year heritage, Capgemini is trusted by its clients to unlock the value of technology to address the entire breadth of their business needs. It delivers end-to-end services and solutions leveraging strengths from strategy and design to engineering, all fueled by its market leading capabilities in AI, generative AI, cloud and data, combined with its deep industry expertise and partner ecosystem. The Group reported 2024 global revenues of €22.1 billion.

Get the future you want | [www.capgemini.com](http://www.capgemini.com)



This presentation contains information that may be privileged or confidential and is the property of the Capgemini Group.

Copyright © 2025 Capgemini. All rights reserved.