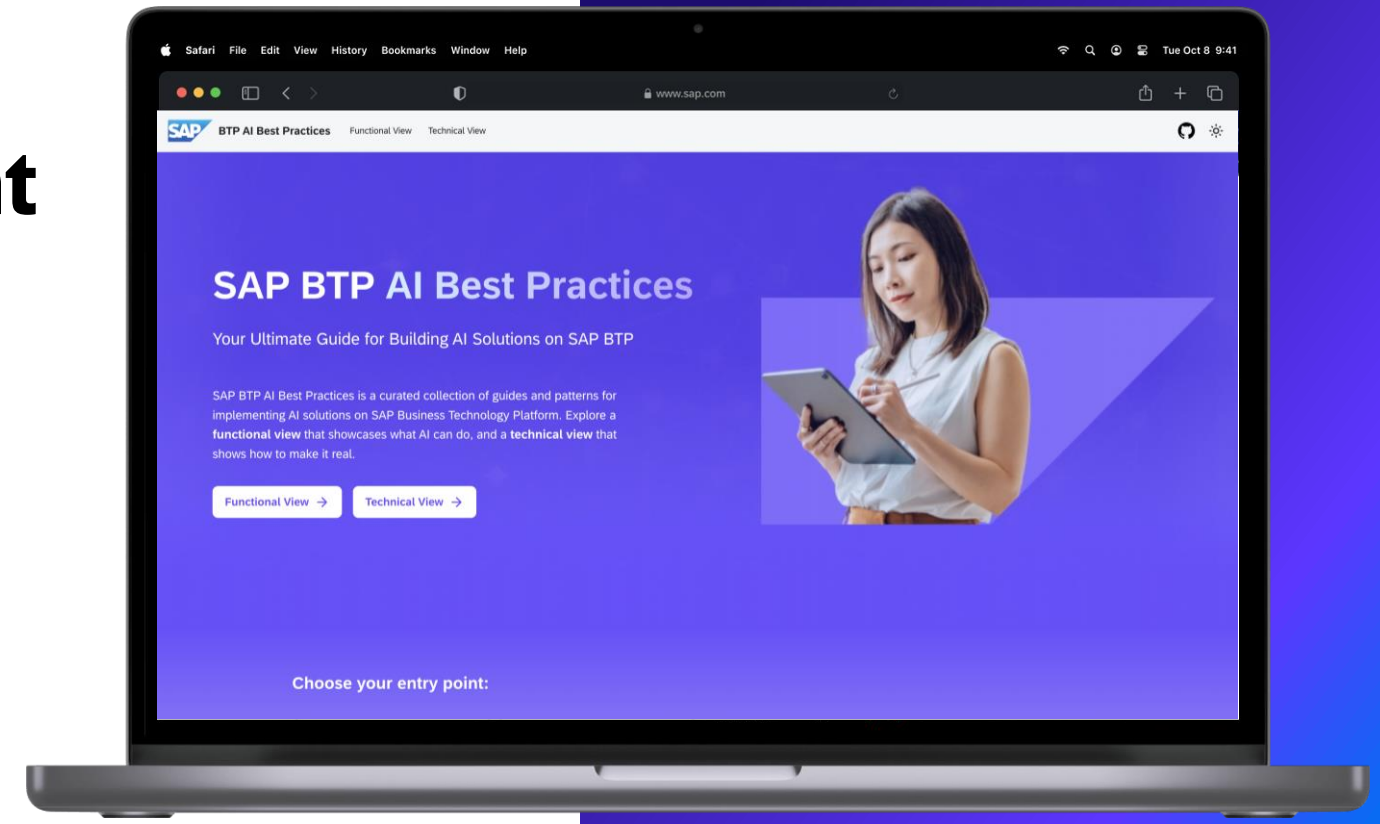
SAP BTP AI Best Practices

Training and Deployment of Custom AI Models

Deploy Custom models in SAP AI Core



BTP AI Services Center of Excellence

29.09.2025

Steps

1 Overview

2 Pre-requisites

3 Key Choices and Guidelines

4 Implementation

Training and Deploying Custom Predictive Models

Building your custom predictive models within SAP BTP is designed to augment the company's existing capabilities by introducing tailored AI solutions. This document explores best practices for training and deploying custom models, focusing on integrating them seamlessly into existing business processes.

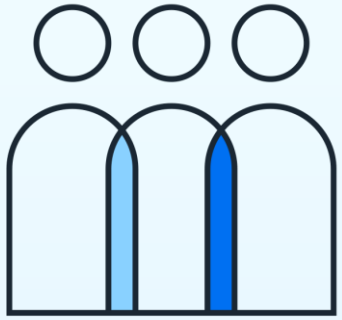
Building a custom predictive model involves an initial assessment of existing solutions in the company's SAP Business AI portfolio. Before you go for a custom predictive model development, it is recommended to review the available solutions in the SAP Business AI portfolio to find out-of-the-box offerings that suit your business needs. If you haven't found a suitable solution, then it's time to consider the development of Custom AI Models on the SAP BTP platform.

Expected Outcome

Effectively build and integrate Custom AI model into existing business solution
Optimize AI operations and Scale AI Applications

Key Benefits

Why build and deploy Custom Predictive Models on AI Core

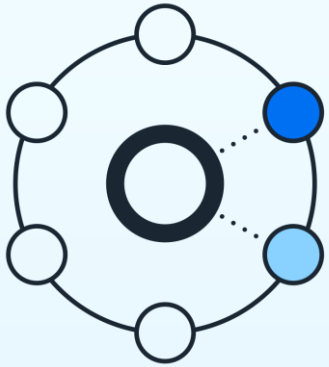


Improved Operational Efficiency and Data-Driven Decision Making

Custom AI Models help making in Data Driven Decision to improve operation efficiency in Business Processes

AI@Scale with Seamless Integration

Scale your AI Apps on AI Core with integration with SAP Apps, SAP Data sources and APIs.



Personalization at Scale

AI models enables you hyper-personalized experiences for your business users

Key Concepts

- **Narrow AI (Domain-Specific AI):**

  A subset of Artificial Intelligence that focuses on solving specific business problems (e.g., demand forecasting or payment default prediction) using historical data

- **Model Lifecycle Management**

  Involves stages like data acquisition, preprocessing, feature engineering, model training, validation, deployment, and monitoring

- **Explainability and Trust**

  Responsible model development should offer interpretable insights (via SHAP, LIME, etc.) to ensure business stakeholders can trust the predictions.

- **Real-Time and Batch Inference:**

  Real-time inference refers to immediate prediction based on the input while Batch inference takes a batch of input and perform the inference together.

Example Use Cases

- Payment Default Prediction
- Sales Forecasting and Deal Conversion
- Custom Demand Forecasting by Region or Channel
- Supplier Lead Time Risk Prediction
- Predictive Maintenance for Custom Assets
- Employee Exit Risk Modelling (Custom Factors)
- Custom Product Return Prediction
- Churn Risk for Subscription-Based Services
- Warehouse Inventory Risk Modeling

Pre-requisites

Business

- SAP Business Technology Platform subaccount
- Subscription/Instance of below services-
  - SAP AI Core
  - SAP AI Launchpad
  - SAP HANA Cloud (Optional)

Technical

- SAP Business Technology Platform subaccount (Setup Guide)
- Create an instance of SAP AI Core (Setup Guide)
- Subscribe to SAP AI Launchpad (Setup Guide)
- Create an instance of SAP HANA Cloud (Setup Guide)

SAP Business Technology Platform (SAP BTP)

- SAP Business Technology Platform (BTP) is an integrated suite of cloud services, databases, AI, and development tools that enable businesses to build, extend, and integrate SAP and non-SAP applications efficiently.

SAP AI Core

- SAP AI Core is a managed AI runtime that enables scalable execution of AI models and pipelines, integrating seamlessly with SAP applications and data on SAP BTP that supports full lifecycle management of AI scenarios.

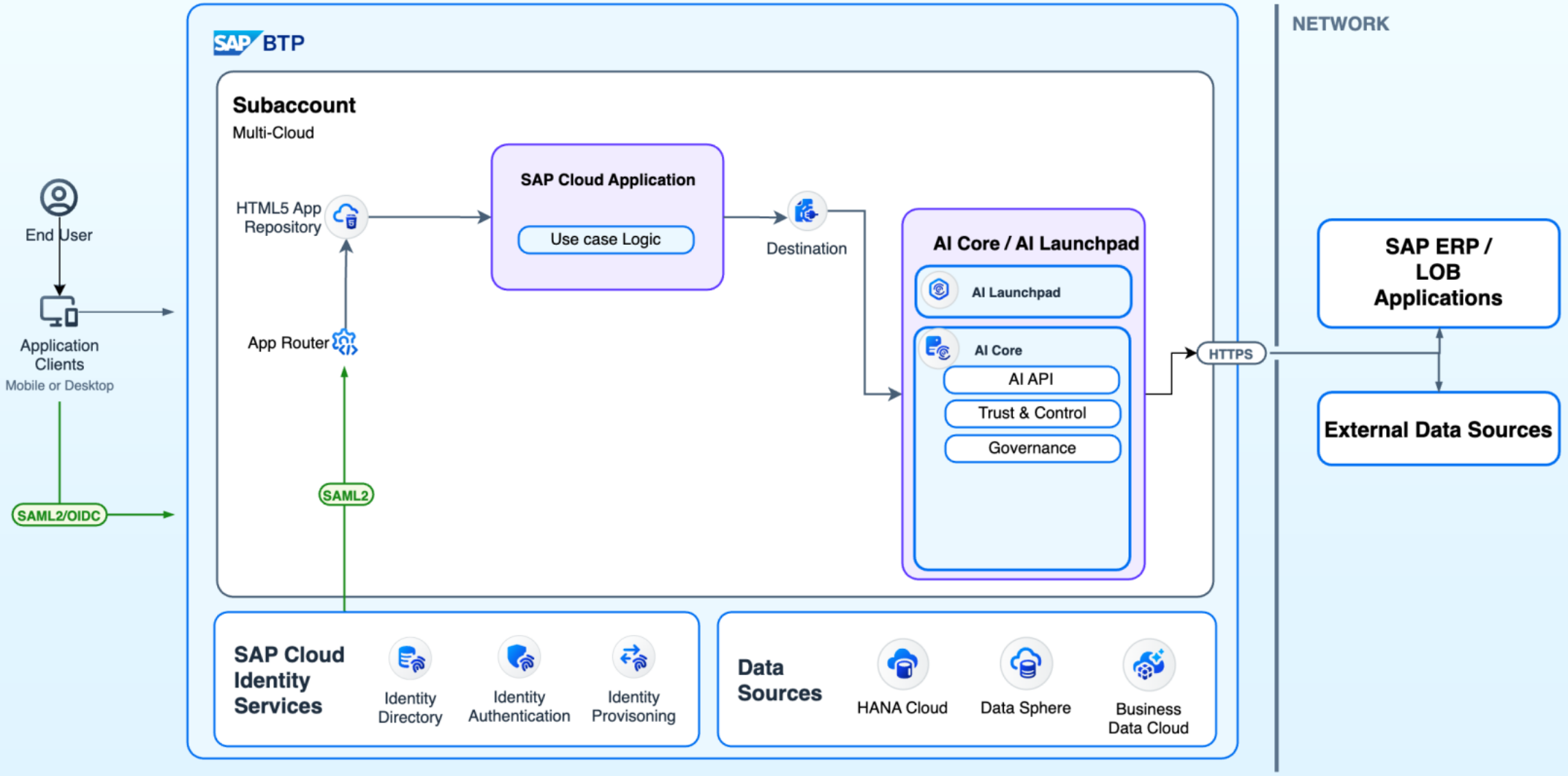
SAP AI Launchpad

- SAP AI Launchpad is a multitenant SaaS application in SAP BTP. Customers can use SAP AI Launchpad to manage AI use cases (scenarios) across multiple instances of AI runtimes.

SAP HANA Cloud

- SAP HANA Cloud is a database as a service that powers mission-critical applications and real-time analytics with one solution at petabyte scale. Use relational, property graph, spatial, vector, and semi-structured data along with embedded machine learning to power intelligent data applications.

High Level Technical Architecture



Key Choices and Guidelines

Problem Definition & Business Alignment

- Clearly define the ML problem and ensure it aligns with business goals to drive measurable value.
- Engage stakeholders early for domain-specific insights and success criteria.

Data Strategy

- Collect high-quality, representative data while ensuring ethical sourcing and regulatory compliance.
- Preprocess and engineer features to maximize model effectiveness and fairness.
- Try to push the AI/ML model development close to data source helps to reduce data movement and latency. It is recommended to consider HANA based data pre-processing and ML algorithms if your data is already in SAP HANA.

Model Development

- Select appropriate algorithms based on problem complexity, interpretability, and available data.
  Start with simpler algorithms or smaller models and then increase the complexity as needed, supporting your business case.
- If you are using deep-learning for your use case, consider pretrained models for faster development cycles.

Model Evaluation & Validation

- Validate models using robust metrics, cross-validation, and edge-case testing.
- These evaluation and validation will help you build robust and reliable models.

Key Choices and Guidelines

Choosing the right resource plan

- Choosing the right resource plan helps to speed up your training cycle, reduce latency on inference yet being cost effective. You may use below guideline to choose the best resource plan for your use case.
  - Small training resource plans that are suitable to train your classic regressors, trees, XGB, and simple neural nets.
  - Small inference resource plans to serve your classic regressors, trees, XGB, and simple neural nets.
  - Larger inference resource plans to serve LLMs, vision models, and deeper/larger neural nets.
  - Note: Large training resource plans allow for distributed training/fine-tuning of larger models like LLMs, vision models, and deeper/larger neural nets, but they may be available only for SAP internal purposes.

Deployment & MLOps

- Use MLOps tools for reproducible deployment, scalability, and automated monitoring.
- Ensure MLOps pipeline or CI/CD pipelines are in place for efficient and secure model releases.
- Containerised (docker based) is a very good choice for deployments to scale you AI applications.

Key Choices and Guidelines

Monitoring & Retraining

- Continuously monitor model performance, data drift, and user feedback.
- Set up retraining workflows to maintain accuracy and relevance over time.

Security & Privacy

- Safeguard data and models through encryption, access control, and protection against adversarial attacks.
- Ensure privacy using anonymization, federated learning, or confidential computing where applicable.

AI Ethics & Responsible AI

- Safeguard Embed fairness, transparency, and accountability throughout the model lifecycle.
- It is advisable to use explainability tools and documentation (like model cards) to support ethical governance using anonymization, federated learning, or confidential computing where applicable.

Implementation

Use BTP Services to Deploy your Customer AI Models

Using AI Launchpad

Tools

- SAP AI Launchpad

Learning Journeys-

Below steps in sequence will helps you deploy your modes in SAP AI Core.

- Landscape Setup (setting up AI Core /AI Launchpad) : (one time activity)
- Setup for Git Repositories - (one time activity)
- Setup you environment for container ready development
- Build your AI/ML Model in a container ready environment :
  - Build a House Price Predictor Model
  - Yoga Pose Image Classification : Use training and inference pipeline
  - Utilizing GPUs in SAP AI Core
  - Ingest data into your AI model with SAP AI Core
- Generate Metrics and Compare Models in SAP AI Core.
- Make Predictions for your AI model with SAP AI Core

Using CLI in VS Code

Tools

- SAP AI Core Toolkit for VS Code

Learning Journeys

- SAP AI Core using VS Code Toolkit

Contributors

Gupta, Chirag

Behera, Bhagabat Prasad

Thank you