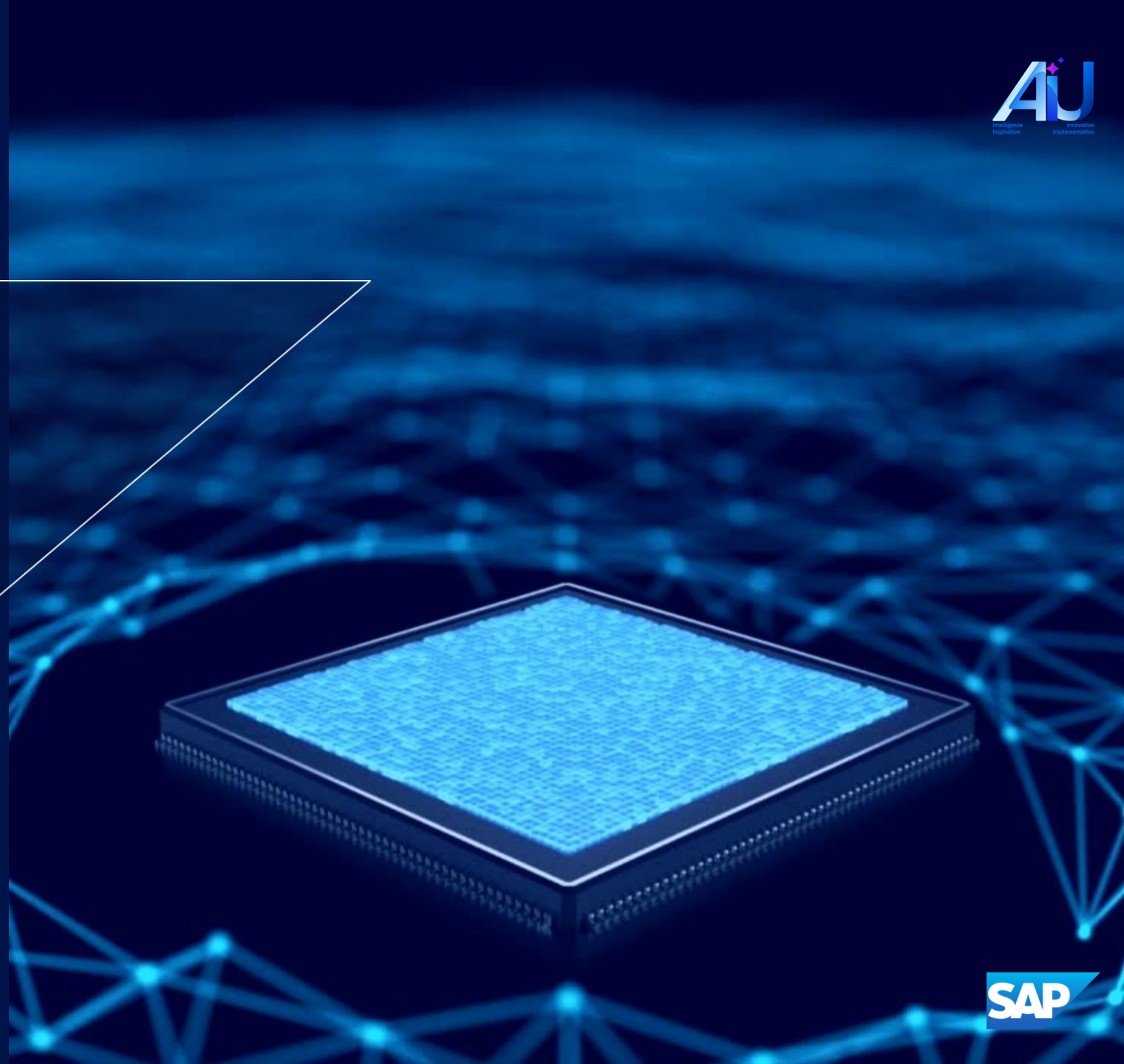
April 14, 2026

SAP BTP AI Best Practices

From Strategy to Implementation

Marco Cigaina, BTP AI Services CoE
Luis Marques, BTP AI Services CoE



Disclaimer

The information in this presentation is confidential and proprietary to SAP and may not be disclosed without the permission of SAP. Except for your obligation to protect confidential information, this presentation is not subject to your license agreement or any other service or subscription agreement with SAP. SAP has no obligation to pursue any course of business outlined in this presentation or any related document, or to develop or release any functionality mentioned therein.

This presentation, or any related document and SAP's strategy and possible future developments, products and or platforms directions and functionality are all subject to change and may be changed by SAP at any time for any reason without notice. The information in this presentation is not a commitment, promise or legal obligation to deliver any material, code or functionality. This presentation is provided without a warranty of any kind, either express or implied, including but not limited to, the implied warranties of merchantability, fitness for a particular purpose, or non-infringement. This presentation is for informational purposes and may not be incorporated into a contract. SAP assumes no responsibility for errors or omissions in this presentation, except if such damages were caused by SAP’s intentional or gross negligence.

All forward-looking statements are subject to various risks and uncertainties that could cause actual results to differ materially from expectations. Readers are cautioned not to place undue reliance on these forward-looking statements, which speak only as of their dates, and they should not be relied upon in making purchasing decisions.

Why do we need BTP AI Best Practices?

Rapid External AI Advancements

The **AI landscape** is evolving at an unprecedented pace. Staying current requires adaptable approaches and best practices.

Dynamic BTP AI Stack

SAP BTP AI stack is rapidly evolving. Best practices ensure we leverage new features effectively.

Information Fragmentation

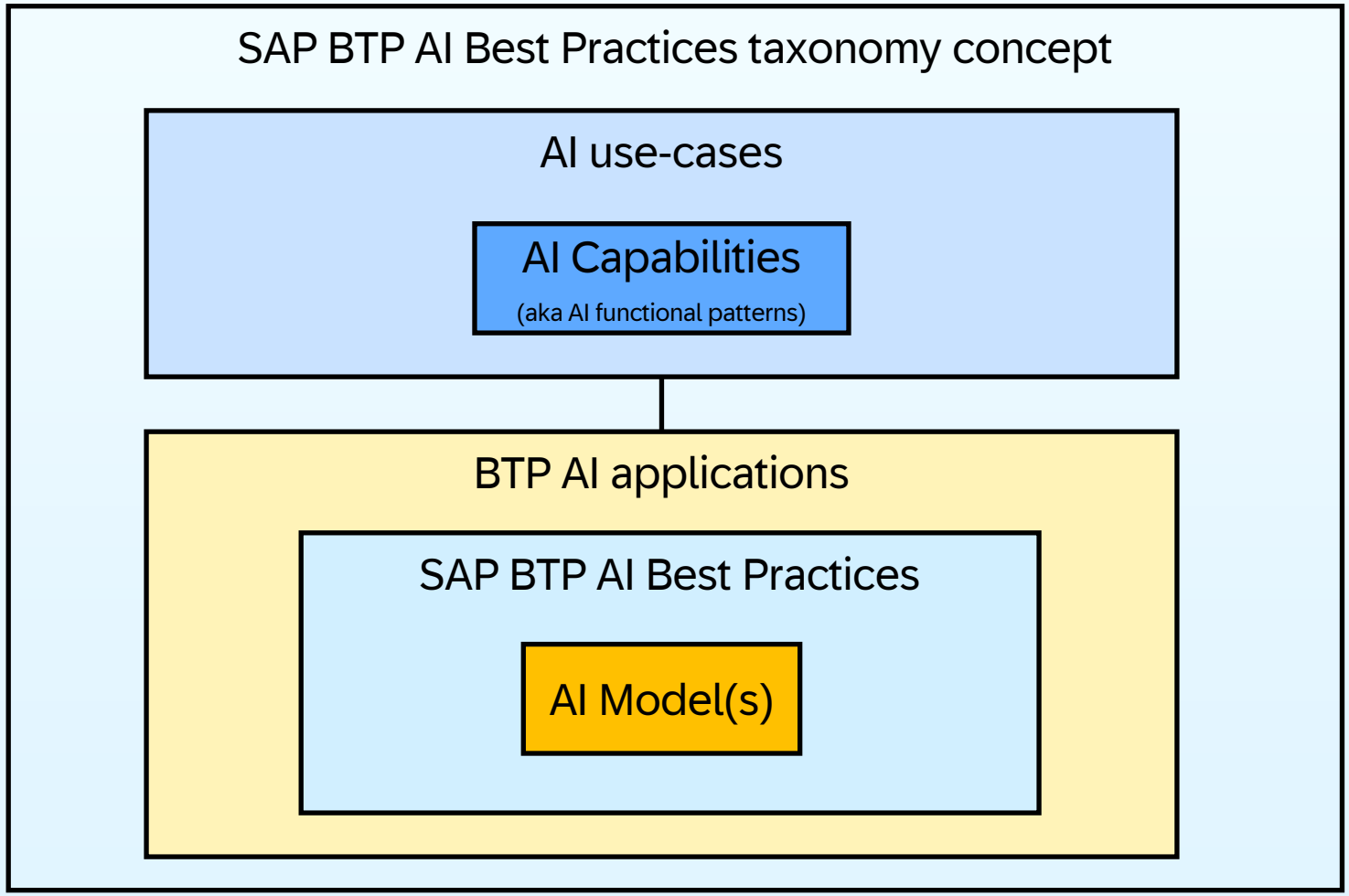
AI **knowledge** is spread across various sources. Consolidated best practices provide a single, reliable point of reference.

What are BTP AI Best Practices?

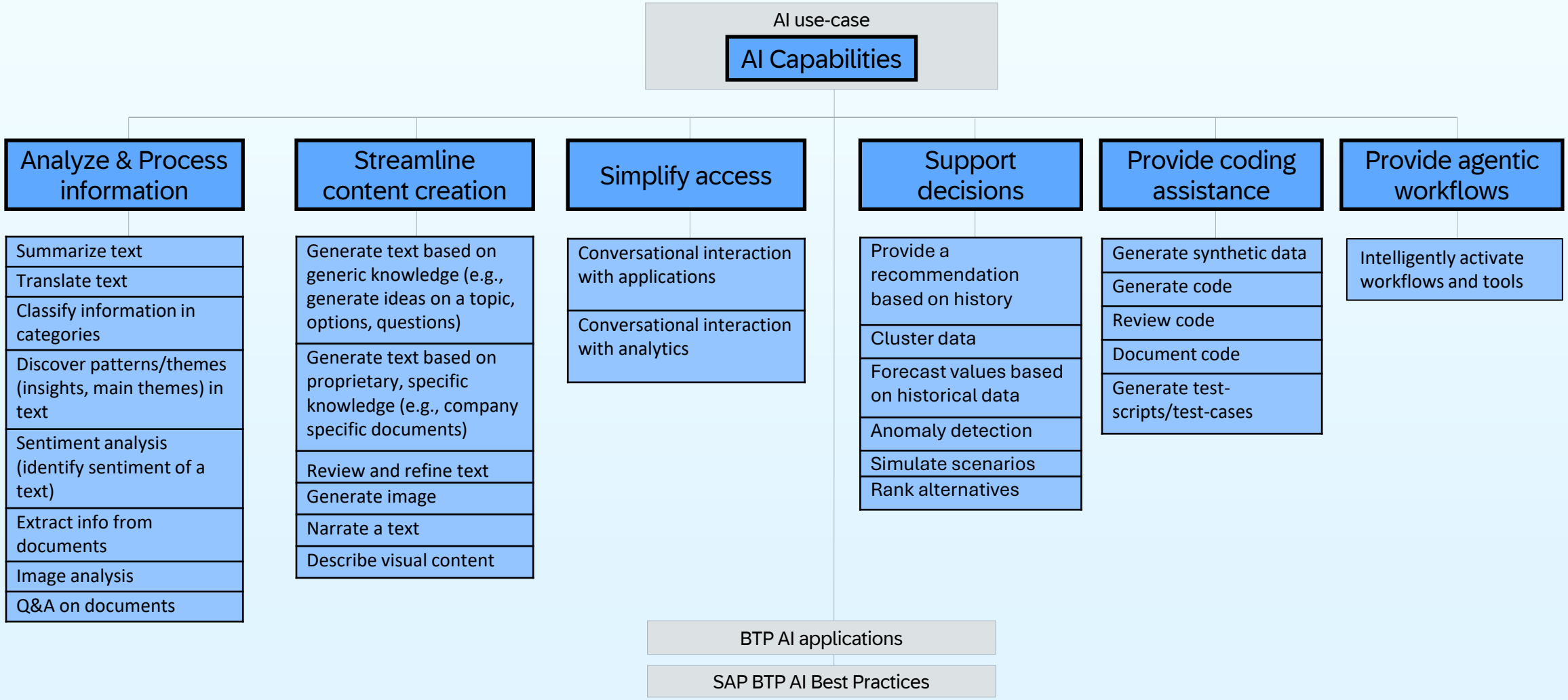
- Best Practices are a well-organized, consistent and comprehensive set of proven **development guidelines** (including all key design and implementation choices) to develop AI-powered BTP-based applications

- **Examples** of specific Best Practices: size of chunks for vector-based Retrieval Augmented Generation, data masking in LLM access, way to deploy custom AI models, etc.

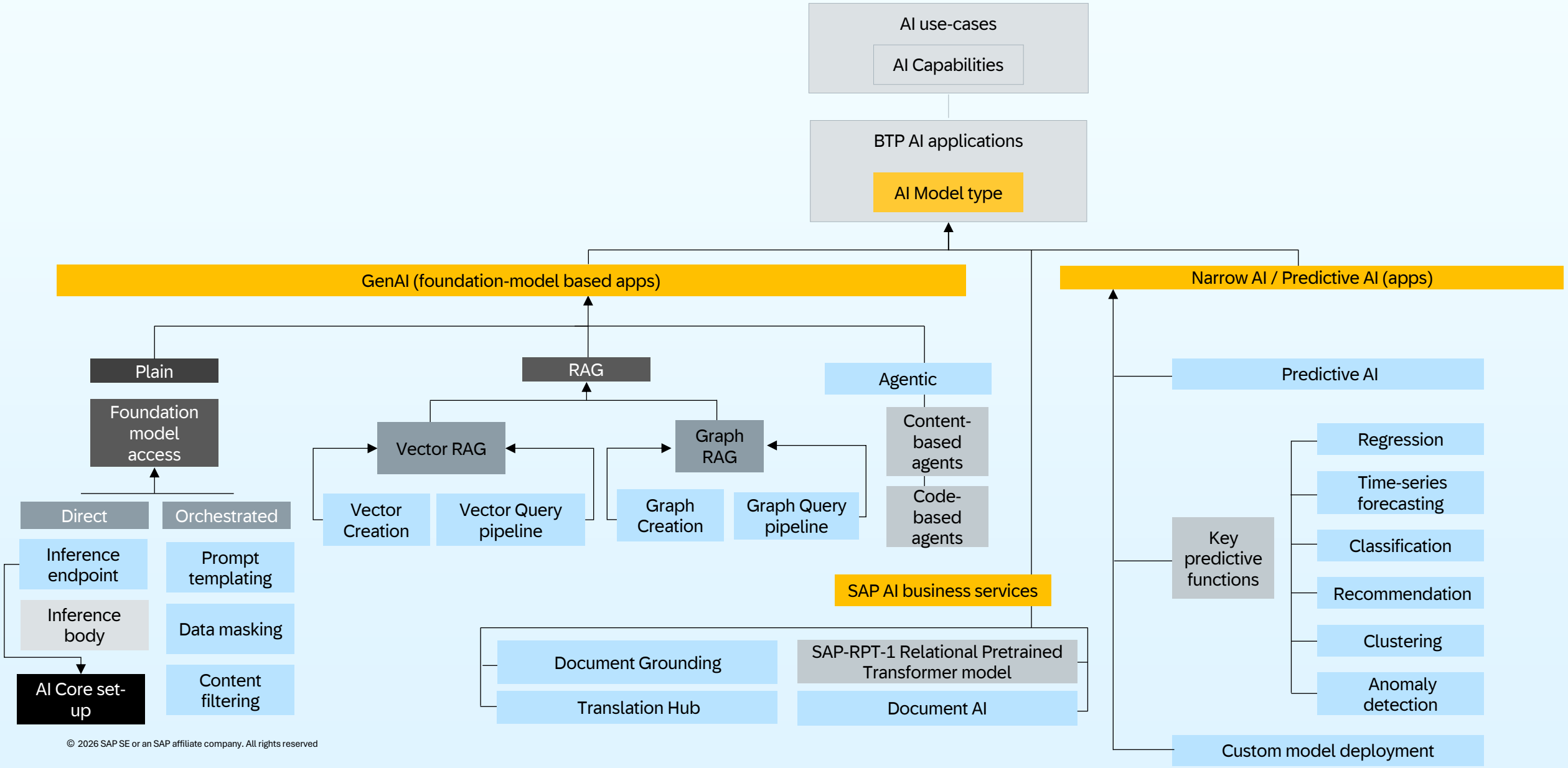
SAP BTP AI Best Practices taxonomy concept



SAP BTP AI Best Practices taxonomy concept



SAP BTP AI development Best Practices



Released BTP AI Best Practices

1. Access to Generative AI Models
2. Prompt Templating
3. Data Masking
4. Content Filtering
5. Vector-based Retrieval Augmented Generation – embedding
6. Vector-based Retrieval Augmented Generation – query pipeline
7. Knowledge Graph based Retrieval Augmented Generation – creation
8. Knowledge Graph based Retrieval Augmented Generation – query
9. Regression
10. Time-series forecasting
11. Anomaly detection
12. Classification
13. Clustering
14. Training and Deployment of Custom AI Models
15. Document AI
16. Custom Joule Skills
17. Document Grounding
18. SAP Translation Hub

https://btp-ai-bp.docs.sap

Access to Generative AI Models

Simple Consumption of Generative AI Models

Interacting with deployed Generative AI models involves **sending requests** to **retrieve generated responses**.

In the request, we can define the **prompt**, the **model**, and the **parameters** that control the response generation.

Expected Outcome

- Provide secure and efficient **access** to Generative AI models.
- Enables applications to **leverage** the power of **AI models** and provide a wide spectrum of functionalities.

Prompt templating

Simple Prompt Reusability for Generative AI Models

Method used to create **structured** and **reusable** prompts for **Large Language Models**

This approach allows developers to **program**, **store**, and **reuse** prompts efficiently across different tasks and applications

Expected Outcome

- Create **reusable templates** for specific tasks
- Ensure **performance, reliability and efficiency** of LLM responses.

Data Masking

Protect sensitive information

Data masking is a crucial technique used to **protect** sensitive information when interacting with Large Language Models.

It involves **replacing** identifiable or confidential data within prompts with placeholder text, ensuring that such information is **not exposed** to third-party models.

Expected Outcome

- The expected outcome after implementing data masking is to ensure that **sensitive information is protected** and not exposed during data processing and analysis.
- This helps in **maintaining data privacy and security**, which is crucial for **compliance with various regulations and standards**.

Content Filtering

Filter the input and output based on content safety criteria

Content filtering in SAP AI Core is designed to help ensure the **safety and appropriateness** of both the **input** sent to AI models and the **output** generated by them.

SAP AI Core's content filtering provides a mechanism to add a layer of **content moderation** to your AI applications by leveraging established content safety services

Expected Outcome

- By implementing content filters, you can ensure that the generative AI models in SAP AI Core produce **safe and appropriate content**, adhering to the specified content safety criteria thus **ensuring compliance and alignment** is achieved in generated text.

Vector-based Retrieval Augmented Generation Embedding

Make unstructured information ready to be queried by LLM-based applications

Store data in **high-dimensional vector** formats, capturing the semantic **meaning** of the content.

Embedding allows for **efficient similarity searches** that help **extending the context** of large language models.

Expected Outcome

Unstructured information **ready to be queried** by LLM-based applications

Vector-based Retrieval Augmented Generation Query Pipeline

Effectively improve LLM responses with augmented context

Retrieve

Fetch relevant Info from knowledge base

Augment

Provide additional Info-chunks to enrich context to LLM

Generate

Content generation based on provided context

Expected Outcome

- **Broadened** LLM knowledge base with domain/organizational specifics.
- Generation of **more accurate and contextually relevant content** via LLMs.

Graph-based Retrieval Augmented Generation: Knowledge Graph Creation

Create a knowledge graph as foundation for Retrieval Augmented Generation

By creating a knowledge graph using best practices, you'll get a clear and organized view of your data that shows how key concepts are connected

Knowledge Graphs help the Retrieval Augmented Generation system find more relevant information and give better answers

Expected Outcome

Structured, contextually rich information, leading to more accurate, relevant, and coherent responses

Graph-based Retrieval Augmented Generation: Query Pipeline

Generate contextually rich responses by leveraging structured knowledge from a Knowledge Graph

Graph-based queries ensure precise and semantically rich retrieval, improving the relevance of LLM responses.

Use appropriate querying approaches and optimize for RAG workloads!

Responses can be linked back to specific triples, entities, or paths in the graph, making results interpretable and auditable.

Expected Outcome

Contextually rich, accurate, and explainable responses to natural language questions by leveraging structured knowledge from a Knowledge Graph

Regression

Simple approach to explain phenomena.

Regression is the process of finding the best fit through a set of data points, essentially summarizing the relationship between two variables through **SAP HANA ML**

In the SAP ecosystem, this involves leveraging tools within SAP HANA ML (PAL, *hana_ml*) to uncover relationships between variables

Expected Outcome

- To predict the value of the dependent variable based on the independent variable(s).
- Allowing for the identification of which independent variables are most influential in predicting the dependent variable.
- Optimize business processes by helping to identify factors that influence key performance indicators (KPIs), test hypotheses, and find optimal input values.

Time Series Forecasting

Using historical time-ordered data to predict future values

Time series forecasting involves identifying patterns like trends, seasonality, and cycles in data collected at regular intervals (e.g., daily, monthly).

In the SAP ecosystem, this entails leveraging tools within SAP HANA ML (PAL, *hana_ml*) to predict future values using historical time-ordered data

Expected Outcome

- Predicting future values based on past observations.
- Helping identify patterns like trends, seasonality, and cycles in data collected at regular intervals
- Optimizing business processes operations by enabling data-driven planning and resource allocation

Anomaly Detection

Simple Identification of Anomalies

Anomaly detection is the process of identifying data points, events, or patterns that **deviate significantly from the expected or normal behavior** within a dataset

In the SAP ecosystem, this involves leveraging tools within **SAP HANA ML** (PAL, *hana-ml*) to find unusual patterns

Expected Outcome

- Successfully identify and flag unusual behavior or outliers in various types of data.
- Enable proactive responses to potential risks through early detection of fraud, system failures, or compliance violations.
- Optimize business processes by identifying operational inefficiencies and understanding unexpected variations.

Classification

Powerful approach to categorize data.

Classification is a fundamental machine learning technique aimed at organizing input data into distinct classes through **SAP HANA ML.**

In the SAP ecosystem, this involves leveraging tools within SAP HANA ML (PAL, *hana_ml*) to automate categorization of data.

Expected Outcome

- To recognize patterns in the training data and use these patterns to classify new data into one of the pre-defined classes.
- Optimize business processes by helping to group similar items together while distinguishing them from different ones, leading to improved understanding, prediction, and decision-making.

Clustering

Simple approach to explain phenomena.

Clustering aims to partition a dataset into subsets (clusters), such that data points within the same cluster exhibit high intra-cluster similarity, while points in different clusters exhibit low inter-cluster similarity.

In the SAP ecosystem, this involves leveraging tools within SAP HANA ML (PAL, *hana_ml*) to partition dataset into groups that share similar characteristics

Expected Outcome

- The outcome of a clustering process is a set of groupings (called clusters) where each data point is assigned to one cluster based on similarity. The goal is to discover natural groupings in data without using labels.
- It helps to identify patterns, segment data, or detect anomalies

Training and Deploying Custom Predictive Models

Integrate Custom AI model into existing business solutions

Custom model development and operationalizing is a multi-step process

AI Core supports container-based deployment, in order to deploy your model

Expected Outcome

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

Document AI

Extracting information from your business documents effectively

SAP Document AI service is an end-to-end document processing solution under SAP Business Technology Platform which helps you automate extraction of structured and unstructured information from your business documents.

This service leverages Machine Learning techniques as well as generative AI to extract the data from your documents.

Expected Outcome

Help extract information from your business documents
Reduce manual effort to process document

Custom Joule Skills

Practical extension of Joule digital assistant

Extending **Joule** functionality creates endless opportunities for tailoring to specific business needs and context.

Custom Skills enable easy implementation with **SAP systems** and external APIs alike

Expected Outcome

- Create custom functionality to extend Joule.
- Leverage external resources directly through single entry point - Joule .

Document Grounding

a ready-to-use capability for generative AI applications grounded in trusted, company-specific knowledge

Directly connect enterprise document sources — such as Amazon S3, SharePoint, SAP Document Management Service, SAP Work Zone, or SFTP file servers.

The service automatically handles document ingestion, chunking, embedding creation, and retrieval

Expected Outcome

- Enable enterprise-grade Retrieval Augmented Generation without managing your own pipelines.
- Seamlessly integrate internal documents and business knowledge into generative AI applications.
- Reduce the risk of hallucinations and improve accuracy in model-generated responses.
- Accelerate AI project delivery with built-in ingestion, vectorization, and lifecycle management.
- Ensure data security, governance, and compliance aligned with SAP’s enterprise standards..

SAP Translation Hub

Simplifying Multilingual Enablement with AI

Document Translation:

- Translate documents with AI using machine translation engines or large language models.
- Automatically detect the source language from the input.

Software Translation:

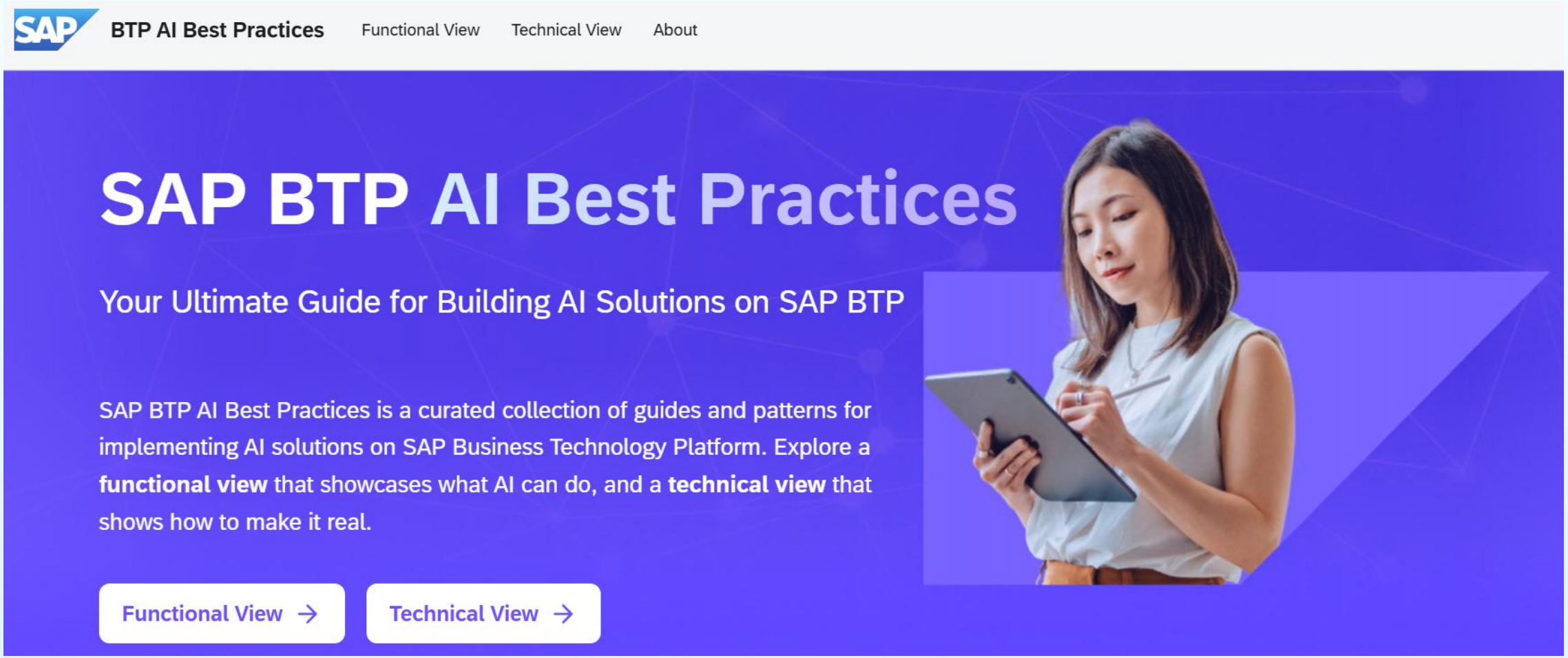
- Translate software and user interface texts using multiple sources.
- Faster localization with consistent enterprise terminology across languages

Expected Outcome

Help to translate software and business documents at scale

How do I access?

https://btp-ai-bp.docs.sap/

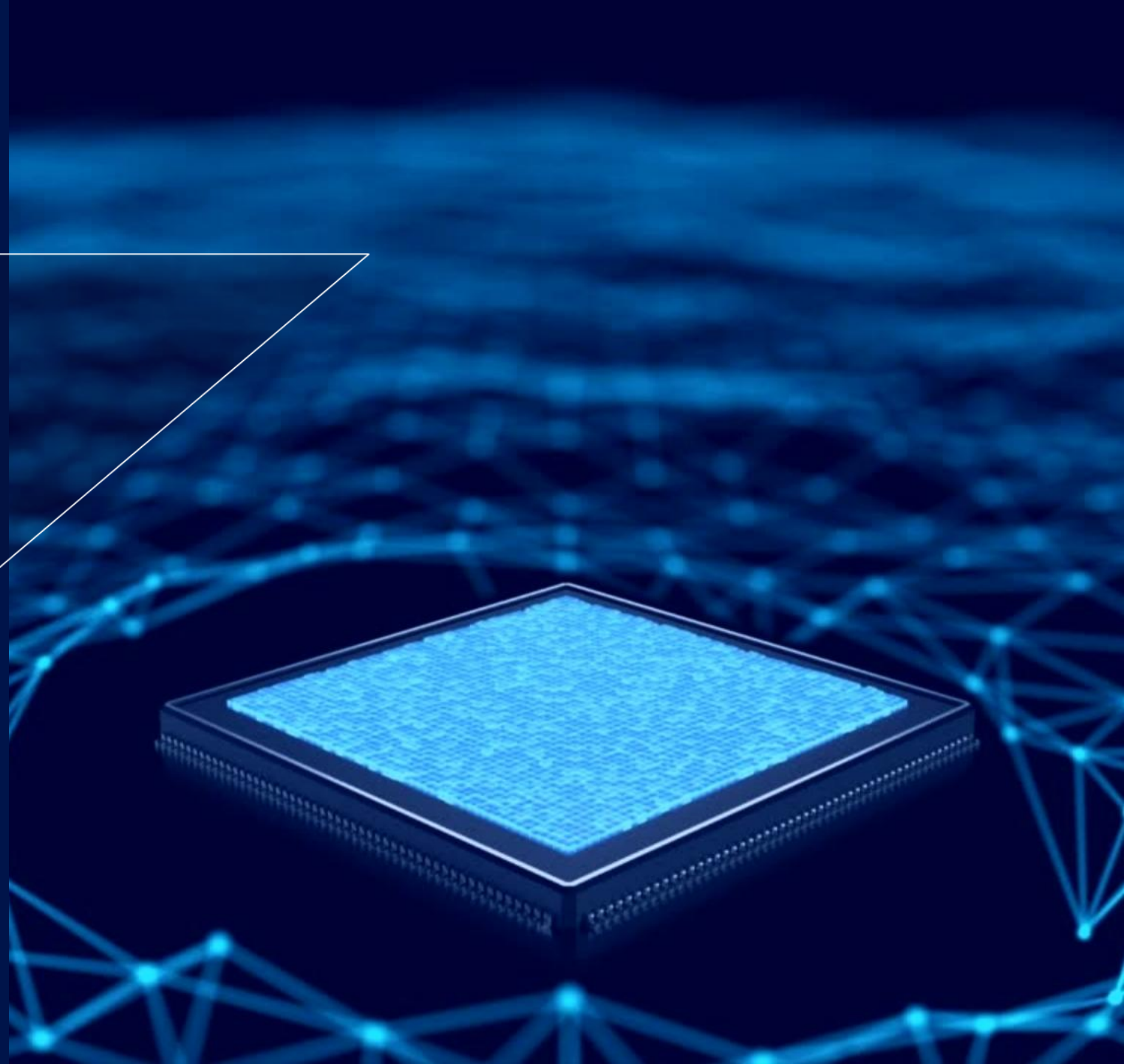


Find it on SAP Discovery Center - SAP BTP Guidance Framework

Thank you!

Contacts:

marco.cigaina@sap.com
l.marques@sap.com



Follow us

www.sap.com/contactsap

THE BEST RUN SAP