From: pieter.van.ommeren@vattenfall.nl
Subject: Fw: Vattenfall IT Architecture AI Guidelines
Date: 4 September 2025 at 14:31
To: Patrick Motsch p.motsch@valueon.ch, Jeroen Haverkorn van Rijsewijk jeroen@h3.company

AI!

Confidentiality: C2 - Internal

From: Ommeren Pieter van (SB) <pieter.van.ommeren@vattenfall.nl>
Sent: Thursday, September 4, 2025 2:30 PM
To: Ommeren Pieter van (SB) <pieter.van.ommeren@vattenfall.nl>
Subject: Vattenfall IT Architecture AI Guidelines

VATTENFALL GROUP IT ARCHITECTU…

Vattenfall IT Architecture AI Guidelines

Hoogenboom Tim (YICA)
IT Architecture

- Architecture Principles
- AI Council
- Vattenfall AI Platform
- Front and Backend Integration
- M365 CoPilot
- Application AI
- Red AI Services
- CoPilot Studio

Document Details

Document ID	VIT-GDL-Architecture-AI
Document Type	Guideline
Document Owner	Andreas Diede

Valid From	June 1, 2025
Revision History	1
Confidentiality Class	C2 - Internal

Purpose

This document provides all Vattenfall Business Areas to select and implement the appropriate AI solutions within Vattenfall.

Validity

Scope related to different IT organizations/divisions within Vattenfall Group

Organization/Division	In Scope/Out of Scope	Remarks
Vattenfall IT	In Scope	Group IT supply unit that supports BA/BU/OU/SF for systems and infrastructure in Purdue Level 4 and 5
Embedded IT	In Scope	IT Supply unit within BA/BU/OU/SF *outside* staff function Vattenfall IT (source: FI31)
Business OT	Out of Scope	BA/BU IT supply unit that supports BA/BU/OU/SF for systems and infrastructure in Purdue Level 1, 2, and 3

Who shall read this document?

This document is intended for a broad audience within Vattenfall. Anyone requiring architecture support or having a responsibility to select, build, or deploy will find it useful.

Document Context

Terms and Definition

Context

The four megatrends decarbonization, digitalization, decentralization, and democratization that utilities, and as such also Vattenfall, are subjecting to, are changing the energy landscape. To thrive in this new energy landscape modern technology capabilities will be needed to ensure security, affordability and sustainability of energy supply that require utilities to become more data driven in the physical processes of generating and distributing energy.

One of these modern technology capabilities within Vattenfall, is leveraging Artificial Intelligence or AI.
This document follows the same definition as defined by the European AI Act stating that AI is ‘a machine-based system designed to operate with varying levels of autonomy and adaptiveness after deployment, generating outputs such as predictions, recommendations, or decisions that influence physical or virtual environments’.

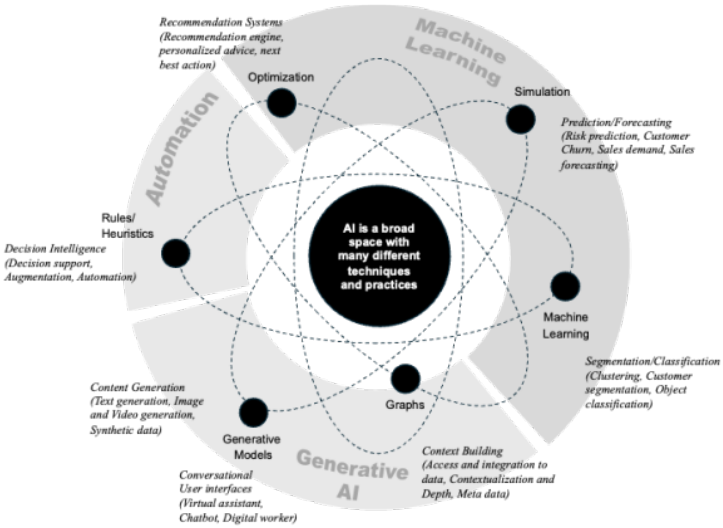
AI will be a crucial enabler and a necessity for the transformation of the energy system. By professionally applying AI, it can contribute to the optimize integration of data, and establishment of a much smarter and interactive energy system are necessary to balance renewable energy, demand response, energy storage, electric mobility, and large-scale energy sources with customer demands, weather conditions, and market prices. This, in turn, enables new services and business models and empowers consumers and citizens. Digitalization can enhance the efficiency, flexibility, and reliability of the energy system. Furthermore, it can foster innovation within the energy sector and European industry.

AI Techniques

AI is a broad field, and generative AI is only one piece of the much broader AI landscape, and most business problems require a combination of different AI techniques. They can often be combined in a way that delivers better accuracy, transparency and performance, while also reducing costs and need for data.

We distinguish three techniques that are most in play when it comes to AI

Area	Description
Generative AI	Refers to AI techniques that learn a representation of artifacts from data, and use it to generate brand-new, unique artifacts that resemble but don't repeat the original data.
Machine Learning	Refers to a technical discipline to feed data into algorithms in a supervised, unsupervised, or semi-supervised manner to get back classifications or predictions. In supervised learning, models are trained on labeled data, meaning the input data is paired with the correct output. Unsupervised learning, on the other hand, deals with unlabeled data, and the model tries to identify patterns and relationships within the data on its own. Semi-supervised learning combines both approaches.
(Cognitive) Automation	Refers to rule-based automation by applying man-made rules to store, sort, and manipulate data. Cognitive automation is a summarizing term for the application of machine learning technologies to automation to take over tasks that would otherwise require manual labor to be accomplished.



AI Techniques and Applications

Value Pools for Vattenfall

Vattenfall has pinpointed four primary value areas of digitalization and AI at Vattenfall

Value Pool	Description	Examples
Employee Productivity	This involves process optimization, and automation further contributes to productivity by reducing the manual effort required for routine tasks, allowing employees to focus on more strategic activities. This is particularly relevant for transforming shared services from a resource and task-driven model to an incubator for digitalizing transactional processes through AI.	
Customer Experience	This involves improving system availability and reducing downtime by utilizing self-healing infrastructure based on sensor-level monitoring to detect events that inform decision-making. This ability to automate and operate autonomously will be essential in managing the growing complexity of assets and systems.	
Asset Development, Construction, and	This involves improving system availability and reducing downtime by utilizing self-healing infrastructure based on sensor-level monitoring to detect events that inform decision-	

Maintenance	inform decision making. This ability to automate and operate autonomously will be essential in managing the growing complexity of assets and systems.	
Asset Value Optimization	With increasing unpredictability in power generation, market volatility, and changing demand patterns, the energy system is rapidly evolving. Real-time systems with higher responsiveness are the foundation for maintaining balance and efficiently managing resource allocation. These systems are fed with faster and detailed forecasts and data of consumption, weather, assets or other conditions to address questions like grid instability, meet emissions reduction targets, or allow for higher grid capacity. This presents opportunities to maximize revenues and enhance asset value.	

General

1. AI solutions processing Vattenfall data (C2 or higher) shall be reviewed by Vattenfall IT Architecture and Security.
2. All Vattenfall AI solutions shall meet requirement as stipulated in Vattenfall Governance AI.
3. All Vattenfall AI solutions shall meet requirements as stipulated in Vattenfall AI Security Standard.
4. All Vattenfall AI solutions shall adhere to principles of

Responsible AI.

a. *Fairness* – AI systems should be fair in their treatment of all people.
b. *Reliability and safety* – AI systems should operate in a reliable and secure manner.
c. *Privacy and security* – AI systems should be secure and should respect the privacy of individuals.
d. *Inclusiveness* – AI systems should be about empowerment and inclusion for all people, regardless of their background.
e. *Transparency* – AI systems should be comprehensible.
f. *Accountability* – Humans should be accountable for AI systems.

Vattenfall developed/hosted AI solutions

1. Vattenfall AI solutions shall run in Primary Zone IT according to ISMS110.
2. Vattenfall data processing by AI solutions shall be under control of Vattenfall.
3. Vattenfall cloud hosted AI services shall adhere to Vattenfall Cloud Principles.
4. Vattenfall cloud hosted AI services shall be implemented with so data cannot be exfiltrated or exposed publicly.
5. Vattenfall AI Platform (VAIP) shall be used for development, hosting, management, and operation of cloud AI-driven solutions.
    a. *Remark: VAIP is designed as a preapproved and prequalified service that meets principles G, V1-V4.*
6. Any services that need to connect to the AI services shall be fully integrated into the virtual network to communicate only over private IP addresses.
7. DevOps tools shall be used for continuous integration and continuous deployment (CI/CD) for AI models to ensure integration, version control, repeatability, etc.
8. Vattenfall AI Services shall be hosted in:
    a. *Vattenfall Cloud* – By default up to C3, GDPR Normal – and GDPR High via additional controls).
    b. *Outside Cloud/External* – By default up to C1, or unless higher levels approved via IT Architecture and Security Review by adequate implementation of ISMS controls).
    c. *On-premises* – By default up to C4, Swedish sensitive data, or export-controlled workloads).

Commercially developed/hosted AI solutions

1. Activation of application embedded-AI solutions or extensions (despite previously ITAS reviews for the application) shall be approved by Vattenfall IT Architecture and Security as mostly AI services rely on third party services outside Vattenfall IT control.
2. External (SaaS) solutions shall have technical controls contractually agreed so data cannot be exfiltrated or exposed publicly.

Within Vattenfall IT AI solutions are governed by the AI Council to create an authoritative body that ensures overview, oversight, and guidance within Vattenfall IT to successfully enable AI to Vattenfall BA/BU/SF.

The AI Council is a executive leadership forum (N2/N3 that are accountable for delivering AI solutions, note that the calander when meetings take place is published at the bottom of this section. Search for **AI Council** in case you want to find the **next meeting** and/or **submit demand** via your AI Council Represenative.

Purpose - Business Enablement

The purpose of the AI Council is to

1. Create a portfolio overview of incoming/ongoing AI solution delivery towards the BA/BU/SF. The inventory is published here)
2. Create a common overview of AI Solutions that are in use of Vattenfall, Decision Trees, Governance to be able to enable a decentralized delivery fitting each BA/BU/SF against common principles.
3. Guide new incoming requests that don't have a clear solution, towards the right channel.
    a. Note: AI requests follow similar to all IT requests the BIO/SPOC dialogue for realization.
4. Follow that new AI deployments follow defined ethical guidelines

4. Follow that new AI deployments follow defined ethical guidelines and supports corporate social responsibility (CSR) goals.
5. Monitor adherence to legal frameworks (e.g., GDPR, EU AI Act) and implement necessary adjustments to ensure compliance.

Purpose - Technology Enablement

Next to the governance aspect, the AI council also services a technology aspect, to advance enterprise, standardized and secure usage of AI services by potentially limiting access to models, services, settings – as per default all is open e.g.

1. *Prioritize enablement topics*, e.g. What AI services should be enabled within the central AI solution space (see next Header).
2. *Ensure IT standards*, e.g. What large language models to use for code development (as there are output and performance difference is code quality).
3. *Ensure connectors*: What connectors do we control as they might run risk of C3 data being exported to non- or under-classified environments.
4. *Model explore*: What controls to implement to uncontrolled exposure of credentials to third party LLM/AI services (e.g. Hugging Face models might have ability for remote code execution by design) and especially when using agentic features also credentials are needed to connect to backend systems.
5. *AI Usability*: Collect BA/BU/SF feedback on usability and deployability of AI services and continuous improvement.

Meeting Calender for AI Council

Find the next meeting for the AI Council, by searching for **AI Council** and see when the AI Council meets and/or submit new AI enablement demand via your AI Council Representative (follow the OLC logic)

Occurs the last Monday of every month from 16:00 to 17:00.

Members

Composition of AI Council consists of:

1. Managers from business facing IT units for Customer, Asset and Corporate delivering AI-based solutions
2. Experts from supporting units Architecture, Data Governance,

Compliance, Security involved in AI policy setting.
3. Current Chair is Head of IT Architecture

AI Council Representatives per OLC

Meijer Edwin (YIAN)
Customer IT Netherlands

Representative YIA AI delivery and ena blement

Wickramasinghe Yasith (YIAS)
Customer IT Nordic

Representative YIA AI delivery and ena blement

Grabowski Michal (YIPB)
Asset Analytics &IoT

Representative YIP AI delivery and ena blement

Malucha Michał (YIPC)
Asset Software Engineering

Representative YIP AI delivery and ena blement

Krüger Matthias (YICAD)
Digital Platform Enablement

Representative YIC AI delivery and ena blement

Hoogenboom Tim (YICA)
IT Architecture

Representative YIC AI (and Chair)

Stockmann Ramon (YIOC)
Public Cloud Platform Services

Representative YIO AI delivery and de mand

Kramer Hans (YIOD)
Data Centre

Representative YIO AI delivery and de mand

AI Council Experts

Diede Andreas (YICAB)
Business Facing

Lead Architect for AI

Kaklij Sunil (YICAB)
Business Facing

Lead Architect for Data and Analytics

Kaus Sebastian (YICAD)
Digital Platform Enablement

Lead Architect for Data Governance

Anjum Muhammad Naveed (YICMC)
Common Solutions Management

Lead for CoPilot (M365, Agents, SP, St udio)

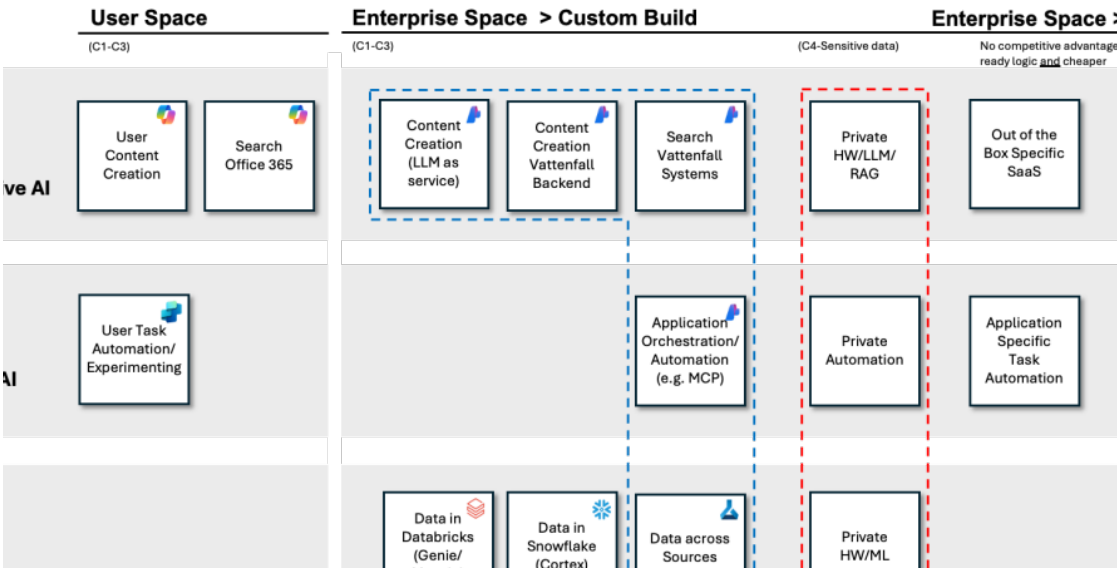
Nordström Elisabeth (YIFG)
General Compliance

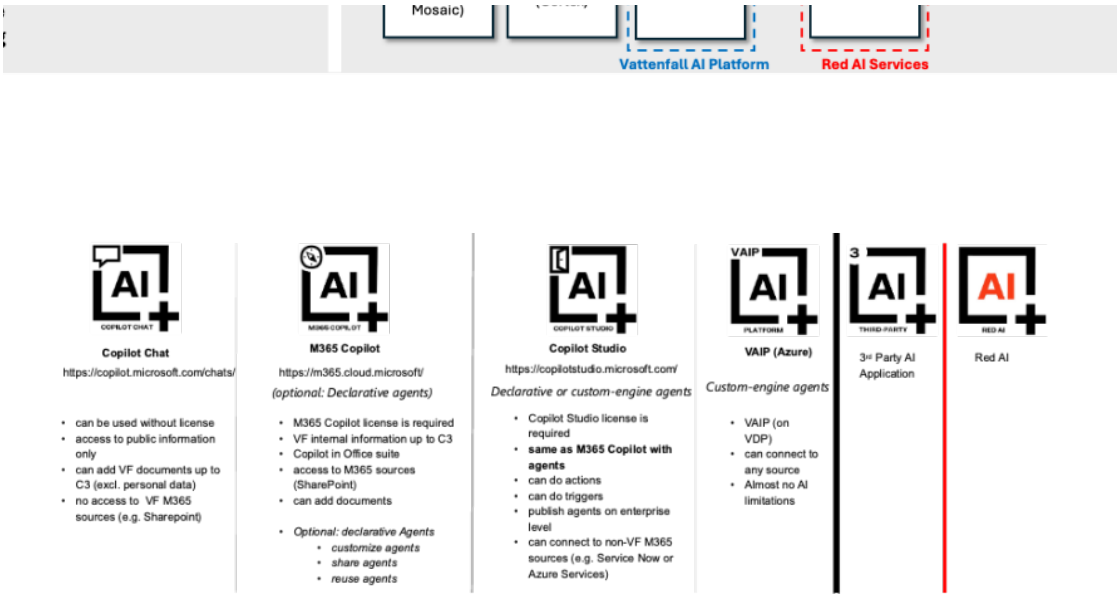
Compliance and Governance for AI

Kültür Murat (YISG)
IT Security Strategy Services

Lead Security for AI

Within Vattenfall we enable various solutions to support AI based tasks. The AI solution space looks like.





Employee AI

For employee productivity Vattenfall enables out of the box

CoPilot Bing

CoPilot Bing, similar to ChatGPT, enables all employees in Vattenfall to query publicly available data via natural language.

M365 CoPilot

M365 CoPilot, similar to ChatGPT, this service enables licensed employees in Vattenfall to query next to publicly available data, also Vattenfall data accessible via Microsoft Graph (which is mostly Microsoft Office 365 apps, like SharePoint, PowerPoint, Teams) based on the users' permissions. Excluded are shared and delegated mailboxes, archived mailbox and SharePoint data.

CoPilot Agents or SharePoint Agents enable end users to build a more context specific agent, for example you can pinpoint the agent to ground replies on selected policy documents or annual reports, to ensure the accuracy of the answers, by excluding web searches, irrelevant versions of documents or exclude team or emails.

Enterprise AI Workloads

Vattenfall AI Platform (VAIP)is an offer from Vattenfall IT to developers to deploy enterprise ready and secure AI Projects via Microsoft Enterprise Scale ML and AI Factory(ESML+AI see GitHub).

Microsoft Enterprise Scale ML + AI Factory

ESML+AI is a plug and play solution that automates the provisioning, deployment, and management of AI projects on Microsoft Azure with a template way of working.

VAIP modified Microsoft AI Factory scripts so that these also work within the specific Vattenfall cloud setup, called Vattenfall Digital Platform (VDP), e.g. Vattenfall has a hub-spoke network topology in the cloud, using central Check Point firewall as virtual security appliances, whereas the previous Microsoft AI Factory scripts could only be deployed with (functionally less advanced) Microsoft native Azure firewalls.
VAIP follows a same concept as Vattenfall Analytics Platform (VAP) – whereby underlying infrastructure and networking is centrally readied, so IT developers can focus on simply developing AI services and connecting them to their resources.

Due to its pro-code nature VAIP offers maximum flexibility as it can connect to both on premises and cloud resources, has many (custom) connectors, and can leverage an extensive catalog of models and AI services (ranging from large language processing, model inference, agentic solutions, traditional ML, image recognition.

Microsoft VAIP creates dedicated resource groups that contain key Azure AI services.

Microsoft Foundry

Microsoft AI Foundry is an integrated platform for Developers and IT Administrators to design, customize, and manage AI applications and agents by offering an integrated GUI with pre-built model catalog, pre-built templates, development environment and interactive playground to do hands-on experimentation.

Microsoft AI Services

Microsoft AI Services are market-ready AI services to build applications

Microsoft AI Services are market-ready AI services to build applications to support natural language processing for conversations, search, monitoring, translation, speech, vision, and decision-making that can be found in AI Foundry.

Enterprise Security and Architecture by Design

VAIP additionally adds Microsoft Well Architected Framework best practices, by adding enterprise security, e.g.

- It reuses existing Microsoft Landing Zone Architecture and best practices from Azure Cloud Adoption Framework and Well Architected Framework.
- Provides full Dev, Test, Production environments, which can by default scale up to 200-300 projects in each environment.
- Full private networking: Private endpoints for all services such as Azure Machine Learning, private AKS cluster, private Container registry, Storage, Azure data factory, Monitoring, etc. to avoid data exfiltration as they are not as per default publicly exposed.
- Dynamically create infra-resources per team, including networking dynamically, and RBAC dynamically.
- VAIP is project based (to ensure for each Team cost control, privacy, scalability per project) and provides multiple templates besides infrastructure templates (e.g. there are template to have turnkey data lake and AI services).
- As all VAIP services are deployed in common AI workload subscriptions (similar to VAP for standardized analytic services like Azure Databricks) this also enables us to centrally deploy new feature sets that will work for all, enforce security policies – but also to test them upfront to ensure zero unforeseen impact.

CoPilot Studio currently excluded to build enterprise AI solutions

Note that CoPilot Studio is a low-code tool for building agents to automate steps. It can be seen as an extension to CoPilot agents, but then with the ability to add agency, implying the ability to interpret and execute simple steps by calling APIs in surrounding applications.

As the underlying architecture of CoPilot Studio is based on Power Platform and Office 365, currently there are various technical limitations that limit its ability to serve enterprise use cases (e.g. limitations in number of documents to be searched < 3000, inability to search documents older 1.5 years and only using meta data of its

sources to create an index and not the actual data, lump sum cost charging only as cost control and pay per use features are lacking).

Vattenfall Group IT Architecture recommends CoPilot Studio as experimentation environment for individual usage, whereas any enterprise deployments shall be realized via VAIP.

Regulated AI workloads

Next to cloud-based AI services, Vattenfall IT also delivers Red AI Services.

These provide on-premises GPU hardware that utilizes open-source large language models that can be trained for specific use cases, to extend sensitive applications that cannot be hosted in or connected to cloud to be augmented with AI services.

Note: At this point of time VIT provides on-premises AI services, that can be evolved into regulated AI services by applying additional organizational controls.

Commercial/Application inbuilt AI

Many application vendors offer application embedded AI, to enhance the functionality of their application.

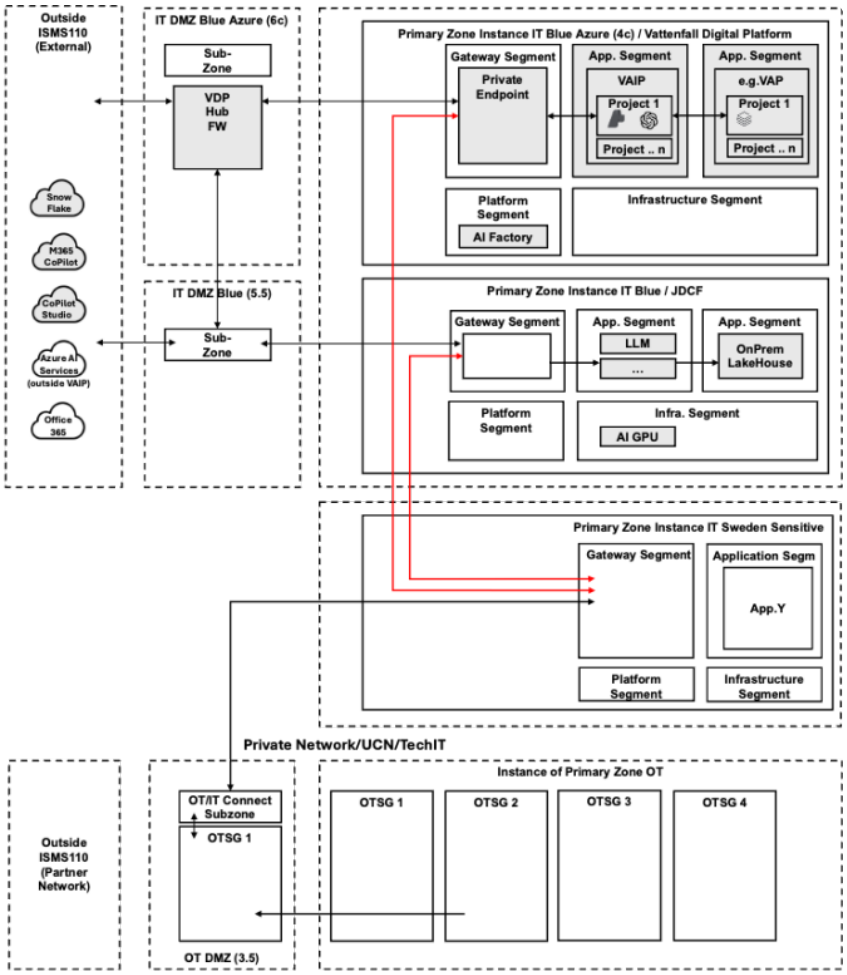
As these enhanced AI features typically utilize external services that might process Vattenfall specific data (Models, hosting providers, etc.) and thereby are always subjected to IT Architecture and Security Review.

Typically, application embedded AI is trained and optimized against the specific context of the application insights, process logic and data model and thereby might render better results than CoPilot and/or VAIP. Typically, these AI services also come with out of the box connectors or API's that can be leveraged, shortening their deployment time. These solutions are mostly limited to their specific application context, and thereby not suitable to be combined with data sources outside the solution, here more generic solutions like VAIP offer benefits.

Next to application embedded AI all Data Lakehouse providers, like Databricks and Snowflake also enrich their data platforms with AI features. The benefit of these data platform providers is typically that the AI features are well-integrated as part of their SaaS offering and can be easily used against the stored data, and thereby again storing their deployment time. Nevertheless, similar to VAIP and CoPilot they are not trained against business logic and data, rendering them good for experimentation and fact finding, but not for enterprise solutions.

The current overview outlines Vattenfall Zone Model with Primary Zones and outlines where what AI service is running. Current deployments of AI Services are being configured in External Zone (SaaS) and Primary Zone IT – both in Azure and on-premises.

Please note this is a simplified overview intending to display in line with ISMS110 Vattenfall Zone Model and Primary Zone Modell IT VITSMS111, where what AI services are deployed. More details per AI solution is conveyed in the next sections.



Vattenfall Primary Zones and deployment of AI Solutions

CoPilot Architecture

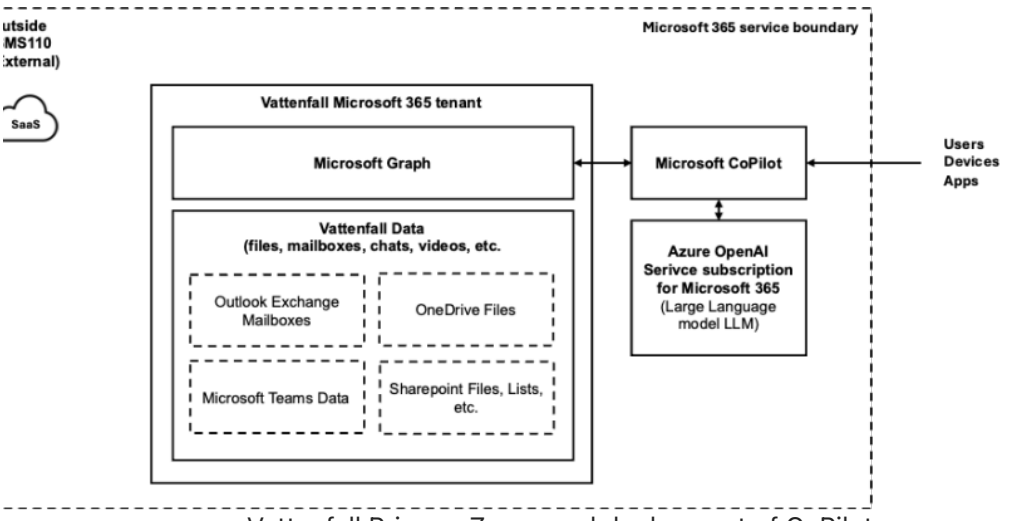
Below CoPilot architecture is based on [Microsoft 365 Copilot Architecture](#).

Vattenfall has a tenant within the Microsoft 365 service boundary where Vattenfall data (in SharePoint, Outlook etc.) resides, which according to Vattenfall ISMS110 is considered External/SaaS. However, M365 is IT Architecture and Security reviewed.
CoPilot is a shared service outside the Vattenfall tenant, but in the Microsoft 365 service boundary.

Copilot only accesses data that an individual user is authorized to access, based on, Microsoft 365 role-based access controls controlled by Vattenfall.

CoPilot activity history (prompts and responses) can be deleted by individual users.

CoPilot Agents are a further low code simplification of CoPilot Studio (see below) to direct CoPilot to act within a more bounded context to improve user experience and limit hallucination (e.g. by limiting search results to specific directories or subject or conversational).



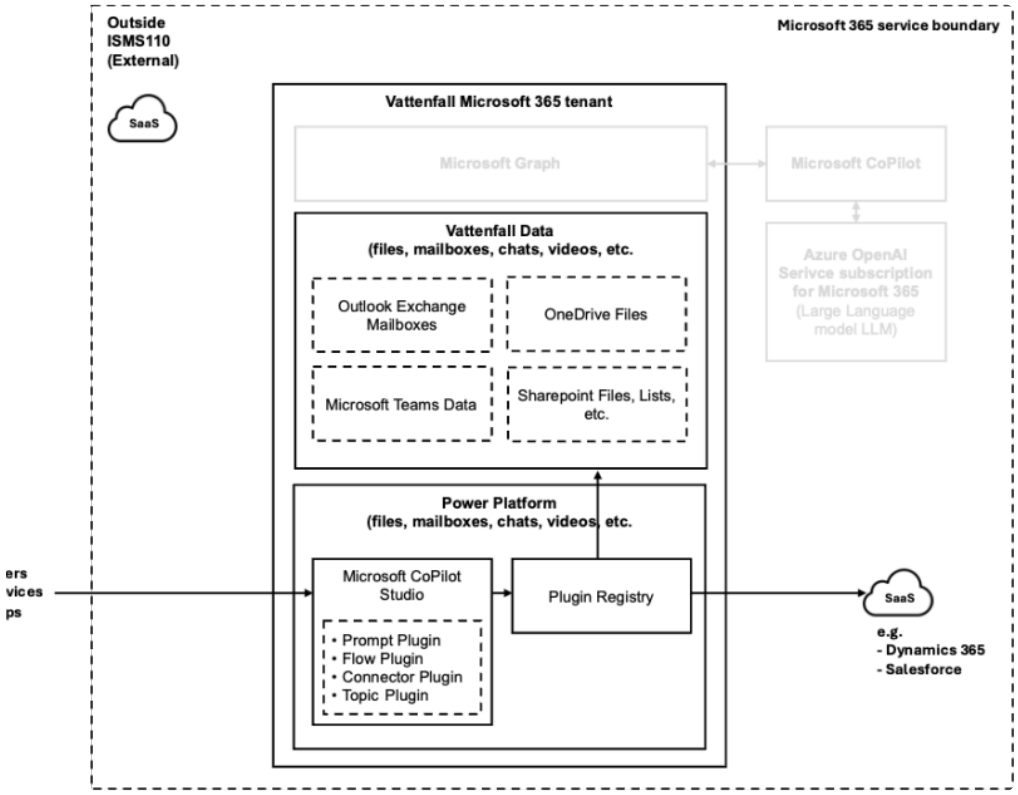
Vattenfall Primary Zones and deployment of CoPilot

CoPilot Agents and/or Studio Architecture

Below CoPilot Agent and Studio architecture is based on [Microsoft 365 Copilot Studio Architecture].

Although CoPilot Studio carries the same name, the underlying architecture is based on the Microsoft Power Platform to support agentic automation of (simple) task execution via standard connectors to Power Platform supported ecosystem or Dataverse ecosystem.

These connectors allow for vendor specific approved agentic AI processing towards selected backend systems via (Power) Automate by using via natural language.



Vattenfall Primary Zones and deployment of AI Solutions.

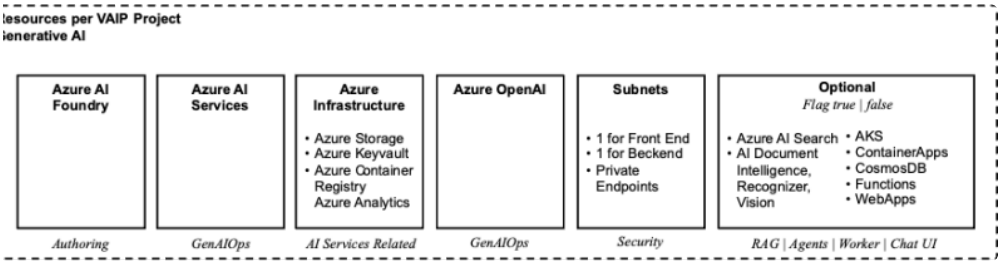
Vattenfall AI Platform Architecture

Vattenfall AI Platform (VAIP) is an extension of the Vattenfall Digital Platform (VDP) services, that furthermore comply to the new [YICA-STD-Cloud IT Architecture Standard for Microsoft Azure] (version 2.0) and [VITSMS-406 AI Security Standard] (version 1.0), and an self-service implementation similar to Vattenfall Analytics Platform (VAP) analytical services in Azure.

By extending VAIP as part of VDP updated security policies and architecture guidelines can be centrally enforced and tested upfront to have zero impact (as known configuration, in comparison to BA or IT specific subscriptions).

VAIP provides a Microsoft Enterprise Scale AI Factory-ready-for VDP GenAI template that deploys the most used Generative AI services and infrastructure services as listed below.

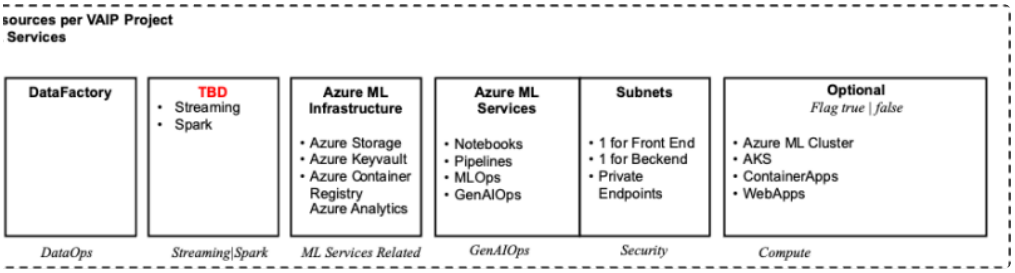
Wave 1 deployment – Generative AI services



Self-service deployable Generative AI services in VAIP

Wave 2 deployment - Machine Learning Services

For the next wave more Machine Learning services will be made available.



self-service deployable AI services in VAIP (next release) – to be updated based on input Andreas

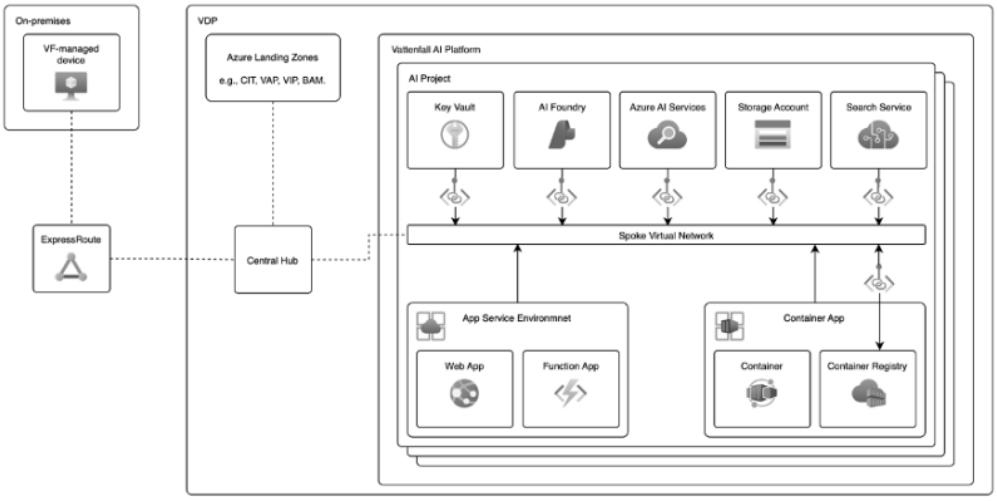
Important is that VAIP template also provides centralized provisioning of network settings (creation of virtual network, creation of subnets, creation of routing tables, linking to private endpoints, creation of private DNS zones) without involvement of persons. Furthermore, complete hub spot integrations ensures that azure resources can frictionlessly communicate between Hub and spoke/subscriptions in VDP.

Furthermore, VAIP services are integrated as part of the same cost charge out of VDP, e.g. External SaaS solutions are integrated for one and can be used by all, AI cost charge out is integrated in VDP billing.

VAIP Technical Blueprint

The services are provided in central subscription (akin to VAP) for decentral AI development by all IT teams due to that

1. AI Governance (policies) approved by AI Council can be implemented centrally (e.g., limiting LLM whereof we know they allow remote code execution)
2. As we aim to control/avoid data exfiltration Azure default network settings need to be reset – which can only be done, checked, and maintained centrally and within the VAIP subscription. Note this is the first subscription where **all** Azure services are only exposing private endpoints – which is following Azure Well Architected Framework and will be the norm for future VDP blueprints.
3. Decentral deployment of AI Foundry will not scale at the enterprise level (effort to propagate all network settings to other subscriptions with may be other conflicts / dependencies)
4. Decentral approach is simply not supported by Microsoft Enterprise AI Factory script, due to aforementioned concerns.



Front End integration towards VAIP AI Services

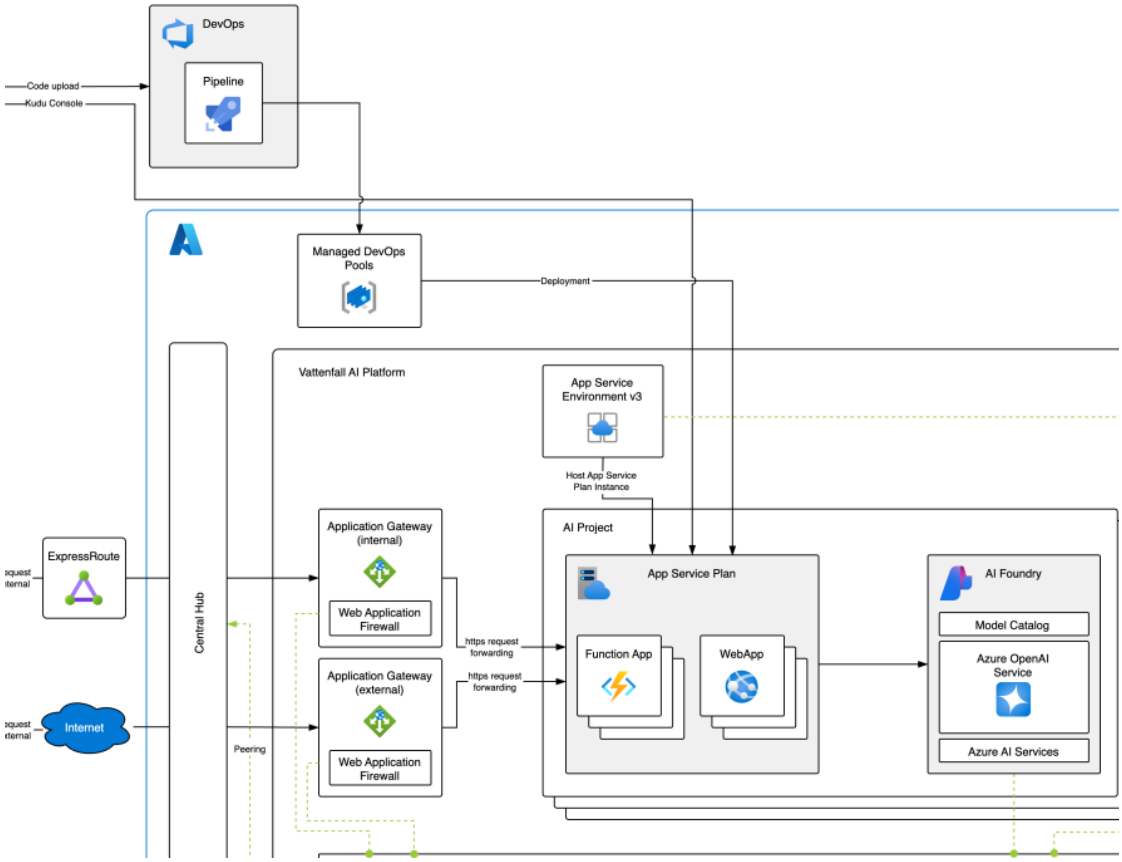
For front end integration we will expose CAPIM for internal and external

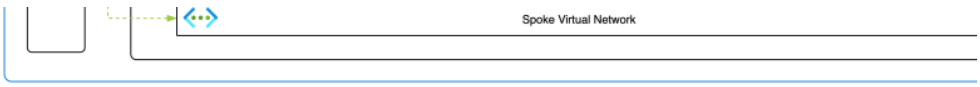
services, to reach via private link endpoints exposed by data bricks or snowflake. These endpoints needs to be exposed by Databricks and Snowflake, and support **querying** purposes by Microsoft Foundry/GenAI services or an MCP server.

VAIP offers the resources to build generative AI services, typically the Vattenfall data it needs is residing in Databricks or Snowflake instances.
To connect VAIP Azure AI foundry instances to Databricks, via Databricks Unity Catalog an *AI agent endpoint* needs to be created to expose the data externally – and if needed execute custom logic and perform specific tasks.

As both the VAIP Generative AI services and VAP Analytics Services (including Databricks) are running within VDP, Azure AI Foundry in VAIP can connect via service credentials or managed identities to these Databricks AI agents, and as such LLM's can query these Databricks workspaces.

Furthermore, the current VDP App Service Environment to build web apps, logic apps, container apps will be privately integrated to connect to Microsoft AI foundry/Gen AI services.





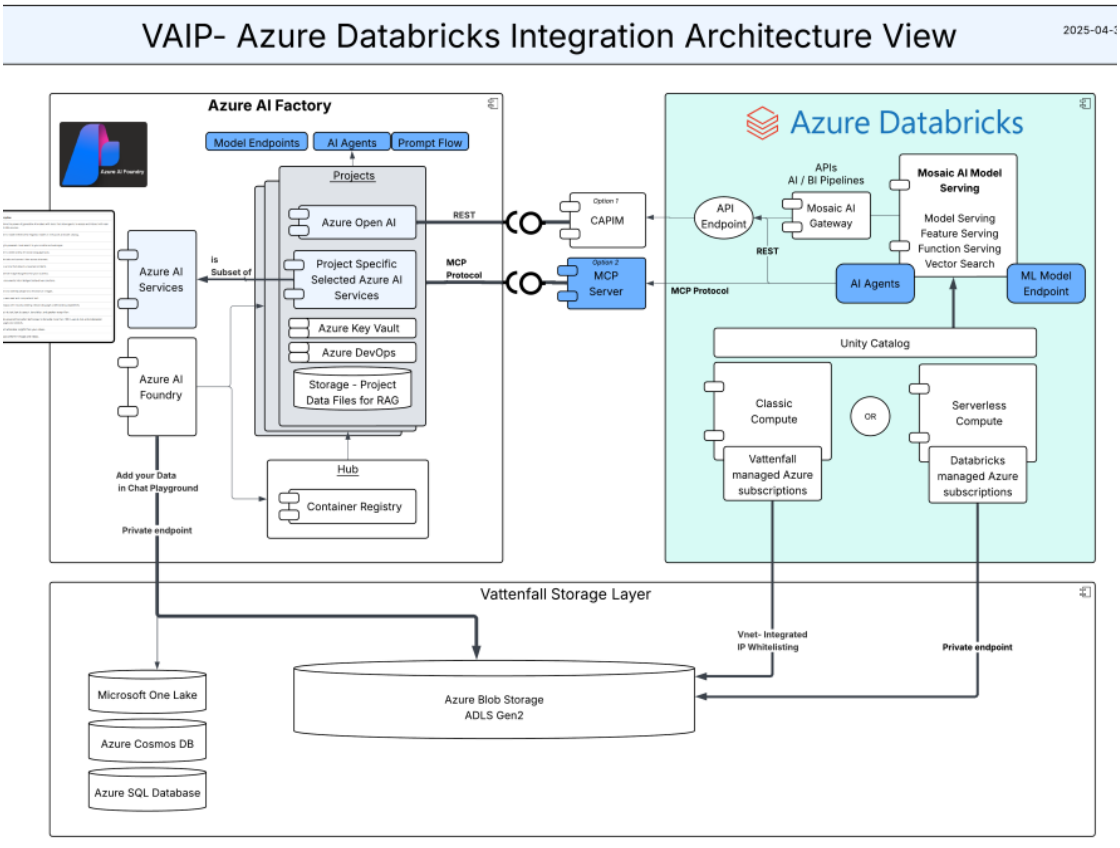
Backend End integration towards VAIP AI Services

To connect to Databricks for machine learning applications, similarly Unity Catalogue needs to expose a *model serving endpoint* needs to be created to expose the data externally.

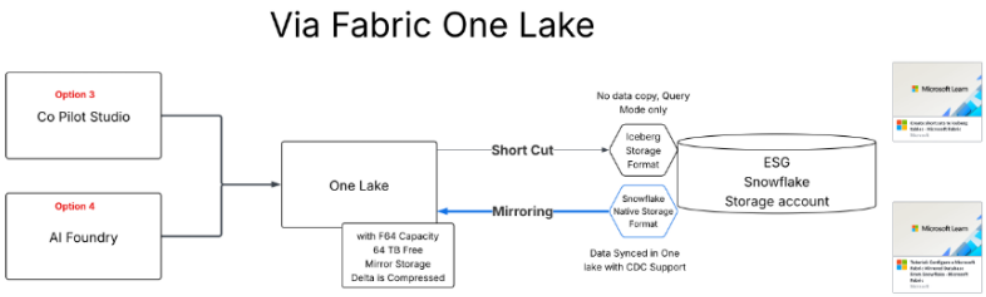
In VAIP Wave 2 Machine Learning Services will be made available, so that Azure ML Workspace to train models to run on data in Databricks workspaces (alternatively models can also be directly trained in Databricks)

When it comes to Snowflake running on Microsoft Azure ADLS (Gen2) (like Vattenfall) Microsoft AI Foundry and CoPilot Studio has inbuilt connectors to connect to Snowflake tenants via service credentials or managed identities.

SharePoint connector is under development.



In case of need for increased performance, Microsoft AI Foundry and CoPilot can also be connected via Microsoft Fabric/OneLake towards Snowflake storage accounts.



AI powered Enterprise Search (solution) running on VAIP

The idea of Enterprise Search is: users should be able to ask any question, and using their identity, a large language model (LLM) will query the appropriate knowledge sources within Vattenfall and try to find the needed information. At the end, the LLM summarizes the search results, and gives this to the user as answer, together with citations to the original source documents.

Such a setup, where an LLM connects to data sources to enrich its inference-time knowledge, beyond what's been learned during training, is called RAG (retrieval-augmented generation). For Enterprise Search in VAIP, we have an AI Foundry Hub and a project within it, where we create model deployments, e.g., for GPT-4o. Enterprise Search then connects the model to external systems that are able to search a source (like SharePoint) reliably based on a user's question. These systems are currently being built for SharePoint and our Intranet. Essentially, they can be thought of as other agents, which the main Enterprise Search orchestrator can talk to.

Responsibility for the "R" (retrieval) in "RAG" is therefore delegated to sub-systems, rather than centralizing this task and putting all of Vattenfall's data into a single Enterprise Search vector store. This has the advantage of allowing owners of their respective systems to design more freely and not force them to dump all their data into our subscription / vector stores. The only requirement is to adhere to an API standard which is currently being developed, so that Enterprise Search can query all systems in a consistent manner.

Below, we explain on a high level which sub-systems are connected to

Below, we explain on a high level which sub-systems are connected to Enterprise Search. At the time of writing, these systems are also in active development and cannot yet be used.

Vattenfall SharePoint GPT Extensions

The system which Enterprise Search queries to get insights about SharePoint data is called Vattenfall SharePoint GPT Extensions, formerly Lookout GPT. This system, when finished, comprises of several key features, not all of which are directly related to Enterprise Search:

- Crawling a given SharePoint document library and enriching the embeddings with metadata about the documents, such as BA/BU, language, author, key findings, etc.
- A web part (plugin in SharePoint) via which the user can chat with an assistant that is scoped on a selected set of document libraries. Rather than searching all of SharePoint, the user gets to choose which scopes make sense for their query.
- An API to search through the currently indexed data; this is what Enterprise Search will invoke to get search results from SharePoint.

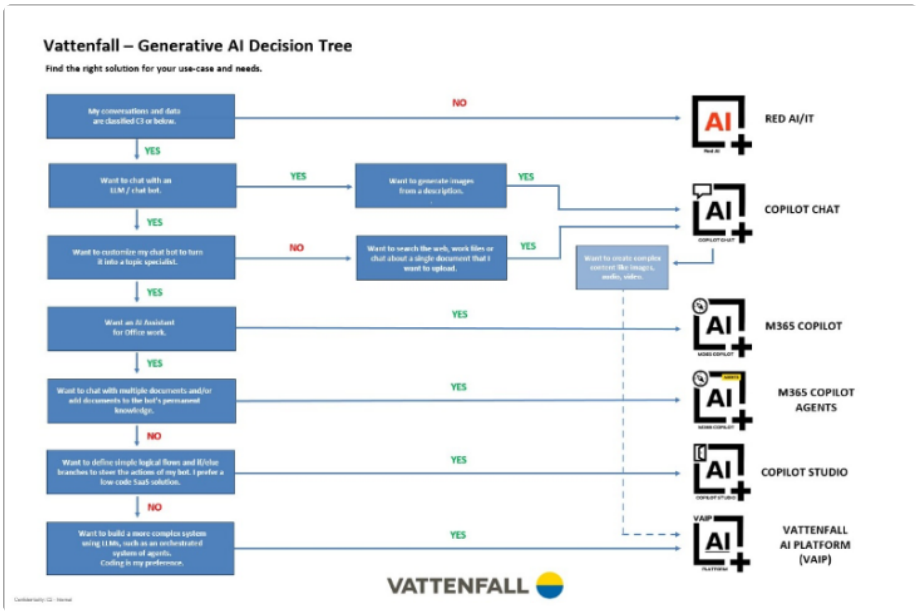
Intranet Search

This project aims to deliver a new search mechanism for Vattenfall's Intranet, powered by AI. Intranet content is crawled into an Azure AI Search service, so that hybrid queries - semantic and keyword search - can be performed on the content. Key challenges include crawling all the content into an LLM readable format (such as Markdown), incrementally rebuilding the index periodically when Intranet content is updated, and handling security trimming, i.e., filtering searchable content based on the user's identity.

In the complex landscape of Vattenfall's systems and applications, structured decision-making is essential for choosing the right AI solutions. The "Vattenfall – Generative AI Decision Tree" is a useful tool, designed for this purpose, guiding users through a series of questions to identify the most suitable AI environment for their needs. This decision tree helps to select the right environment to manage confidential or sensitive data, interact with AI assistants, generate images from descriptions, summarize text, and automate tasks. By providing a clear and systematic pathway, it ensures users can efficiently navigate the diverse AI offerings within Vattenfall's

efficiently navigate the diverse AI offerings within Vattenfall's ecosystem. This approach not only enhances decision-making but also aligns with Vattenfall's AI Architecture Guideline, fostering a cohesive and optimized deployment of generative AI tools across the organization.

The decision tree starts by determining whether the conditions and data are confidential, branching out based on the answers. Users are guided through requirements such as AI interaction, image generation, text summarization, and task automation. It directs them to relevant Vattenfall AI tools like RED AI/IT, Copilot Chat, M365 Copilot Agents, Copilot Studio, and the Vattenfall AI Platform (VAP), ensuring alignment with Vattenfall's AI Architecture Guideline.



Generative AI Decision Tree

Summary

This document will help the user to get started with an AI project in the development environment of the Vattenfall AI Platform.
The document contains step-by-step instructions on how to order an AI project within the Vattenfall AI Platform (VAIP).

Author: Andreas Diede (YICAB)
DRAFT
Valid from: May 14

Revision History

The following table shows the document's revision history.
Table 1: Revision History

Revision	Description of Change	Responsible	Date
0.1	Initial internal draft	Andreas Diede	24.04.2025

Introduction

The Vattenfall AI Platform development environment can now be used to start developing AI solutions. In the first step, the following services are automatically deployed within the corresponding resource group.

1. Azure AI Hub
2. Azure AI Foundry
3. Search Service
4. Azure AI Service
5. Bing Search
6. Azure AI Vision
7. Azure Speech
8. AI Doc Intelligence
9. App Service Plan
10. Function App (Does not work now. Only with CA Client (VDP Windows Client) related to Proxy Ticket.)
11. Web App (Does not work now. Only with CA Client (VDP Windows Client) related to Proxy Ticket.)

More services will be launched in the coming weeks. Please keep in mind that the Vattenfall AI Platform is also in a development phase at the moment and is not yet available in its entirety.

How to order an AI project on the Vattenfall AI Platform

To set up a dedicated resource group, also called a project, within our development environment at the Vattenfall AI Platform (VAIP), it is necessary to follow the steps outlined below.
First, please note that the enumerated information is mandatory and must be included in the ITSP ticket.

1. **CAT4 Number** – A valid cost allocation element is required to charge the Azure run costs.

2. **Owner** – A dedicated person who is responsible for the requested AI project.
3. **Description** – A brief overview of the AI use case.

Secondly, please create an ITPS ticket according to the following instructions.

1. Open the Web browser on the VF device and go to the IT Service Portal (ITPSP).
2. In the search field, search for "VDP" and click on the first hit in the list "Vattenfall Digital Platform (VDP)".
3. In the new window that appears, enter the following information as shown in the screenshot below.
    a. **Oder Options**: Requests, Changes and Enhancements
    b. **Short description**: AI-Project – AI Foundry
    c. **Level 1**: Azure services
    d. **Level 2**: New implementation
    e. **Resource Group**: (not needed, just add an "- " Resource Group will be created automatically during our deployment)
    f. **Subscription**: (not needed, just add an "-" We will use our Vattenfall AI Subscription)
    g. **Resource:** (not needed, just add an "- " You get resources which are listed above)
    h. **Error message: "-"**
    i. **Description:** CAT4 – WBS Element, Owner + Member of the AD Group which give access to the resources.
4. Click **Next**
5. Click **Submit**

Support Request

As mentioned above is the Vattenfall AI Platform is also in a development phase and therefore, some issues may occur at any time. If this is the case, please create a new VDP ticket via IT Service Portal (ITPSP) with a detailed description of the problem and, if possible, a screenshot of the error. The VAIP project team will then investigate the problem and try to resolve it as quickly as possible.

View in SharePoint

How was the formatting of this email?

controlled by Vattenfall AB.