NATIONAL GAS TRANSMISSION ASSET CORTEX AI

Cortex Final Report

National Gas Transmission plc

Report no.: 001, Rev. 2
Document no.: 3005769
Date: 2026-06-05

Project name:	National Gas Transmission Asset Cortex AI	DNV Services UK Limited Energy Systems
Report title:	Cortex Final Report	Digital Trust
Customer:	National Gas Transmission plc, National Grid House Warwick Technology Park Gallows Hill CV34 6DA Warwick United Kingdom	5th Floor Vivo Building Vivo Building 30 Stamford Street SE1 9LQ London United Kingdom GB 440 60 13 95
Customer contact:	Sabia Sadiya	
Date of issue:	2026-06-05	
Project no.:	10591706	
Organization unit:	Digital Trust	
Report no.:	001, Rev. 2	
Document no.:	3005769	

Applicable contract(s) governing the provision of this Report:

Objective:

Report the findings of Asset Cortex proof of concept project.

Prepared by:	Verified by:	Approved by:
Gabriel Griffin Booth	Rachel Hassall	Graham Faiz
Gabriel Griffin Booth Principal Consultant	Rachel Hassall Data Science and AI Team Lead	Graham Faiz Head of Digital Energy
B.Ross		
Benjamin Ross Graduate Consultant		
Andrew Lawrie		
Andrew Lawrie Graduate Consutant		

Internally in DNV, the information in this document is classified as:

	Can the document be distributed internally within DNV after a specific date? No	Yes
☐ Open ☐ DNV Restricted	--	--
☒ DNV Confidential ☐ DNV Secret	☒	☐

Additional authorised personnel for distribution within DNV:

Keywords Artificial Intelligence, Agentic AI, Information Modelling

Rev. no.	Date	Reason for issue	Prepared by	Verified by	Approved by
2	2026/06/05	First complete issue	Gabriel Griffin Booth, Benjamin Ross, Andrew Lawrie	Rachel Hassall	Graham Faiz

TABLE OF CONTENTS

1 EXECUTIVE SUMMARY

The Asset Cortex project is an innovation proof-of-concept (PoC) developed for National Gas Transmission (NGT) to demonstrate how generative AI can be used to enhance asset data quality and coverage. Focusing on the [Redacted] site, the project aimed to address two key limitations in the existing asset register: sparse coverage and insufficient hierarchical information. These gaps constrain maintenance efficiency, automation, and the ability to support data-driven decision-making. Throughout the project, the scope shifted to prioritise identification of missing entities, instead of classification against the ISO 14224 standard.

The PoC implemented a semantic-enabled AI system that integrates data sources such as defect history and maintenance records. The system iteratively constructs a knowledge graph representing site assets and their relationships, enabling robust reasoning over data to enrich the existing asset register. The system architecture comprises three core processes: ingestion of structured datasets, extraction and analysis of unstructured data, and synthesis of missing entities through AI reasoning, supported by SME knowledge. The solution was developed within a secure Azure environment, adhering to NGT governance requirements for data security, residency, and access control.

Results demonstrate that the AI system improved asset register coverage and integrity. The best performing model identified 314 entities, relative to 426 entities present in a manual survey of the site. Importantly, 92 of the identified entities were not included in the survey validation dataset but were manually validated by DNV as likely to exist, highlighting the system's capability to discover missing assets. The system was also shown to improve data quality, providing more detailed and precise attribute values, relative to manual records. Key limitations are reliance on text-based data (excluding engineering drawings), bias towards well-documented assets, and limitations in the utility of the survey validation dataset.

The PoC demonstrated high cost-efficiency relative to manual analysis, with total model execution costs for the PoC remaining negligible. While higher-capability models currently require more expensive provisioned deployments, the cost remains small. Relative to manual analysis to produce comparable outcomes, the AI system deployment was far more cost efficient. The AI system was estimated by DNV to be at least 3 orders of magnitude more cost-effective.

DNV recommend next steps to further improve model performance. These next steps include incorporating vision models to extract data from engineering drawings, expanding SME knowledge capture through ontologies, revisiting asset classification against ISO 14224, and migrating the solution to a scalable multi-agent architecture with graph data stores. Integration into NGT's IT environment and further development of probabilistic reasoning within the knowledge graph are also identified as priorities. The flexible structure of this solution further enables roll out onto other existing sites, with St Fergus identified as a priority by NGT.

2 INTRODUCTION

Asset Cortex is an innovation initiative to demonstrate a method to enhance asset data for National Gas Transmission (NGT). At the selected test site ([Redacted]) the existing asset register coverage is sparse, containing just 41 records out of the 426 identified by a manual survey approximately 10% coverage. Furthermore, this register is limited to four hierarchy levels (plus a functional location). These factors constrain maintenance efficiency, automation, and data-driven decision-making. By leveraging generative AI grounded in expert knowledge, this project sought to demonstrate the ability of an agent to mine supporting asset data for the [Redacted] site and propose an enriched asset register, enabling improved maintenance strategies and operational insights.

This document outlines the system architecture of the Proof-of-Concept (PoC), approach to data and cyber security, the data landscape, project results, a cost-benefit analysis for the use of an AI system and proposed future work. This PoC has focussed on the [Redacted] site, validating AI-driven classification against a manually curated dataset, and documented findings to inform future scalability and integration.

2.1 Project Scope

The initial project scope of Asset Cortex was to develop a proof-of-concept generative-AI approach to classify NGT's asset data in alignment with ISO 14224. DNV were advised that existing NGT asset datasets were of variable quality. In particular, the existing asset register was known to be incomplete, with some entities absent from the records. Identifying and classifying these missing entities was included in the initial scope.

During the project, NGT advised the focus should shift to prioritise identifying missing entities, instead of classifying entities into the ISO 14224 hierarchy. This scope change was due to the sparse asset register coverage and lack of manual classification data to enable performance scoring of the AI system's skill at hierarchy classification.

This new scope focussed development to compare generated outputs against a manual site survey conducted by Premtech and the subsequently inferred physical reality, and on assessing the AI system's capability to identify missing assets and enhance attribute-level detail. Development of solutions to the initial asset classification and alignment to ISO 14224 hierarchy scope were subsequently paused, in favour of focus on asset identification and enrichment.

2.2 Concepts

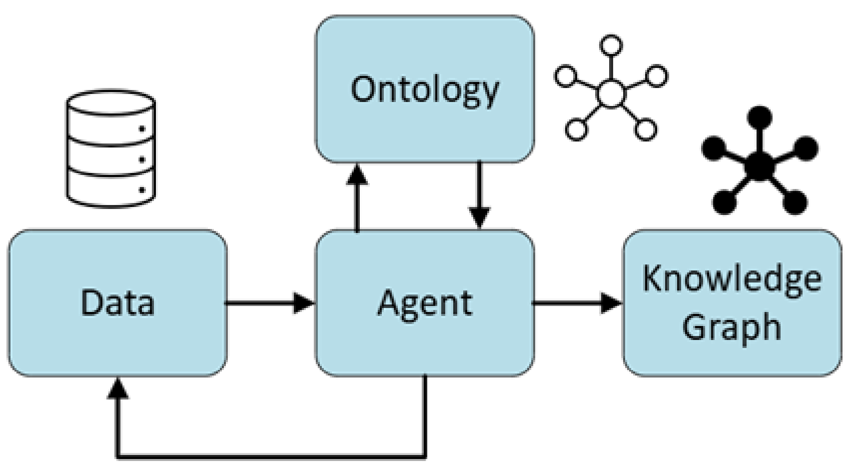


Figure 1 A high-level overview of the AI system.

The proof of concept was a semantic-enabled AI system that operated over a set of structured and unstructured data. A semantic-enabled system is one which is designed to understand the meaning of data, rather than simply process the raw data.

By defining machine-readable descriptions of both the data landscape and DNV subject matter expert (SME) knowledge, the project team enabled the AI system to reason around and identify the physical entities that may be present on the [Redacted] [Redacted] site.

The system sequentially ingested data and constructed a knowledge graph of the physical entities that may be present on the [Redacted] site.

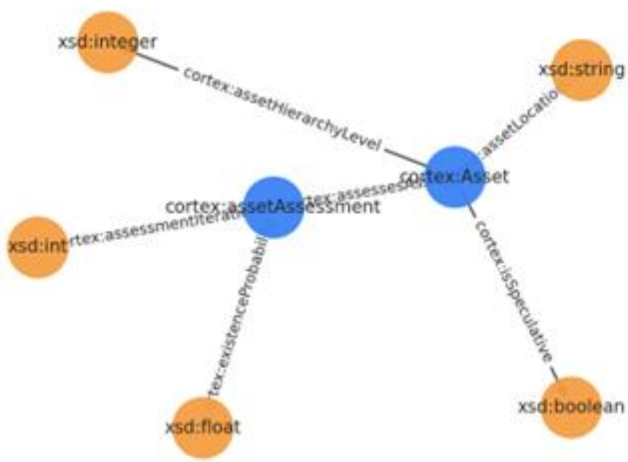


Figure 2 A small extract of the asset ontology.

An ontology is a formal, machine-readable specification of the concepts, properties, relationships, and constraints within a domain. The example given above displays an asset linking to asset attributes and their datatypes – *assetLocation*, *assetHierarchyLevel*, and *isSpeculative*. Additionally, the asset node is linked to the *assetAssessment* node – the class used to store the evidence trial for an asset existence.

SME knowledge and data schema can be encoded in machine-readable forms such as ontologies, which may support AI reasoning, or define the rules used to structure a knowledge graph. The machine-readable format used throughout this project was TTL (‘turtle’) files, storing Resource Description Framework (RDF) triples. For example, the definition of the *cortex:isSpeculative* attribute is show in Figure 3 below.

```
cortex:isSpeculative a owl:DatatypeProperty ;
    rdfs:comment "A boolean representing if the asset is assumed to be present by an agent"@en ;
    rdfs:domain cortex:Asset ;
    rdfs:range xsd:boolean .
```

Figure 3 Extract from Ontology TTL file.

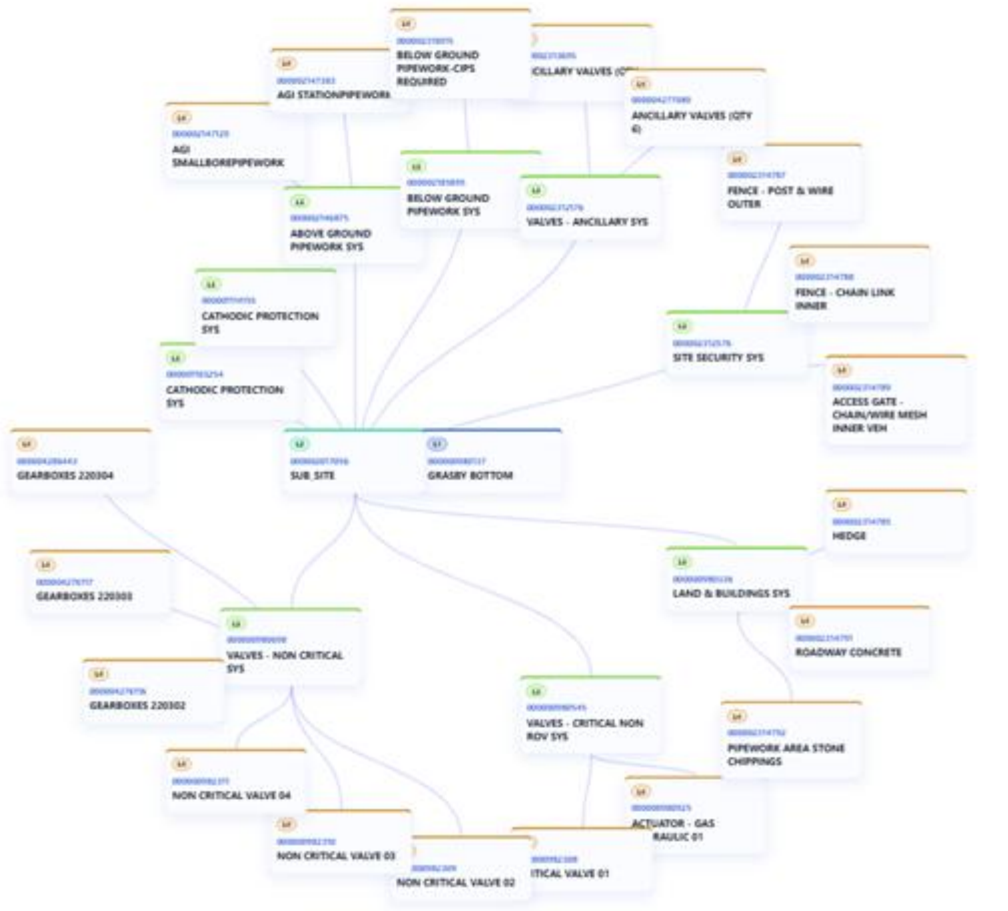


Figure 4 A knowledge graph extract of the initial asset register for the [Redacted] site.

A knowledge graph is a graph-based data representation of entities and their relationships, typically used to store and query interconnected information about a domain. An ontology can provide the schema for a knowledge graph. The example given above is the original [Redacted] asset register, pre-enrichment, with ~10% coverage.

2.3 Technical Constraints

The following constraints define the design limits and architectural choices for the Asset Cortex PoC:

Platform & Environment:

The PoC will operate within a controlled Azure environment in the DNV tenant, with the future possibility to replicate in the NGT tenant under a future Statement of Work. All data storage and processing must remain within UK/EU regions to meet residency requirements.

Model Approach:

The initial design will use a multi-prompt architecture with an optional agentic workflow. The Azure resource type “Document Intelligence” may be used to extract information from provided unstructured data. A dedicated vector database (traditional RAG) will not be provisioned for the PoC.

Technology Stack:
All code will be developed in Python and designed to integrate with Microsoft Azure services. No additional enterprise integration (e.g., SAP, Maximo) is in scope for this PoC.

Governance & Monitoring:
The solution must comply with DNV AI governance standards. National Gas' Design Authority and Gas Design Authority (GDA) will review the DNV design, using a stage gate system.

Data Retention & Disposal:
Define and implement retention and disposal policies for all project data in line with DNV governance requirements.

Semantic Layer:
A semantic model will be used, in the format of TTL files, to provide context to the AI system.

3 DATA LANDSCAPE

Identifying and ingesting high quality, relevant, datasets was essential for the successful implementation of the PoC. This section details the source and properties of each dataset ingested.

The provided unstructured datasets include engineering drawings, maintenance records, and images. Only text data is within scope for this project, vision models are explicitly excluded. Text that can be extracted from the unstructured data without a fine-tuned vision model, for example using optical character recognition (OCR), is ingested by the system. However other information, for example entities visualised in expanded diagrams, have been excluded from this project, though would likely improve the quality of the results if included in a further project.

3.1 Structured Data

The model ingests three structured datasets: the initial asset register, site defect history, and PM records. All structured datasets were provided in XLSX or CSV format.

3.1.1 Initial Asset Register

The initial asset register provided was a Maximo extract. The register contains 48 records of entities on the [Redacted] AGI site. All records contain essential information, including but not limited to: *Asset ID, Location Code, Name, Parent Location Code, Equipment Class*. The *Equipment Class* encodes the hierarchical level of the entity within the existing NGT structure. The *Location Code* and *Parent Location Code* attributes encode the parent-child relationships between entities.

Through conversations with NGT, DNV identified 18 records in the initial asset register which were created as artifacts of system maintenance. These records are identified as containing 'JOBF' or 'SYSTEM_ASSET' in place of the asset descriptions.

3.1.2 Site Defect History

A dataset of defect records was provided, covering history of defects for the [Redacted] , [Redacted] , and [Redacted] sites. The dataset contains a total of 153 records, of which 6 are applicable to [Redacted] .

Each defect record is linked to an asset within the initial asset register. NGT have advised these records are linked to 'system-level' assets at level 3 in the pre-existing NGT hierarchy, rather than the specific asset with the defect.

3.1.3 PM Jobs

An extract of PM job records (maintenance policies) has been provided, covering activities across the [Redacted] , [Redacted] [Redacted] , and [Redacted] AGI sites. The dataset contains a total of 144 records, of which 12 are identified as relevant to [Redacted] [Redacted] . DNV were advised in the later stages of the project that these records are likely to be the repeated jobs that are carried out on site, but not the specific instances of the work carried out and thus exclude potential valuable information. NGT indicated that this omitted information may be available and provided in future phases of work.

The extract includes key attributes associated with each maintenance activity, such as job status, job description/details, asset location, and both current and next scheduled completion dates. The dataset comprises 14 attributes, providing a broad view of maintenance planning and execution across the sites.

3.1.4 Non-Ingested Data

DNV were provided with other structured datasets, which were used to provide context to the project, but were not ingested by the model as they do not contain references to entities on the [Redacted] site. The datasets not ingested are:

- [Redacted] _2025117.xlsx
- EU Taxonomy v2.12 Jan 2025 Codified for use.xlsx
- EU_Mapping_input_files_EU_MAPPING_TEMPLACE_V2.77.24.xlsx
- [Redacted] _AGI_20251117.xlsx
- Maximo Data Dictionary V2.xlsx

3.2 Unstructured Data

The unstructured data provided to DNV has been classified into four groups: Material Take Offs (MTOs), maintenance records, engineering drawings, and images. As image data is excluded from the project scope, this group is not ingested by the model but instead used for human-in-the-loop validation by DNV.

3.2.1 Material Take Off

DNV were provided five MTO documents, in PDF format. All documents were “construction issue”. Each document contains an itemised list of entities with various attributes, including but not limited to *pipe size*, *pressure rating*, *standard name*, and *quantity*. Typically, the entities listed in the MTOs are at low levels of the asset hierarchy, and do not contain larger components, such as complete systems or equipment units.

As the MTOs are in PDF format, information passed to the AI system is first extracted by performing optical character recognition (OCR), retrieving the text.

3.2.2 Maintenance Records

DNV were provided maintenance records by NGT. The format of the maintenance records varies between documents. All documents contain free-text descriptions of maintenance work carried out on site. The documents are all either PDF or DOCX file types. OCR was performed on the PDF files to extract text for further processing.

3.2.3 Engineering Drawings

DNV were provided detailed engineering drawings of the [Redacted] site. Inspection of these drawings with a vision model is outside of the scope of the project, and therefore information embedded in the drawings was not extracted.

During the project, DNV investigated the capabilities available with language models. These models have poor vision capabilities, and as such engineering drawings are not ingested in the final AI system. Vision models are available via the cloud resources used throughout this project, and thus inclusion of vision models in a future project scope forms a natural next step, supporting more advanced missing entity inference.

4 SOLUTION DESIGN

This section defines the architecture of the final AI system and associated cyber security controls.

4.1 System Architecture

The system architecture was grouped into three processes: structured data ingestion, unstructured data ingestion, missing entity synthesisation.

4.1.1 Structured Data Ingestion

Sequentially ingesting the three structured datasets was a deterministic process. First the initial asset register was processed, followed by the PM records, then finally the defect records.

For each dataset, the following steps were performed:

1) **Fetch dataset from Azure Storage Blob**
   To ensure sensitive NGT data isn't stored on local machines, all information was stored in a secure Azure storage blob, which was only accessible via the DNV Network, programmatically retrieved, and removed once processing is complete.

2) **Fetch entity properties and map attributes to ontology properties**
   Some properties are common between datasets with different attribute labels, for example "LOCATION" in PM records and "Location Code" in the initial asset register. A mapping from these attributes to a common vocabulary was specified within the structured data ontology created by DNV in the second work package.

3) **Update knowledge graph with discovered information**
   Once the information was in a common format, the knowledge graph can be updated. DNV understand the location code of an asset is a unique identifier, so checking if an entity was already present in the asset register was a simple look-up.
   If an entity could be identified in the asset register, the evidence trail for the asset instance can be updated with the new observation. If the entity was not present in the asset register, a new record is created. New entities discovered from the structured data were not flagged as 'speculative', as the deterministic nature of this process gives high confidence in the discovered entities.

4.1.2 Unstructured Data Ingestion

The unstructured datasets were ingested sequentially. First all MTOs were processed, then the maintenance records. The process to ingest the data is as follows:

1) **Fetch data from Azure Storage Blob**
   To ensure sensitive NGT data isn't stored on local machines, all information as stored in a secure Azure storage blob which was only accessible via the DNV Network, programmatically retrieved, and removed once processing is complete.

2) **Extract Information with Azure Content Understanding**
   To extract the text from the PDF file types, DNV used Azure Content Understand OCR capabilities. This resource

allowed repeatable, efficient processing of text and retains some structural information, such as paragraphs and tables.

3) **Prompt model to extract entities referenced**
The extracted text was injected as context into predefined prompt text, which is submitted to a secure agent deployment in Microsoft Foundry. The prompt defined the task for the agent as extracting entity references from the document text.
The MTOs contained itemised entity references, along with several relevant attributes. In the maintenance documents, references to entities were more implicit, and the agent must infer the properties of the attributes.
For example, Asset Name was a property of all entities. The MTOs provide a standardised, explicit name, whereas this wasn't typically present in maintenance records. An example extract from an MTO document is shown below.
Following entity extraction and identification, the agent was prompted to check if an entity extracted from the data exists in the knowledge graph. The agent must be invoked as there is no unique identifier present in any unstructured data records to lookup the entity in the knowledge graph. If the entity existed, the evidence trail was updated with new information. If not, a new asset was created and flagged as 'speculative'.

ITEM NO. (1)	PIPE SIZE (2)	PRESS RATING (3)	SPEC (4)	STANDARD NAME (5)	VOCAB NO. (6)	QTY (UNIT OF MEASURE) (7)	ISSUE CHANGE REF. (7A)	B.G.C. FREE ISSUE N/Y (8)	SITE REQUIREMENT DATE (9)	W.P. W.E. NO. (10)	INSPECT REQ.D N/Y (11)	CERTS REQ.D N/Y (12)	LOCATION FOR DELIVERIES (13)	REMARKS SUGGESTED/NOMINATED SUPPLIER (STATE IF TECHNICAL EVALUATION REQ.D) (14)
50	200	ASA 600	PS/F1 SPEC	FLANGE, WELD NECK CLASS 600 R.T.J TO SUIT 8.18 LX42 PIPE	79-08-6011	2		Y						

Figure 5 An extract from MTO document 3420x3xd-01.

4.1.3 Missing Entity Synthesisation

To further enhance the asset register after ingesting all structured and unstructured data, the following steps were performed to rigorously and repeatably ingest DNV SME knowledge.

1) **Encoded SME knowledge into an ontology**
SME knowledge was translated from a natural language description of the CP system to RDF triples, which encode the hierarchical structure of the CP system and its sub-components.

2) **Agent ingests ontology and reasons around speculative assets on** [Redacted] **site**
The agent was prompted to infer which assets are missing from the knowledge graph, given existing world knowledge and the additional DNV SME knowledge as additional context.

3) **Update knowledge graph with speculative assets**
As in prior steps, the knowledge graph was updated with the required attributes, and the agent reasoning and speculation was added to the evidence trail.

4.2 Cyber Security

The cyber security controls followed throughout the project are detailed below.

4.2.1 NGT Governance

DNV has completed all governance requirements requested by NGT. Supporting documentation, including ITAR, SPIG, and Privacy Impact Assessment, was provided and subsequently reviewed with approval granted to progress through the Gas Design Authority stage gate and develop a proof-of-concept model. Productionised development will require returning to the GDA.

The project has been established within Azure Foundry to support appropriate security controls for data handling. This includes the implementation of clearly defined access controls, which follow the least privilege principle, controls over data processing and storage, restrictions on external access, code vulnerability scanning, and encryption of NGT data.

4.2.2 Access rights

To limit potential risk exposure, access to Azure AI Foundry resources was restricted to a small group consisting of Andrew Lawrie, Benjamin Ross, Gabriel Griffin Booth, and Finlay Brierton. Within Azure, the principle of least privilege was applied, ensuring that users were only granted the permissions required for their roles, with owner-level access minimised where possible.

In addition, API key access was disabled, meaning that all access to, and interaction with, the project required authentication via Microsoft Entra ID. This ensured that access remained tied to specified users only.

4.2.3 Data Storage & Processing

To ensure alignment with NGT guidance, all data provided by NGT was stored within a UK South Azure Storage Account. Where data processing occurred outside the UK, this was conducted strictly in accordance with NGT requirements and limited to UK South and Sweden Central.

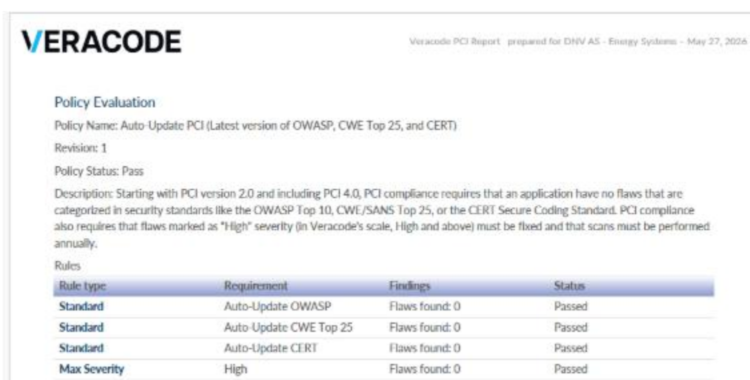
Due to the potential for internal DNV Copilot to access and utilise information stored within SharePoint, all project data was classified as DNV Confidential (Protected). This classification prevents Copilot from accessing or processing the NGT data, ensuring that sensitive information remains restricted and protected.

4.2.4 Code Security

To ensure that the code created is kept secure, automated code scanning was implemented to identify potential vulnerabilities and weaknesses. This includes the use of static application security testing (SAST) and software composition analysis (SCA) to review both developed code and third-party dependencies.

This scanning covered a range of security risks, including common coding vulnerabilities (e.g. Improper validation), known vulnerabilities in third-party components (e.g. publicly disclosed CVEs), potential exposure of sensitive information such as hard-coded credentials, and configuration issues that could weaken security.

The purpose of this testing is to support the security and integrity of data associated with both DNV and NGT by identifying and flagging issues for remediation.


VERACODE

Veracode PCI Report prepared for DNV AS - Energy Systems - May 27, 2026

Policy Evaluation

Policy Name: Auto-Update PCI (Latest version of OWASP, CWE Top 25, and CERT)

Revision: 1

Policy Status: Pass

Description: Starting with PCI version 2.0 and including PCI 4.0, PCI compliance requires that an application have no flaws that are categorized in security standards like the OWASP Top 10, CWE/SANS Top 25, or the CERT Secure Coding Standard. PCI compliance also requires that flaws marked as "High" severity (in Veracode's scale, High and above) must be fixed and that scans must be performed annually.

Rules

Rule type	Requirement	Findings	Status
Standard	Auto-Update OWASP	Flaws found: 0	Passed
Standard	Auto-Update CWE Top 25	Flaws found: 0	Passed
Standard	Auto-Update CERT	Flaws found: 0	Passed
Max Severity	High	Flaws found: 0	Passed

Figure 6 Veracode scan results for the codebase

5 RESULTS

This section outlines the AI system results and includes illustrative examples of selected entities to aid analysis and interpretation of the outputs.

Three sets were defined to assist in evaluation;

- "*Asset Cortex*" - the outputs of the AI System.
- "*Validation Data*" - 426 entities contained within a manual survey of the [Redacted] site performed by Premtech.
- "*Physical Reality*" - entities that are physically on the site.

The survey validation data was not a complete dataset of all entities physically on the site. As such the known physical reality set is the validation dataset, plus all other entities predicted by the AI system, which were manually verified by DNV as present on the [Redacted] site. This set is our best estimate of all assets on site.

5.1 Model Evaluation Criteria

Due to the change in scope during the project, the success metrics defined in the design document required change. Furthermore, the AI system identified entities which were not present in the survey validation data. Subsequently the survey validation data reports only partial coverage of all entities on site, so classic metrics, such as accuracy, do not provide complete results. Additional metrics have been provided to evaluate the performance of the model.

The AI model was evaluated on the following criteria:

- **Recall** – "How many of the survey validation data entities did the system identify?"
  The value was calculated as the number of entities correctly identified by the AI system, as a percentage of the number of assets in the survey validation data.
  The recall cannot be calculated for the physical reality set. Because no exhaustive asset register exists, the number of entities in this set is not known.
- **Precision** – "How many entities that the system identified are correct?"
  The value was calculated as the number of entities correctly identified by the AI system, as a percentage of the number of predicted assets by the AI System. Note that this figure was reported for predictions that are in 1) the survey validation data only and 2) the known physical reality.
- **Entity Discovery Rate** – How many new entities did the system discover?
  The value was calculated as the number of newly discovered entities, as a percentage of the number of entities in the survey validation data.

For the purposes of evaluation, DNV assumed all entities recorded in the survey validation data are present on site. Therefore, there are no entities in the survey validation data which were considered to be included as a result of human error.

Some entities were present in the initial asset register and survey validation data as a single record, but with a quantity defined in the record description. For evaluation, DNV have considered these to be multiple unique entities. For example, "STUDBOLT" records are consistently followed by "pack of X" in the validation data, these were considered to be X unique entities.

Throughout the evaluation process, it was consistently noted that the AI System carries bias from the datasets ingested. As many input datasets are related to maintenance or completed work procedures, the model was biased against entities which were less likely to be mentioned in these documents, notably smaller assets and welds. Through discussion with NGT, DNV identified In Line Inspection data as a potential method to mitigate this bias, though this data could not be sourced within the project timeframe.

5.2 Quantitative Results

The results are reported for two LLM architectures – GPT 5.1 and GPT 5.4. GPT 5.1 was used because it was the most capable LLM available throughout the development process, and GPT 5.4 was used as it was the most capable LLM available towards the end of the project.

The survey validation data includes 172 references to entities that DNV have classed as '*connection*' type, predominantly "welds" and "threads". Therefore, there were 254 non-connection type entities. The count of non-connection type entities has been reported separately, as non-connections were not the focus of the initial scope of the model.

5.2.1 GPT 5.1

The AI System, running the GPT 5.1 model, identified 294 entities. 167 of these entities were present in the survey validation data, 104 were entities missing from the validation data but have been manually validated as likely to physically exist or contain important hierarchy information and so are included in the Known Physical Reality data, and 23 were either incorrect hallucinations or too generic to provide value.

The evidence supporting the 104 entities that are absent from the validation data but identified by the AI system is as follows: MTOs – 35, DNV SME – 9, Initial Asset Register – 40, Maintenance Documents – 20, Defect Records – 2. Examples of the assets missing from the validation data are given in Section 5.3 Qualitative Examples.

Table 1 Evaluation metrics for agent running GPT 5.1

Metric	Value (2 sf)	Definition
Recall [0, 1], higher is better	167/254 = 0.66 (excl connections) 167/426 = 0.39 (incl connections)	How many of the survey validation entities did the model find?
Precision [0, 1], higher is better	167/294 = 0.57 (validation data) 271/294 = 0.92 (physical reality)	How many entities found are present in the survey validation data or can be manually validated as know physical reality?
Entity Discovery Rate Higher is better	104/426 = 0.24	What percentage increase in number of entities does the model facilitate?

The recall, excluding connections, value of 0.66 for the system was considered an average result. In the Next Steps section, additional steps to improve this score are suggested – for example incorporating a vision model, or additional datasets.

The precision, considering physical reality entities, value of 0.92 was considered a strong result – the system precisely identified entities present on the [Redacted] site, and had a relatively low rate of hallucination. A high precision value justifies confidence in the system going forward and suggests the underlying model successfully inferred information about the [Redacted] site.

Additionally, the entity discovery rate of 0.24 was another strong result. This value demonstrated the AI system inferred or identified a significant proportion of entities that were physically on the [Redacted] site but omitted from the validation data. This characteristic of the system provides significant value to NGT.

Finally, the hallucinated assets are typically generic references to collections of assets on site. For example, the model predicted '*Gas pipeline risers*' but couldn't infer appropriate additional details from the context in the document, as such this entity was marked as a hallucination.

5.2.2 GPT 5.4

The agent, running the GPT 5.4 model, identified 314 entities. 198 of these entities were present in the validation data, 92 were entities missing from the survey validation data but have been manually validated as likely to physically exist or contain important hierarchy information, and 24 were incorrect hallucinations or too generic to provide value.

The evidence supporting the 92 entities that are absent from the survey validation data but identified by the AI system was as follows: MTOs – 22, DNV SME – 18, Initial Asset Register – 40, Maintenance Documents – 10, Defect Records – 2. Examples of the assets missing from the survey validation data are given in Section 5.3 Qualitative Examples.

Table 2 Evaluation metrics for agent running GPT 5.4

Metric	Value (2 sf)	Definition
Recall [0, 1], higher is better	198/254 = 0.78 (excl connections) 198/426 = 0.46 (incl connections)	How many of the validation entities did the model find?
Precision [0, 1], higher is better	198/314 = 0.63 (validation data) 290/314 = 0.92 (physical reality)	How many entities found are present in the survey validation data or can be manually validated as know physical reality?
Entity Discovery Rate Higher is better	92/426 = 0.22	What percentage increase in number of entities does the model facilitate?

The recall, excluding connections, value of 0.78 for the system was considered an excellent result. This value was strong evidence to suggest the agent could infer entities from the input datasets.

The precision, considering physical reality entities, value of 0.92 was considered a strong result. This value provided evidence for the inference capability of the agent, enabling effective identification of missing entities, whilst minimising hallucinations.

The entity discovery rate of 0.22 was another strong result. This value demonstrated the AI system inferred or identified a significant proportion of entities that are physically on the [Redacted] site but omitted from the validation data.

5.3 Qualitative Examples

This section presents two examples of entities identified by the AI system, with a focus on those not present in the validation dataset, as these offer the greatest value to NGT.

5.3.1 Cathodic Protection System

When synthesising entities missing from the knowledge graph, DNV focussed on the cathodic protection (CP) system to facilitate a proof of concept, to demonstrate these missing entities can be identified by the agent. In the initial asset register, and in the validation data, there are no CP system entities, beyond the system itself. As the CP system isn't frequently referenced throughout the datasets ingested, the agent is biased against identifying entities belonging to the CP system.

After ingesting the DNV-defined ontology describing the CP system, the AI System inferred entities which may be present on the [Redacted] site. In particular, the AI System identified a test post and an AC coupon, both belonging to the CP system, not present in either the initial asset register nor the validation dataset. Based on DNV SME understanding of the [Redacted] site, an AC coupon and CP system test post must be present on the AGI, to allow NGT to follow the internal T/SP/ECP/7 standard. These two entities provide strong evidence for the AI System's ability to reason around AGI system components using a SME-defined ontology.

5.3.2 Civils

The initial asset register provided to DNV includes several civil entities, none of which are present in the survey validation data. The AI system enriched the asset register further by incorporating additional civil entities referenced in maintenance records. The agent was able to identify these entities, and update the knowledge graph accordingly.

For example, the initial asset register includes a record of the "Fence - Chain Link Inner", this asset is not contained within the validation data. The initial asset register doesn't reference the 'Anti-Burrowing Mesh' which is referenced in the maintenance records. The agent was able to identify this entity and update the knowledge graph accordingly with the missing entity for the asset register. This proposal of physical security measures existing at the site, missing from manual survey validation data, would enhance the ability for associated security teams within NGT to understand the risk profile of their sites.

The agent also identified sandbox installations in the maintenance records. The extract from the maintenance records note that these installations are new, not replacing existing assets, and as such may not have been included in the initial asset register. The new sandbox installations are included in the knowledge graph, enriching the asset register.

5.3.3 Plugs

The GPT 5.4 agent proposed the existence of 7 "plugs" that were not observed in the survey validation data. Upon manual verification, 1" and 1 ½" carbon steel plugs can be seen in MTO documents. The DNV Gas Technical SME has advised that these may be omitted by a human surveyor if they're considered to be part of a larger valve assembly. However, these components have the potential to cause gas leaks if they corrode.

5.4 Evaluation

Identification of Missing Entities - A major strength of the AI system is the ability to consistently identify entities missing from the validation dataset. Section 5.3 Qualitative Examples reports examples of these entities, highlighting both Cathodic Protection items required for compliance with T/SP/ECP/7, alongside physical security measures of high value to understanding the risk profile of these sites. These entities are identified from all sources of data. DNV have been able to manually validate that these entities extracted are likely present on the [Redacted] site. Identifying entities absent from both the survey validation data and the initial asset register, will enable NGT to optimise data-driven decision making, comply with Ofgem and ISO standards, and improve operational performance.

AI system increases integrity of asset register - It was observed that the entities present in the validation data may be described with higher accuracy in the AI system predictions. For example, there are 23 valves present in the survey validation data. The survey validation data includes an attribute to flag critical valves *'G2045_CRITICAL VALVE*'. The survey validation data marks all 23 valves as non-critical. However, the initial register identifies a critical valve on the [Redacted] site, which has been validated by DNV SME. The AI system therefore provides an effective tool to enhance the integrity of the final asset register, flagging any inconsistencies between datasets.

A further example of the enhancement of the asset register is the precision of reporting of diameters between the AI system outputs, and the survey validation data. It was empirically observed that the AI system reports numerical values to an extra decimal place, when compared to the manual survey. These values are not hallucinations – they are pulled directly from the evidence trail, typically the MTO document.

GPT 5.4 outperforms GPT 5.1 – The evaluation metrics of the GPT 5.4 model were consistently higher than those of the GPT 5.1 model, except for Entity Discovery Rate (0.22 vs 0.24). This higher entity discovery rate was attributed to the GPT 5.1 model providing more generic, but still valuable, entity predictions. For example, the GPT 5.1 model identified a "*set of four vent line valves*", which are omitted from the GPT 5.4 outputs.

The higher precision and recall of the GPT 5.4 model was not unexpected, as this model is widely considered more powerful, with more advanced reasoning and inference capabilities. Notably, DNV observed that the GPT 5.4 model handled entity persistence and duplication more accurately. When the agent was prompted to decide if a specific entity already existed in the knowledge graph, the GPT 5.4 model distinguished more accurately, which improved both the precision and recall of the agent outputs.

Validation Dataset Limitations – Importantly, the results presented in this section do not provide a fully comprehensive overview of the model performance. Model performance cannot be completely understood as the survey validation dataset has several limitations that constrain its validity as the physical reality. Notably, the survey validation data does not provide complete coverage of all assets present on site, meaning some valid entities identified by the AI system are absent from the dataset.

Additionally, the survey validation data lacked hierarchical structure aligned to standards such as ISO 14224, preventing direct comparison of asset relationships or classification depth. Many assets are recorded in an aggregated format (e.g. quantities embedded within a single record), or do not contain any unique identifying information (e.g. all records of "*threads*" are identical). Subsequently, many predicted entities cannot be identified on a one-to-one basis against the validation data.

6 COST BENEFITS ANALYSIS

The use of an AI agent in this scenario enabled DNV to utilise a system to rapidly query structured and unstructured data using mapped semantic connections between datasets. Whilst an initial work package was needed to begin to map the links between supplied NGT datasets, this is a reusable activity. The produced ontology, whilst not a complete data model, can be built upon over time. This is because the data storage systems in NGT will only change slowly over time, so the concepts extracted will not need to be fully reworked in every project.

The main ongoing cost is the AI tokens used when prompts are sent to the Large Language Model, or the provisioned model when a token-based model was not available in the UK data zone, to comply with NGT data residency requirements. These costs are defined in Table 3.

The use of the AI agent reduces the level of manual effort required to review, interpret and correlate large volumes of data. Manual analysis would require significant human resource to perform document review and cross-reference activities, particularly across unstructured formats such as PDFS and images. The AI-assisted approach enables these activities to be completed more efficiently, with human oversight focused on validation rather than initial data processing. Estimates from DNV Gas Technical SMEs familiar with Feeder 28 and St. Fergus data are provided, to demonstrate the scale of additional manual effort required.

As a result, even at higher usage levels, the cost of AI processing remains lower than the equivalent cost of sustained manual analysis, with the primary benefit being a reduction in time and resource required to perform data inspection and interpretation.

The design of this proof-of-concept system was to be sufficient to demonstrate the concept had value to NGT. Some additional work will be required to productionise the logic used into a robust replicable agent-based framework to cover multiple sites. These next steps are detailed in Section 7 Next Steps.

6.1 Model Type Overview

Azure Foundry offers two broad model categories: base models and fine-tuned models. Fine-tuned models are out of scope for this assessment. For base models, capacity can be provisioned in two ways: Standard and Provisioned Throughput. Standard deployments rely on shared Azure cloud capacity, whereas Provisioned Throughput models reserve dedicated model capacity. The reserved resource capacity is given in Provisioned Throughput Units (PTUs). The two methods of provisioning utilised are summarised below:

On-Demand Foundry AI Models: Standard (on-demand) models are charged on a pay-per-token basis, with no fixed hourly reservation cost. These are suitable for general workloads; however performance will vary depending on demand. Due to UK Zone data requirements, the latest GPT models were not available with this type of model deployment.

Provisioned Throughput Models: Provisioned throughput models charge a fixed hourly rate for a specific reserved capacity (given in PTUs), regardless of usage. These deployments enabled access to newer GPT models (e.g. GPT-5.4), and provided predictable performance and latency, even during periods of high demand.

Table 3 Cost Comparison of example model deployments.

Deployment Type	Model Type	Cost per Hour	Minimum PTU
Regional	Throughput	$100	50 PTU
Data Zone	Throughput	$16.50	15 PTU
Standard	Throughput	$15	15 PTU
Data Zone	Standard (on-demand)	Pay-per token	None
Standard	Standard (on-demand)	Pay-per token	None

6.2 Cost Implications

During the proof of concept, use of the GPT-5.1 standard regional (pay-per-token) deployment resulted in a total cost of less than 100 NOK (<£8) for the full processing workload. In contrast, the equivalent capability through a provisioned throughput deployment would incur a cost of up to $100 (~£74) per hour based on regional pricing.

Between these options, regional throughput is significantly more expensive due to its higher hourly rate and larger minimum capacity requirements. Data zone and standard throughput options reduce this cost substantially, while still enabling access to higher-capability models through the reserved capacity. For this project, only regional UK or Sweden Central regional models meet NGT requirements. The alternative option is non-throughput (standard on-demand) deployments which operate on a pay-per-token basis. These deployments are a more cost-effective solution but were not available for any model above GPT-5.1.

Manual data extraction for the uprating of Feeder 28 provides a useful narrative to understand the costs of manually scaling data extraction (as opposed to the use of this Agentic AI system.) The methodology applied to Feeder 28 uprating was to identify only those assets directly associated with the safe containment of gas, as well as grouping multiple iterations of identical components. Therefore, of the 342 [Redacted] items identified in the manual site survey, this method would have selected approximately 1/7th of the total, ~50. The Feeder 28 uprating process identified 1497 items (i.e. 30x uplift of [Redacted] at the same item selection criteria, or scaled up to match manual survey coverage, ~10,500 items, with Llanwrda, an analogous site to [Redacted] , representing 5% of records) where the Data Recovery Task ran to a total cost of £130k, plus an additional £30k data quality assessment. The feeder and its associated sites were constructed in 2007/2008 and were considered to have excellent data availability.

Gas Technical SMEs were also consulted regarding the St. Fergus site data but highlight that data completeness and quality at this site is considerably lower than experienced at Feeder 28, with additional complex underground infrastructure. Whilst estimates of item counts are at ~4,000 gas facing components (scaling up to ~30,000 items to match survey criteria), the effort to establish which version of reference data is appropriate for St. Fergus based on experience of working with this site's data is considered to take significantly more time and cost per item. SME order of magnitude estimates of costs range in the £1million-£10million for a manual data extraction and validation activity.

The use of this AI system does not provide perfect coverage, as detailed in Results, and does require human-in-the-loop verification. However, it does provide an initial view of items on site, alongside identifying items that are missing from the manual survey such as Cathodic Protection tokens, physical security measures, and steel plugs. Token costs were <£10 for a single run of the agent relating to the [Redacted] site. Scaling this up to larger sites will require a more complex architecture that was outside the scope of this PoC (Vision Models, mature RAG approach) and further iterative loops, with specialised agents working in coordination. However, even scaling token counts by several orders of magnitude to represent the additional complexity of the site and system required, these represent token costs in the hundreds to thousands of pounds, rather than millions as anticipated for a manual data process.

7 NEXT STEPS

DNV have compiled a list of proposed next steps. These next steps build on capabilities and biases identified in the PoC model and incorporate features to increase AI system robustness.

Vision Model - A vision-based model could be introduced to extract symbols and features directly from engineering drawings, enabling identification of assets that are not explicitly labelled in text. This inclusion would complement the existing text-based pipeline and reduce bias toward well-documented assets, improving coverage where data is sparse or inconsistently labelled.
For example, in engineering drawing *3420x3x37*, 23 valves were aidentified, with several attributes – for example *manual / actuated*, *open / closed*. However, the valves are not identified by text in this engineering drawing, only visual data (see Figure 7 An extract from engineering drawing 3420x3x37 below).

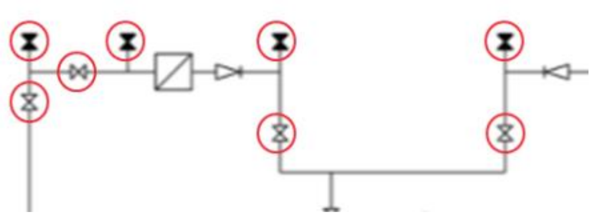


Figure 7 An extract from engineering drawing 3420x3x37

Capture Additional SME Knowledge - The initial use of this technique proved the concept was useful for predicting absent items from the Cathodic Protection system. Capturing additional subject matter expert (SME) knowledge in structured, machine-readable ontologies would allow the agent to incorporate further domain expertise more consistently, improving classification accuracy and supporting more reliable interpretation of ambiguous or incomplete asset data.

Refocus on ISO14224 classifier - Future work could revisit classification of entities into the ISO 14224 taxonomy, as in the original scope, rather than focusing on identifying missing entities. DNV have been advised validation data for this classification isn't available, so methods to work around this challenge would need to be defined.

Migrate code logic into multi-agent framework and utilise Azure graph data stores – The focus of this PoC was to demonstrate the methodology rather than building a production ready system. Code artifacts produced were designed to run for the [Redacted] site, using ontology and knowledge graphs stored in TTL files. A productionised RAG style architecture using graph data stores was explicitly out of scope. A multi agent framework within Azure AI Foundry, with the graph data stores to enable efficient queries for agent context, is required to scale this system to larger sites.

Migrate Cloud Resources to GCP – To facilitate integration into NGT IT estate, the AI system will need to be migrated. NGT has identified Google Cloud Platform (GCP) as its strategic cloud solution. Most components of the solution are portable, for example the Python scripts and Ontology TTL files. However, if the use of NGT's existing Azure estate is out of bounds, non-portable resources will need to be identified and a suitable solution defined. Example non-portable resources are Microsoft AI Foundry, Azure Storage, and Azure Content Understanding.

Further Development of Probabilistic Knowledge Graph – Many assets were referenced only once between datasets, meaning small, nuanced updates did not influence the final results. As a result, probabilistic updates to the knowledge graph were not a focus of development. If further stages of the project ingest far larger datasets, for example at a larger AGI site, more rigorous probabilistic updates to the knowledge graph are a desirable next step.

Ingest Additional Available Datasets – Specific maintenance records and ILI data were identified as likely sources of valuable entity references. These datasets could not be sourced during the project, so are included as a next step for potentially "low hanging fruit".

About DNV

DNV is the independent expert in risk management and assurance, operating in more than 100 countries. Through its broad experience and deep expertise DNV advances safety and sustainable performance, sets industry benchmarks, and inspires and invents solutions.

Whether assessing a new ship design, optimizing the performance of a wind farm, analyzing sensor data from a gas pipeline or certifying a food company's supply chain, DNV enables its customers and their stakeholders to make critical decisions with confidence.

Driven by its purpose, to safeguard life, property, and the environment, DNV helps tackle the challenges and global transformations facing its customers and the world today and is a trusted voice for many of the world's most successful and forward-thinking companies.