

G-Cloud 15

Service Definition Document

Lot 3: Cloud Support

Generative AI Platform Implementation (AWS & Azure)

Provider:	Zartis
Framework:	G-Cloud 15 – Lot 3: Cloud Support
Document Version:	1.0 – January 2026
Contact:	ricky.hill@zartis.com

Table of Contents

1. Executive Summary	3
2. Service Overview	4
3. Delivery Methodology	6
4. Team Structure and Roles	9
5. Case Studies	10
6. Tools and Technologies	11
7. Service Capabilities	11
8. Pricing	12
9. Support Arrangements	13
10. Security and Compliance	14
11. Engagement Governance	15
12. Social Value	16
13. About Zartis	17

1. Executive Summary

Zartis provides Generative AI Platform Implementation (AWS & Azure) to public sector organisations through the G-Cloud 15 framework. As an Anthropic Partner, we bring expertise in responsible generative AI implementation. We help organisations harness the potential of large language models and generative AI while establishing appropriate governance frameworks and ensuring safe, ethical deployment.

Our service is delivered by experienced consultants who combine deep technical expertise with proven delivery experience across regulated industries. We hold ISO 27001 and Cyber Essentials certifications, and our methodologies align with NCSC Cloud Security Principles.

Key service highlights:

- Proven delivery methodology with defined phases and deliverables
- Experienced team with regulated industry track record
- Flexible engagement models including Time & Materials and fixed-price options
- Comprehensive knowledge transfer ensuring client self-sufficiency
- ISO 27001 and Cyber Essentials certified delivery

This document provides a complete description of the service, including our delivery methodology, team structure, case studies demonstrating relevant experience, support arrangements, and governance framework.

2. Service Overview

2.1 Service Description

Zartis provides Generative AI Platform Implementation (AWS & Azure) to public sector organisations seeking to harness the potential of generative AI and large language models. As an Anthropic Partner, we bring expertise in responsible AI implementation and the specific considerations of deploying generative AI in public sector contexts.

We help organisations implement generative AI platforms on AWS and Azure, establishing the infrastructure, integration patterns, and governance frameworks needed for safe, effective deployment. Our consultants guide teams through use case identification, platform selection, and the responsible AI practices essential for public sector adoption.

The service is designed for organisations exploring generative AI opportunities while managing risk appropriately. We help teams move from experimentation to controlled production deployment.

Our consultants bring experience across multiple sectors including healthcare, financial services, education, utilities, and logistics. This cross-sector experience enables us to apply proven patterns and avoid common pitfalls, while respecting the specific regulatory and operational requirements of each industry.

2.2 Target Audience

This service is designed for:

- Central government departments
- Local authorities and councils
- NHS trusts and healthcare organisations
- Educational institutions
- Arm's length bodies and agencies
- Any public sector organisation seeking cloud capabilities

2.3 Key Features

The service encompasses the following capabilities:

AWS Bedrock and Azure OpenAI platform setup

This capability is delivered by our experienced consultants as part of the overall service engagement. Our approach ensures alignment with industry best practices and client-specific requirements.

Claude, GPT, and other foundation model integration

This capability is delivered by our experienced consultants as part of the overall service engagement. Our approach ensures alignment with industry best practices and client-specific requirements.

RAG (Retrieval Augmented Generation) architecture design

This capability is delivered by our experienced consultants as part of the overall service engagement. Our approach ensures alignment with industry best practices and client-specific requirements.

Vector database implementation (OpenSearch, Pinecone)

This capability is delivered by our experienced consultants as part of the overall service engagement. Our approach ensures alignment with industry best practices and client-specific requirements.

Prompt engineering and optimisation

This capability is delivered by our experienced consultants as part of the overall service engagement. Our approach ensures alignment with industry best practices and client-specific requirements.

AI guardrails and content safety configuration

This capability is delivered by our experienced consultants as part of the overall service engagement. Our approach ensures alignment with industry best practices and client-specific requirements.

API integration and application embedding

This capability is delivered by our experienced consultants as part of the overall service engagement. Our approach ensures alignment with industry best practices and client-specific requirements.

Responsible AI governance and monitoring

This capability is delivered by our experienced consultants as part of the overall service engagement. Our approach ensures alignment with industry best practices and client-specific requirements.

2.4 Benefits

Organisations engaging this service can expect:

Access to latest foundation models including Claude

Organisations engaging this service can expect to realise this benefit through our proven delivery approach and experienced team.

Production-ready generative AI infrastructure

Organisations engaging this service can expect to realise this benefit through our proven delivery approach and experienced team.

Secure, compliant AI deployment for public sector

Organisations engaging this service can expect to realise this benefit through our proven delivery approach and experienced team.

Reduced hallucination risk through RAG architecture

Organisations engaging this service can expect to realise this benefit through our proven delivery approach and experienced team.

Built-in safety controls and governance

Organisations engaging this service can expect to realise this benefit through our proven delivery approach and experienced team.

Anthropic Partner expertise in responsible AI

Organisations engaging this service can expect to realise this benefit through our proven delivery approach and experienced team.

3. Delivery Methodology

3.1 Overview

Our delivery follows a structured generative AI methodology designed for responsible and effective AI implementations. This approach ensures appropriate guardrails, accuracy, and user value.

3.2 Phase 1: Use Case Discovery

The discovery phase identifies AI opportunities and establishes governance requirements.

Activities

- AI use case identification and value assessment
- Data source inventory for knowledge base/RAG
- Compliance and governance requirements identification
- Risk assessment and guardrail requirements definition
- Success criteria and accuracy target definition

Gate Criteria

Phase 1 concludes with the AI Use Case Assessment presented for approval.

3.3 Phase 2: Architecture and Design

The design phase creates the AI platform architecture including RAG pipelines and safety controls.

Activities

- Foundation model selection and hosting architecture
- RAG architecture and vector database design
- Document ingestion pipeline design
- Prompt engineering and template design
- Guardrails and content safety control design

Gate Criteria

Phase 2 concludes with architecture and guardrail design approval before implementation.

3.4 Phase 3: Implementation

The implementation phase builds the AI platform and establishes safety controls.

Activities

- AI platform infrastructure provisioning
- RAG pipeline and vector database implementation
- Document processing and embedding pipeline development
- Prompt templates and guardrails implementation
- User interface and integration development

Gate Criteria

Implementation includes iterative accuracy testing. Phase completion requires guardrail validation and accuracy acceptance.

3.5 Phase 4: Validation and Handover

The validation phase ensures AI accuracy and responsible deployment.

Activities

- Accuracy and hallucination testing
- Guardrail and safety control validation

- User acceptance testing and feedback incorporation
- Knowledge transfer on AI operations and prompt management
- Documentation and operational handover

Gate Criteria

Phase 4 concludes with accuracy sign-off, safety validation, and operational handover acceptance.

3.6 Deliverables

The engagement produces the following tangible deliverables:

AI Platform Documentation and Operations

- **AI Platform and RAG Architecture Diagrams:** High-level and granular technical blueprints of the "as-built" Generative AI system, including RAG data flows and model hosting.
- **AI Model Operation and Monitoring Runbooks:** Comprehensive guides for daily system management, troubleshooting, and monitoring of model performance, drift, and data pipelines.
- **Responsible AI Governance Handover:** Documentation and formal transition kit for responsible AI governance, safety controls, and continuous prompt management to the client's internal support teams.

Core Technical Assets

- **Codified Platform Infrastructure:** Production-ready IaC templates (Terraform, CloudFormation, or Bicep) for deploying AWS Bedrock, Azure OpenAI, and associated vector databases.
- **AI Pipelines and Automation:** Fully functional CI/CD pipelines for streamlined deployment of LLM configurations, RAG indexing processes, and application integration layers.
- **AI Safety and Observability Dashboards:** Custom-built monitoring dashboards to track system health, performance, token usage, and critically, AI safety events and guardrail breaches.
- **Prompt Engineering Artifacts:** Library of optimised prompt templates and orchestration logic to maximize model accuracy and relevance for defined use cases.
- **Content Safety and Guardrail Configuration:** Production configuration files and policies for platform-level AI guardrails and content filtering.

4. Team Structure and Roles

4.1 Engagement Team

Engagements are staffed with experienced consultants in defined roles. Team composition scales based on engagement size and complexity.

AI Engineer

Implements technical solutions and activities. Expertise in relevant platforms and tooling. Executes implementation plans and resolves technical issues.

Solution Architect

Leads technical design and ensures architectural decisions align with business requirements and best practices. Responsible for technical quality assurance.

Cloud Engineer

Implements technical solutions and activities. Expertise in relevant platforms and tooling. Executes implementation plans and resolves technical issues.

DevOps Engineer

Implements technical solutions and activities. Expertise in relevant platforms and tooling. Executes implementation plans and resolves technical issues.

AI Governance Specialist

Provides expertise and delivery capability for the engagement.

4.2 Client Team Requirements

Successful delivery requires client participation. We typically request:

- Executive Sponsor with authority to make decisions and remove blockers
- Technical Lead with knowledge of current systems
- Subject Matter Experts for relevant business areas
- Operations representatives who will support systems post-implementation
- Security and compliance representatives for reviews and approvals

5. Case Studies

5.1 Financial Services: Azure OpenAI Platform

Client Context

A global financial services provider required implementation of an AI-powered assistant platform leveraging generative AI capabilities while meeting compliance requirements.

Challenge

The client needed to deploy Azure OpenAI capabilities in a secure, compliant manner, integrate with existing business systems, and enable business users to benefit from AI assistance in their workflows.

Our Approach

Zartis engineers designed and implemented the Fexco Smart Assist platform using Azure OpenAI services. We established secure integration patterns with existing systems, implemented appropriate data handling controls meeting financial services requirements, and developed user-friendly interfaces.

Outcomes

- Azure OpenAI platform successfully deployed
- Secure integration with existing systems established
- Financial services compliance requirements validated
- Platform enabling AI-assisted workflows across business units

5.2 Enterprise AI: Knowledge Interface Platform

Client Context

An AI consulting company serving enterprises in financial services, healthcare, and insurance required development of AI-powered knowledge interfaces and workflow automation.

Challenge

The client needed to develop enterprise AI platforms that connect directly to existing data systems without complex migrations, enabling rapid deployment of intelligent automation for world-leading enterprises.

Our Approach

Zartis engineers contributed to the development of AI-powered knowledge interfaces and workflow automation capabilities. We implemented integration patterns enabling

direct connection to client data systems while bypassing complex data migration requirements.

Outcomes

- AI-powered knowledge interfaces deployed for enterprise clients
- Workflow automation capabilities enabling operational efficiency
- Integration patterns established for rapid deployment
- Platform serving clients across financial services and healthcare

6. Tools and Technologies

We utilise industry-standard tools and generative AI platforms native services:

We implement and manage high-performance vector databases including Amazon OpenSearch, Pinecone, and pgvector. These are selected based on the client's existing cloud ecosystem and the required scale for document indexing and retrieval.

7. Service Capabilities

This section provides detailed answers to common questions about our service capabilities, technical approach, and delivery practices.

What foundation models have you deployed? (Claude, GPT-4, Llama, etc.)

We have extensive experience deploying a variety of foundation models, specifically Anthropic's Claude family and OpenAI's GPT-4. We also support the integration of open-source models like Llama through managed cloud services to suit specific performance or privacy requirements.

Describe your RAG implementation experience and approach

Our approach to Retrieval-Augmented Generation (RAG) centers on reducing hallucination risks by grounding model responses in private, authoritative data. We design architectures that integrate document ingestion pipelines with vector databases to provide the model with real-time, contextually relevant information before generating a response.

What vector databases do you use? (OpenSearch, Pinecone, Weaviate, pgvector)

We implement and manage high-performance vector databases including Amazon OpenSearch, Pinecone, and pgvector. These are selected based on the client's existing cloud ecosystem and the required scale for document indexing and retrieval.

What AI safety/guardrails do you implement?

We configure robust AI guardrails and content safety controls within the platform to prevent the generation of harmful or inappropriate content. This includes implementing built-in cloud safety features and custom filters to ensure the AI operates within defined ethical and organizational boundaries.

How do you approach prompt engineering?

We utilize iterative prompt engineering and optimization to improve the accuracy and relevance of AI outputs. Our process involves structured testing of different prompt templates and techniques to ensure the model consistently meets user requirements while minimizing token usage and latency.

Describe your Anthropic partnership and Claude expertise in detail

Zartis possesses specialized expertise in the Claude model family. We focus on responsible AI deployment, leveraging Claude's high reasoning capabilities and safety features to build enterprise applications that require advanced document processing and reliable output

8. Pricing

Services are priced on a Time and Materials basis using SFIA-aligned day rates. Fixed-price engagements are available for well-defined projects with clear scope and acceptance criteria. All rates are quoted in GBP excluding VAT.

Please refer to the separate Pricing Document for the full rate card and engagement pricing details.

9. Support Arrangements

9.1 Delivery Model

Services are delivered remotely by UK and EU based teams. Our consultants follow ISO 27001 standards and connect to client environments via secure VPN with multi-factor authentication.

9.2 Working Hours

Standard delivery is conducted during UK business hours: 09:00 to 17:30 GMT/BST, Monday to Friday, excluding UK public holidays. Extended support including 24/7 coverage is available by prior agreement subject to commercial terms.

9.3 Support Channels

Channel	Availability	Use Case
Email	Yes – included	General queries, non-urgent issues, documentation requests
Ticketing System	Yes – included	Issue tracking, change requests, formal support requests
Phone	No	–
Web Chat	No	–

9.4 Support Levels

Standard support includes Priority 1–5 incident management with response times from 1 business hour (P1 Critical) to 5 business days (P5 Service Requests). Enhanced support with accelerated response times and dedicated resources available at additional cost. Support operates via monthly retainer; unused hours roll over monthly but forfeit at engagement end.

9.5 Hypercare

Following go-live, we provide a standard hypercare period. During hypercare, the delivery team remains engaged to monitor systems, resolve issues, optimise performance, complete knowledge transfer, and support transition to client operations.

10. Security and Compliance

10.1 Our Certifications

Zartis maintains the following certifications relevant to secure service delivery:

Certification	Status	Reference
ISO 27001:2013	Certified	34/5700/19/10070
Cyber Essentials	Certified	7f243da7-4b18-402f-973d-90fe65086dc2
Microsoft Solutions Partner	Active	Data & AI, Infrastructure, Digital & App Innovation

10.2 Compliance Frameworks

Our delivery approach aligns with:

- NCSC Cloud Security Principles
- ISO 27001 Information Security Management
- Cyber Essentials
- UK GDPR and Data Protection Act 2018

10.3 Data Handling

During delivery:

- All client data remains within client-controlled environments
- We do not extract or retain client data beyond engagement requirements
- Any data used for testing is anonymised or synthetic
- Data handling procedures are documented and agreed at engagement start
- Data is encrypted in transit and at rest

10.4 Insurance

Zartis maintains appropriate insurance coverage:

Insurance Type	Cover Level	Provider
----------------	-------------	----------

Professional Indemnity	£10,000,000	Hiscox SA
Public Liability	£10,000,000	Hiscox SA
Employer's Liability	£13,000,000	Hiscox SA
Cyber Insurance	£5,000,000	Hiscox SA

11. Engagement Governance

11.1 Onboarding

New engagements begin with a structured onboarding process:

- Kick-off meeting to confirm scope, objectives, and ways of working
- Access provisioning for required systems and environments
- Introduction to key stakeholders on both sides
- Communication channels and meeting cadence established
- Project governance framework agreed

11.2 Governance and Reporting

Throughout delivery, we maintain governance through:

Weekly Status Reporting

Written status reports covering progress against plan, risks and issues, upcoming activities, and decisions required.

Steering Committee

Regular steering meetings with senior stakeholders to review progress, address escalated issues, and make strategic decisions.

Risk Management

RAID (Risks, Assumptions, Issues, Dependencies) log maintained throughout delivery. Risks are assessed for probability and impact, with mitigation actions tracked.

11.3 Change Control

Changes to agreed scope are managed through a formal Change Request process:

- Change Requests documented with description, rationale, and impact
- Impact assessment covering timeline, cost, risk, and dependencies

- Client approval required before change implementation
- Changes tracked and reflected in project documentation

11.4 Acceptance and Sign-Off

Deliverables are accepted through formal sign-off at defined milestones:

- Phase gate reviews with documented acceptance criteria
- Technical deliverables validated against requirements
- Sign-off required before proceeding to next phase
- Final acceptance upon successful hypercare completion

11.5 Exit and Termination

Engagement conclusion includes:

- Knowledge transfer to client teams or successor provider
- Documentation handover
- Access revocation and security credential rotation
- Project closure meeting and lessons learned

12. Social Value

Zartis is committed to delivering social value through our G-Cloud engagements. Our approach addresses the five social value themes defined in the Social Value Model.

12.1 Tackling Climate Change and Reducing Waste

Cloud adoption inherently supports environmental sustainability by enabling organisations to leverage hyperscale providers with strong renewable energy commitments.

Our delivery model further reduces environmental impact through:

- Remote-first delivery reducing travel-related emissions
- Digital-first documentation and collaboration eliminating paper waste
- Optimisation services reducing compute resource consumption
- Guidance on sustainable architecture patterns that minimise energy usage

12.2 Creating Jobs, Skills and Growth

We contribute to skills development and employment through:

- Knowledge transfer programmes that build client team capabilities
- Training and coaching as part of engagement delivery

12.3 Tackling Economic Inequality

As an SME ourselves, we understand the importance of supporting smaller businesses:

- We actively engage SME and VCSE subcontractors where appropriate
- We support local economies through distributed team employment across UK and Europe
- Our services help public sector organisations deliver more efficient services to citizens

12.4 Equal Opportunity

Zartis is committed to diversity and inclusion:

- We maintain inclusive hiring practices and actively work to increase diversity in technology roles
- Our remote-first model enables participation from individuals who may face barriers to traditional office-based work
- We provide equal opportunities regardless of background, disability, or personal circumstances

12.5 Wellbeing

We prioritise wellbeing:

- Flexible working arrangements supporting work-life balance
- Mental health support and employee assistance programmes
- Public sector digital transformation enables better citizen access to essential services

13. About Zartis

13.1 Company Overview

Zartis is a technology consultancy established in 2009, with over 280 professionals delivering cloud transformation, software development, and digital services to

organisations across Europe and North America. We combine deep technical expertise with practical delivery experience to help organisations achieve their technology objectives.

Zartis UK Limited is our UK-registered entity (Company Number: 11207666), through which we deliver services to UK clients.

13.2 Our Expertise

We have particular depth in:

- Cloud platforms: AWS, Microsoft Azure
- Cloud-native development: Kubernetes, containers, microservices
- Data platforms: Data engineering, analytics, machine learning
- DevOps and platform engineering: CI/CD, Infrastructure as Code, SRE
- Application development: Full-stack, mobile, API development

13.3 UK Client Experience

We serve a range of UK-based clients including:

- Kaluza (OVO Group) – Energy technology platform development
- Kooth – NHS-partnered digital mental health services
- SMS Group – Smart metering and energy services
- EcoOnline – Environmental health and safety software
- Nourish Care – Digital care planning software
- EML Payments – Payments technology platform

13.4 Contact Information

Legal Entity	Zartis UK Limited
Company Number	11207666
Registered Address	128 City Road, London, EC1V 2NX
Website	www.zartis.com
G-Cloud Enquiries	ricky.hill@zartis.com

We welcome the opportunity to discuss how Zartis can support your requirements.
Please contact us to arrange an initial consultation.