

AI AGENTS VALIDATION FRAMEWORK & CONTINUOUS MONITORING

Service Definition
G-Cloud 15

REPLY

38 Grosvenor Gardens - London SW1W 0EB - United Kingdom - tel +44 (0) 207 730 6000 - fax +44 (0) 207 259 8600
info@reply.com

www.reply.com



TABLE of CONTENTS

SERVICE DEFINITION	1
TABLE OF CONTENTS	2
SERVICE DESCRIPTION	3
ONBOARDING	6
KICK-OFF MEETING	6
ACCOUNT MANAGEMENT	6
<i>DESIGNATED ACCOUNT MANAGER</i>	6
<i>FORTNIGHTLY REVIEW MEETING</i>	6
OFFBOARDING	6
ACCESS TO DATA	6
SECURITY	7
PERSONNEL SECURITY	7
INFORMATION SECURITY	7
PHYSICAL SECURITY	7
TRAINING & KNOWLEDGE TRANSFER	8
ORDERING AND INVOICING PROCESS	9
CUSTOMER RESPONSIBILITIES	10
TECHNICAL AND CLIENT-SIDE REQUIREMENTS	10
OUTCOMES AND DELIVERABLES	10
ABOUT REPLY	11



SERVICE DESCRIPTION

The AI Agents Validation Framework & Continuous Monitoring service provides a comprehensive approach to ensuring the reliability, safety, and compliance of AI agents throughout their lifecycle.

Leveraging advanced quality assurance methodologies, this service encompasses the design and implementation of robust validation frameworks tailored to the unique requirements of AI-driven systems. Our solution includes automated and manual testing strategies to assess AI agent performance, accuracy, fairness, and ethical alignment.

Continuous monitoring mechanisms are integrated to track AI agent behaviour in real-time, enabling proactive detection of anomalies, drifts, or deviations from expected outcomes.

This ongoing oversight ensures that AI agents remain effective and trustworthy as they evolve and interact with dynamic environments. Our service also supports regulatory compliance and auditability, providing detailed reporting and traceability for all validation and monitoring activities.

By partnering with Reply UK, clients benefit from industry-leading expertise in AI quality assurance, minimising operational risks and maximising the value delivered by AI agents in production.

Details

The AI Agents Validation Framework & Continuous Monitoring service from Reply UK delivers a specialised, multi-layered approach to validating and assuring the quality of AI agents and large language model (LLM)-based systems. Recognising the non-deterministic, evolving, and data-dependent nature of AI agents, our framework goes beyond traditional static testing by implementing a holistic, data-driven validation strategy that spans the entire AI lifecycle.

Multi-Layered Test Scope:

Our framework evaluates AI agents across open knowledge (generalisation), specific knowledge (task/domain expertise), and adversarial knowledge (robustness against edge cases and manipulative inputs). This ensures comprehensive coverage of both expected and unexpected behaviours.



DATA-DRIVEN AI QUALITY ENGINEERING

A testing approach built around curated datasets and controlled experiments, ensuring consistent, reproducible, and scalable evaluation of AI agents.



AGENT-SPECIFIC QUALITY METRICS

A dedicated KPI framework that captures functional behavior, reliability, safety, and orchestration patterns unique to autonomous agents.



NEXT-GEN QA WORKFLOW

A modern, experiment-centric QA process that integrates automated evaluations, live monitoring, and continuous validation across the agent lifecycle.

Staged Validation Workflow:

The process includes discovery, pre-release evaluation, controlled production rollout, and continuous monitoring. Each stage is designed to incrementally build confidence in the AI agent's reliability, safety, and business readiness.



Continuous Monitoring:

Real-time monitoring is embedded to detect anomalies, performance drift, and deviations from expected outcomes in production. This enables rapid response to emerging risks and supports ongoing compliance and operational assurance.

Custom KPI Framework:

We define and track custom key performance indicators (KPIs) tailored to model quality, system reliability, adoption, business operations, and value delivery. These metrics provide actionable insights for continuous improvement.

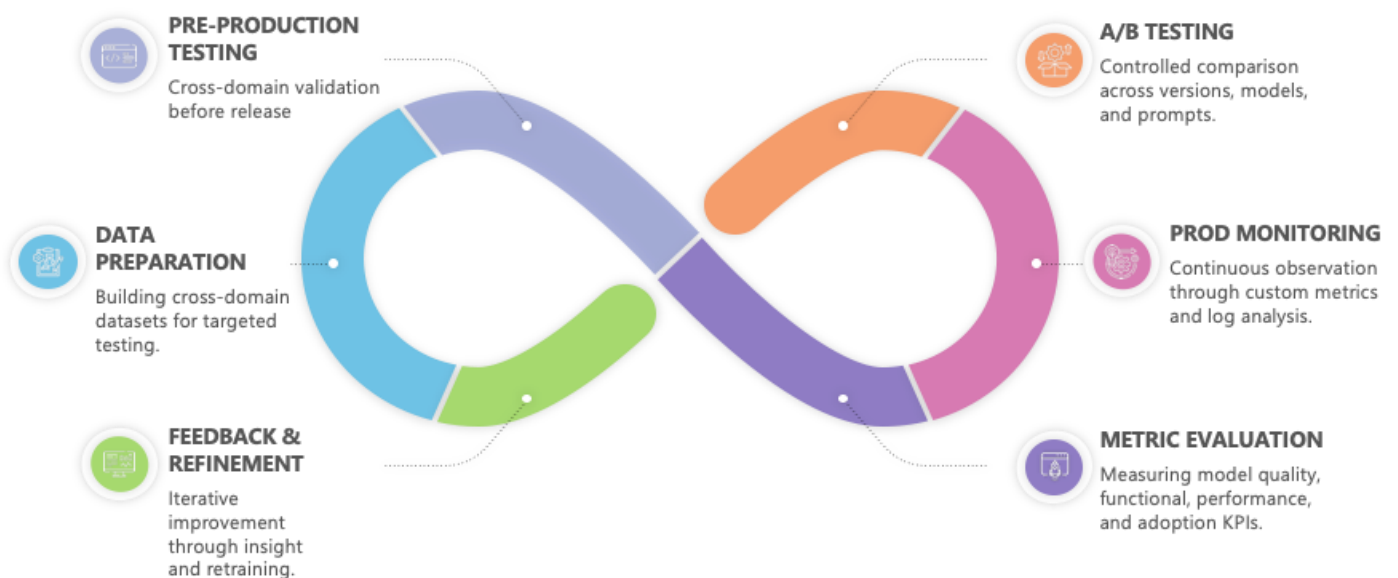


Repeatable Experiments:

Structured, repeatable experiments are conducted to validate consistency, regression, and adaptability of AI agents under varying conditions.

Human-in-the-Loop:

Subject matter experts are involved in reviewing ambiguous or critical outputs, ensuring that ethical, legal, and contextual considerations are addressed.



Diverse Test Datasets:

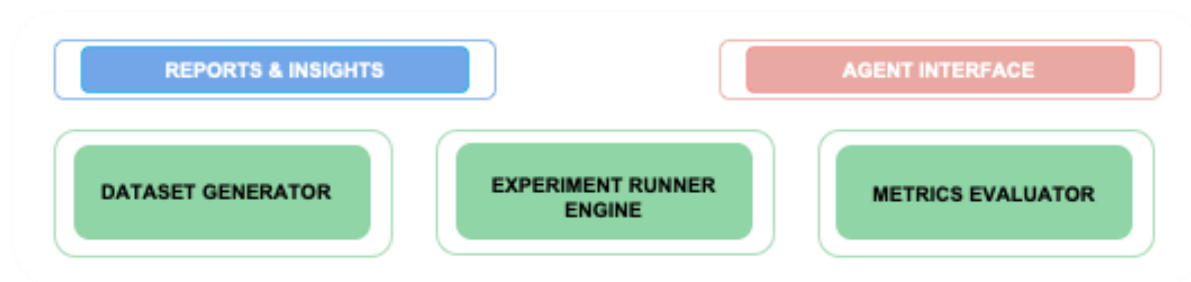
Both open and task-specific datasets are leveraged to ensure robust validation across a wide range of scenarios and data distributions.

Benchmark Rotation:

Regular updates to benchmarks and test datasets minimise the risk of overfitting and maintain the relevance of validation activities.



Architecture and Tools:



Benefits:

- Ensures model and system quality, reliability, and production readiness.
- Reduces operational and reputational risks by proactively identifying and addressing issues.
- Supports regulatory compliance and auditability with detailed reporting and traceability.
- Builds stakeholder confidence in the safe and effective deployment of AI agents.



ONBOARDING

Kick-Off Meeting

When we start an engagement with a new client, we begin with a workshop introducing the team (from both Reply and the client), the project objectives, and our anticipated timelines and approach. This workshop is most impactful when the majority of client stakeholders involved in the project are present and involved (either in person or remotely). Where this is impossible, we can hold separate sessions with teams that cannot participate in the kick-off workshop to ensure all stakeholders are engaged and bought into the project's objectives.

Account Management

Designated Account Manager

Each client has a designated Account Manager, who will be the main point of contact throughout the engagement. This individual is responsible for ensuring that delivery progresses and is executed to the standards that Reply and our clients expect.

Fortnightly review meeting

Alongside regular ceremonies and informal daily interaction between the Reply project team and the client, we hold formal review meetings with key stakeholders, usually fortnightly. This includes critical client stakeholders, senior Reply project team members, and the Reply Account Manager. Where appropriate, these meetings will consist of a review of work completed and demonstrations or walk-throughs of deliverables.

Offboarding

Where an engagement continues to completion, unless the client requests an extension or new engagement, all client IT, documents, artefacts and deliverables will be returned to the client. Our team will ensure a complete handover of data and knowledge to client staff. An ongoing support or retained advisory service may be implemented where appropriate to provide additional continuity.

Access to Data

Ensuring secure access to client data upon exit is paramount to maintaining confidentiality and integrity. We begin by conducting a comprehensive inventory of all client data stored within our systems, encrypting it both in transit and at rest using industry-standard protocols. Strict access controls and authentication mechanisms are implemented to restrict access to authorized personnel only, with permissions regularly reviewed and updated. Client data is segregated from other data within our systems, ensuring isolation and protection. Upon exit, all access rights to client data are promptly revoked, including user account deactivation and permission removal. If requested, we securely delete or anonymise client data in accordance with retention policies and regulatory requirements. Detailed documentation and audit trails of data access are maintained, providing transparency and accountability. Before exiting, we verify with the client that all necessary data access has been revoked. Finally, we adhere to legal and regulatory obligations regarding data protection and privacy, ensuring compliance with contractual agreements and applicable laws such as GDPR or HIPAA. Through these measures, we ensure that client data remains secure throughout the exit process, mitigating the risk of unauthorized access or data breaches.



SECURITY

Personnel Security

As a Specialist Cloud Service, the capability being offered is not limited to specific Impact levels (as it is not infrastructure, software or a platform) and can be used, subject to personal Security Clearance levels. Reply consultants are mostly Security Cleared (SC), some have higher-level Developed Vetting (DV) clearances. We also have a pool of NPPV3 Consultants for Policing work. The majority of our work for both public and private sector clients is at IL2 but we work in the Official, Secret and Top-Secret domains. Our hard and soft information security processes have been designed and approved by independent CESG CLAS accredited consultants.

Information Security

We are ISO27001, Cyber Essentials certified and our Quality Assurance processes are based upon and compliant with our ISO 9001 accreditation. Our Information Security processes are directly guided by our ISO27001 accreditation. We shall adhere to local information and other security policies and will apply local Security Operating Procedures (SyOPs) as may exist. If such do not exist we shall apply our own SyOPs.

Physical Security

The Main Reply Office (London, UK) is a Police Accredited Secure Facility (PASF) with access controls for each point of entry. All Laptops/Phones must be stored securely overnight. When working on a client-site our consultants adhere to client security policies.



TRAINING & KNOWLEDGE TRANSFER

We enable customer teams throughout delivery by embedding knowledge transfer alongside implementation activities. Training is aligned to organisational roles, ensuring relevance for finance users, operational staff, and system administrators.

Knowledge transfer is delivered through:

- A train-the-trainer approach to establish internal subject matter experts.
- Technical training for administrators covering configuration, security roles, and integrations.
- Walkthroughs of system design, data structures, and operational processes.
- Provision of comprehensive documentation, including user guides and runbooks.

Formal training is supported by ongoing access to expert guidance during delivery and early life support. Where required, refresher sessions or additional enablement can be provided to support organisational change or onboarding of new users.



ORDERING AND INVOICING PROCESS

To begin the process, please contact glue.frameworks@reply.com with details of your organisation, role and high-level requirements.

Our Framework Manager will connect you to our experts and from here we will organise an introductory discussion to go into more detail and shape a proposed engagement that suits your needs.

We invoice for both fixed price and time and materials engagements one month in arrears. For fixed price engagements, this takes place upon completion of the deliverable(s), or in stages, if the size of the engagement is greater than £40,000 excluding VAT.

Payment terms are 30 days net of receipt of invoice.



CUSTOMER RESPONSIBILITIES

To deliver our responsibilities as a partner, we require the following:

- Full access to systems and data that are relevant to the specific project
- Reasonable and timely access to key stakeholders and internal SMEs for information gathering, requirements definition and other critical tasks associated with the delivery of our outputs

Technical and Client-Side Requirements

Other than the system access specified above, there are no specific technical or client-side requirements that would limit our capacity to begin work.

Outcomes and Deliverables

Our engagement's specific outcomes and deliverables will be determined during the scoping phase before beginning the project based on your goals and objectives and the level of depth you require. These will include:

- Workshop(s) to discuss and assess the existing architectural design
- Detailed assessment of design with identification of risks and opportunities
- Report with recommendations and proposed action plan



ABOUT REPLY

Reply is a company that specialises in Consulting, Systems Integration and Digital Services with a focus on the conception, design and implementation of solutions based on the new communication channels and digital media.

Reply partners with key industrial groups in defining and developing business models made possible by the new technological and communication paradigms such as Artificial Intelligence, Big Data, Cloud Computing, Digital Communication, the Internet of Things and Mobile and Social Networking. In so doing, it aims to optimise and integrate processes, applications and devices.

Reply's offer is aimed at fostering the success of its customers through the introduction of innovation along the whole economic digital chain. Given its knowledge of specific solutions and due to a consolidated experience, Reply addresses the main core issues of the various industrial sectors.

Through its network of specialist companies, Reply supports some of Europe's leading industrial groups in Telco & Media, Industry & Services, Banks & Insurance, and Public Administration to define and develop business models, suited to the new paradigms of Artificial Intelligence, Big Data, Cloud Computing, Digital Media and the Internet of Things.

We make innovation happen.

We started with a small team and a purpose: to help the digital revolution happen. Today, we are a team of more than 14,600 people in 16 different countries but we still have the same DNA, made up of agile, vertical task forces and an inner passion for innovation.

We are a decentralised network of specialised companies.

Among Replyers, you will find passionate geeks, visionary strategists, and creative minds, each with sharp skills in a specific business, who cooperate together and never stop learning from each other.

Make forward, act sustainably.

We know a sustainable future is possible and technology can be a strong asset to reach it. As leaders in digital transformation, we push for change and operate in full accordance with the highest ethical standards and with respect for the rights of future generations