



AI testing and evaluation

G-Cloud 15

About us

We exist to unlock the full potential of government technology, transforming complexity into clarity and ambition into achievement. We partner with public sector organisations to deliver digital transformation that genuinely changes how citizens experience services, creating measurable impact across departments, communities, and the nation.

We believe the public sector deserves technology that functions brilliantly, not just adequately. Our expertise covers legacy modernisation, data strategy, platform engineering, cloud transformation, and AI implementation, but what truly defines us is our unwavering focus on delivering value. Every architectural decision, line of code, and strategic recommendation is guided by one question: how does this create better value for the people the government serves? This clarity allows us to address the most demanding challenges, from consolidating fragmented services and enhancing observability across complex estates.

Our agile delivery model, supported by a trusted network of over 800 certified, specialist associates, offers the close partnership of a boutique consultancy with the breadth of expertise needed for ambitious programmes. Operating exclusively outside IR35 provides real flexibility for rapid response to emerging priorities. Most importantly, we instil confidence in challenging transformation journeys, demonstrating that with the right partnership, talent, and approach, government technology can become a source of pride, efficiency, and genuine public benefit.



HM Revenue
& Customs



Department for
Science, Innovation
& Technology



Department
for Environment
Food & Rural Affairs



Crown
Commercial
Service



Department
for Education



We're always impressed about how much esynergy knows about our domain, and how they look to solve problems in a pragmatic, methodical and open-minded way. Working with them is a genuine team effort, and the respect they have for each other and their expertise is self-evident.



Department
for Environment
Food & Rural Affairs



Our services



Data



Platforms



Ways of working



Data experience



Platform integration



Technical agility



Data governance



Application modernisation



Organisational agility



Artificial intelligence



New platform build



User experience

AI testing and evaluation

Our AI Testing and Evaluation service offers an independent assessment of AI systems for accuracy, fairness, drift, robustness, and explainability. We provide automated tests, benchmarking datasets, and risk reports that support approval procedures and public sector assurance standards.

Our specialists design and implement comprehensive test suites that assess AI system performance across diverse scenarios and populations. We develop benchmarking datasets that accurately reflect real-world inputs, enabling meaningful evaluation of production performance rather than overly optimistic lab results. We perform fairness checks that evaluate AI outputs across protected characteristics, detecting potential bias before deployment impacts citizens. Our drift detection capabilities monitor model performance over time, alerting us when accuracy declines or outputs change unexpectedly. We perform robustness reviews that stress-test AI systems with adversarial inputs and edge cases, ensuring systems fail gracefully rather than catastrophically. We evaluate explainability to determine whether AI decisions can be understood and communicated to those affected and oversight bodies. Our automated testing frameworks integrate into CI/CD pipelines, enabling continuous evaluation as AI systems evolve. We provide quality reporting that tracks performance metrics and trends over time, supporting ongoing governance. We record risk observations and provide mitigation recommendations to support informed decisions on deployment and operation. Our model comparison features assess alternative approaches to guide selection choices.

Our service enables continuous validation through ongoing monitoring and periodic re-evaluation, ensuring AI systems perform as expected. We provide the evidence necessary for governance approvals and regulatory compliance.

Whether you're assessing new AI systems for deployment or maintaining ongoing assurance for production systems, we provide the testing expertise you need to operate AI with confidence.



Value proposition and benefits

Validate AI systems through our comprehensive AI testing and evaluation service that builds trust for deployment

Trusted outputs

Extensive testing confirms that AI systems generate accurate, reliable results in real-world situations.

Bias detection

Fairness testing recognises potential bias across protected characteristics before deployment.

Robust assurance

Independent evaluations offer confidence to governance bodies, ministers, and regulators.

Approval evidence

Test results and documentation support approval procedures and compliance standards.

Reduced incidents

Identifying issues before deployment prevents costly production incidents and reputational damage.

Governance support

Clear reporting facilitates governance decision-making and continuous oversight.

Model optimisation

Comparison testing determines the most effective methods before committing.

Performance visibility

Dashboards offer continuous insight into AI system behaviour and trends.

Drift detection

Ongoing validation identifies performance decline before it impacts results.

Explainability assessment

Evaluation confirms AI decisions are understandable and can be communicated effectively.

Quality improvement

Testing feedback facilitates ongoing enhancement of AI system performance.

Regulatory readiness

Evidence and documentation underpin emerging compliance requirements.

Training and knowledge transfer

We act as integrated partners, placing our specialists directly alongside your teams in a true one-team model. Through joint decision-making, paired work on complex challenges and shared responsibility for outcomes, your teams gain genuine mastery. Every architectural choice becomes a teaching moment, fostering continuous upskilling in real-world contexts.

Shadowing and reverse shadowing are central to our capability-building approach. Your team members initially shadow our specialists, learning not just what they do but why. As confidence develops, we switch roles through reverse shadowing, with your teams leading while we coach. This two-way process, proven effective at HMRC and DEFRA, transfers not just skills but belief, demonstrating your teams can achieve brilliance with the proper support and space to grow.

Training and knowledge transfer

We establish Communities of Practice as engines for collective learning and ongoing upskilling. These safe spaces enable civil servants across departments to share challenges, celebrate breakthroughs and access real-time coaching. When one department overcomes a complex challenge, that solution becomes accessible to all.

We document as we build, creating living playbooks, design patterns and implementation guides aligned with GDS standards and Data Standards Authority guidance. These actionable training resources empower your teams to tackle new challenges confidently. Through regular demos, playbacks, Q&As and formal training sessions paced to your readiness, we ensure knowledge is absorbed meaningfully. Our mid-contract Internal Team Collaboration Score KPI (target: 9/10) ensures upskilling remains on track, providing early signals for course correction.



Account management

Exceptional transformation starts with detailed onboarding and flourishes through proactive account management. We deliver tangible value from the first interaction, with support that encourages your independence rather than dependence.

Our approach aligns with the Technology Code of Practice and GDS Service Standard principles:

Prioritising user needs, adopting agile methods, and forming multidisciplinary teams that integrate seamlessly with your stakeholders. This ensures engagement starts with clarity of purpose, shared objectives, and governance frameworks that facilitate delivery—reducing ambiguity and speeding up decision-making.



Account management

Following GDS Design Principles guidance to focus on user needs, our onboarding emphasises thorough stakeholder engagement to understand genuine needs and desired outcomes. We run structured workshops that bring together technical teams, business owners, end users, and senior stakeholders to define success metrics, identify constraints, and establish a shared understanding of what done truly means. This prevents costly misunderstandings and scope creep, builds trust essential for an effective partnership, avoids weeks of potential rework, and guides the programme towards outcomes that genuinely matter.

We assemble teams aligned with Service Standard requirements, comprising user researchers, designers, developers, architects, and delivery managers to meet your programme's needs. These specialists integrate directly into your teams, adopting your communication channels, attending your ceremonies, and embedding within your ways of working. This one-team approach accelerates knowledge flow:

Our specialists quickly understand your context while your teams gain immediate access to expert skills. Complex challenges are addressed promptly whilst your internal teams develop capabilities that create lasting organisational value.



Workshops



Delivery governance



One-team collaboration

Case studies

DEFRA: Accelerating digital service delivery through a Core Delivery Platform



Problem

We partnered with DEFRA to develop a Core Delivery Platform (CDP) to modernise infrastructure and streamline digital service delivery. The goal was to create an evolutionary platform without lengthy lead times, complexity, or high costs for digital services. The project's central focus was to implement a set of guidance, tools, technologies, and practices to accelerate delivery, ensure quality, and simplify the operation.



Solution

We introduced standardised toolchains, automated pipelines, and infrastructure-as-code practices, enabling multi-tenant environments and reducing reliance on specialised DevOps resources. Our collaborative, value-driven approach focused on improving communication between siloed teams, removing duplication of efforts across programmes, improving automation visibility, and simplifying governance

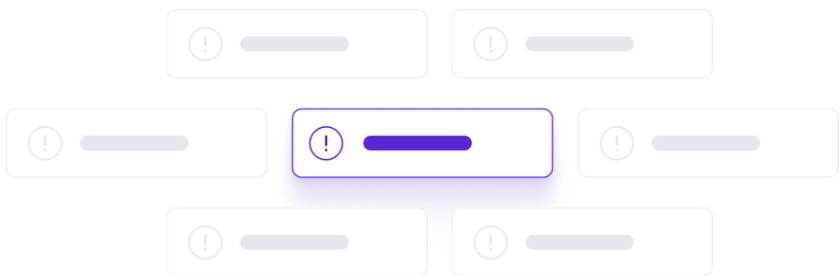


Outcome

Our solution led to significant cost savings, shortened development cycles, and improved compliance standards. By focusing on reusable patterns and self-service capabilities, DEFRA empowered its teams to deliver digital services more efficiently, thereby enhancing the organisation's overall agility and responsiveness.

Case studies

Crown Prosecution Service: Digitising expense claims to enhance the witness experience



Problem

The Crown Prosecution Service (CPS) collaborated with us to modernise its Witness Expenses (WITEX) programme, aiming to streamline witness reimbursement. Previously reliant on manual, paper-based forms, the system was time-consuming and prone to delays, impacting both staff efficiency and witness satisfaction.



Solution

We facilitated a full Government Digital Service (GDS) product lifecycle to develop a user-friendly digital platform that reduced cognitive load and improved accessibility. We conducted comprehensive user research to understand pain points and challenges. In Alpha, we refined the solution through iterative design reviews, leading to Beta for full testing, prioritising inclusivity and continuous improvement.

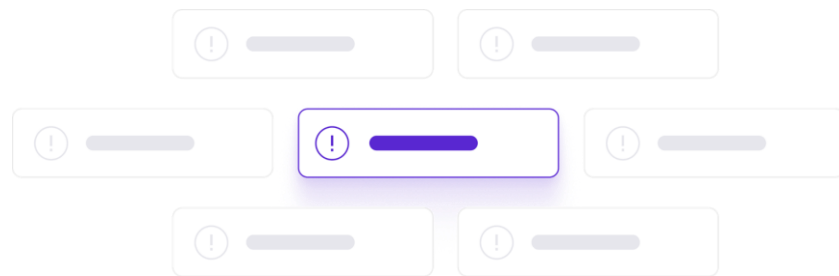


Outcome

The new system achieved a 92% user satisfaction rate, saved approximately 5,000 staff hours annually, and enabled 50% of claims to be submitted within one day of the court appearance. This digital transformation not only enhanced operational efficiency but also ensured timely financial support for witnesses, reinforcing their crucial role in the justice system.

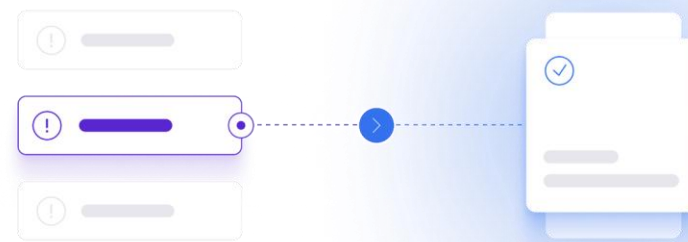
Case studies

HM Revenue & Customs: AI-Driven solution to improve consistency and increase confidence



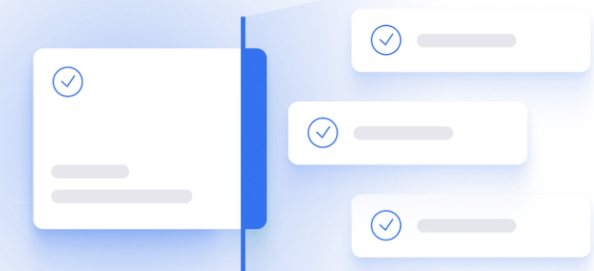
Problem

HMRC manages hundreds of digital services and APIs supporting tax collection, compliance, and citizen services. Teams struggled with fragmented API estates, inconsistent data models, and limited visibility into how APIs aligned with business domains. This slowed impact analysis, increased manual effort during change activities, and made it difficult to answer fundamental questions about which APIs support specific tax journeys.



Solution

We introduced an AI-driven approach to ingest HMRC OpenAPI specifications and documentation automatically. Using large language models and automated parsing, schemas were extracted, normalised, and mapped into a unified domain model. The solution enabled natural-language querying, automated domain classification, and rapid insight generation, significantly reducing manual review while creating a shared, queryable view of HMRC's API landscape for design, assurance, and governance activities.



Outcome

Within two weeks, over 100 API specifications were analysed and structured into a unified model. Manual analysis effort reduced by approximately 60%, while time to answer architectural questions dropped from days to minutes. Early findings identified 25%-30% schema overlap across services, highlighting clear rationalisation opportunities. The approach improved consistency in domain understanding, increased confidence in change impact assessments, and established a scalable foundation for ongoing API governance and modernisation across HMRC.

“

What you've delivered is the best thing we've received in such a long time. Thank you



HM Revenue
& Customs



Reporting and governance

We ensure visibility through transparent reporting aligned with your governance schedules, providing clear insights into progress, risks, dependencies, and value delivered. We offer honest, actionable intelligence to build trust and keep stakeholders confident in the programme's direction. Leaders can make informed investment decisions, teams can prioritise effectively, and your organisation understands how delivery connects to business outcomes.

We actively monitor programme health through ongoing stakeholder engagement, team sentiment assessments, and early warning signals. Minor issues are quickly managed before they escalate into programme - threatening problems, preventing costly crises while maintaining delivery timelines and stakeholder trust.



Reporting and governance

Following the GDS Design Principle of frequent iteration, we include structured checkpoints at Week One, Week Four, Week Eight, and monthly thereafter. We gather feedback, assess what's working, identify friction points, and adapt accordingly. Delivery methods are continually refined, team productivity improves, and outcomes progress as we collectively determine what works best in your specific context.

We adopt your existing development environments, CI/CD pipelines, collaboration platforms, and project management tools instead of imposing our preferences. When opportunities for improvement arise, we suggest changes transparently, supported by clear business cases, and in line with the Technology Code of Practice guidance on open standards, cloud-first strategies, and avoiding vendor lock-in. Your teams continue working in familiar environments, existing investments are utilised, and any new tooling delivers measurable returns.

- ☒ Rapid value delivery
- ☒ Smooth governance
- ☒ Sustained collaboration

The outcome

Within days rather than weeks, joint teams deliver value, governance structures operate smoothly, and the foundation is established for sustained collaboration that achieves outcomes rather than just outputs.

“

Over the years, we've had lots of suppliers say they will put experts on the project, but this is the first time it has been true



Department for
Science, Innovation
& Technology



Quality assurance

We incorporate quality assurance from the earliest stages of engagement, ensuring defects are prevented rather than detected. To meet the GDS Service Standard requirement for end-to-end testing, we employ comprehensive testing strategies that include unit, integration, and end-to-end testing within automated CI/CD pipelines.

Accessibility is integrated from the beginning. We test against WCAG 2.2 AA standards throughout development using both automated tools and manual testing with assistive technologies. To meet Service Standard requirements for making services accessible to all users, our teams include accessibility experts as standard, conduct regular audits, and incorporate inclusive design practices. This prevents costly remediation and ensures compliance with the Public Sector Bodies Accessibility Regulations.

Security testing is continuous and comprehensive. In accordance with the Technology Code of Practice and National Cyber Security Centre guidance, we integrate security testing throughout the entire development lifecycle, from threat modelling during design to automated vulnerability scans and penetration tests before launch. Our teams stay informed of OWASP standards and government security frameworks, ensuring that code is secure by design.



Testing coverage

- Unit test
- Integration test
- End-to-end test
- Automated CI/CD testing



Accessibility

- WCAG 2.2 AAA
- Automated + manual AT testing
- Inclusive design
- Accessibility experts



Security

- Threat modelling
- Vulnerability scans
- Penetration tests
- OWASP, NCSC, government standards

Quality assurance

We maintain full traceability between requirements, tests, and delivered functionality to support audits and ensure each user requirement is properly tested. Combined with regular code reviews, pair programming, and clear definition-of-done criteria, our approach delivers robust, fit-for-purpose services from the start.

Our quality practices ensure that government departments successfully pass GDS service assessments. We integrate assessment readiness into every stage of delivery, documenting decisions, demonstrating user research integration, and providing evidence of how services meet each of the 14 Service Standard points. We support government departments through alpha, beta, and live assessments, building internal confidence while ensuring services genuinely meet the standard in practice.



Traceability & governance

- ▀ Requirements → tests → delivery
- ▀ Code reviews
- ▀ Pair programming
- ▀ Definition of done



GDS assessment readiness

- ▀ Documentation
- ▀ User research evidence
- ▀ Meeting the 14 service standards
- ▀ Passing alpha → beta
→ live assessments

Data protection and security

Security is fundamental to everything we provide. We possess ISO 27001 certification, demonstrating our dedication to information security management through independently audited policies, processes, and controls. We also hold Cyber Essentials Plus certification, meeting the government's standards for suppliers managing sensitive public sector data and defending against the most common cyber threats.

Our security practices align with the Technology Code of Practice requirements for secure design and the National Cyber Security Centre (NCSC) guidance. We implement the NIST Cybersecurity Framework principles—Identify, Protect, Detect, Respond, and Recover—ensuring comprehensive risk management throughout delivery. Our teams apply NCSC's secure development and deployment guidance, building security into architecture decisions from the outset rather than retrofitting controls later.



Data protection and security

Data protection is integrated throughout our operations. We fully comply with the UK GDPR and the Data Protection Act 2018, applying privacy-by-design principles to every service we develop. Our teams understand the specific requirements for handling government data, including proper data classification, secure data-handling procedures, and compliance with data sovereignty obligations.

We conduct regular security assessments, including penetration testing, vulnerability scanning, and code reviews, in line with OWASP standards. Our staff receive security awareness training and are vetted to suitable levels for government work. We maintain robust incident response procedures to detect, contain, and report security incidents in accordance with NCSC guidelines and contractual obligations.

Our approach enables government departments to confidently fulfil their obligations under the Government Security Classifications policy and broader public sector security standards, with comprehensive audit trails and documentation that support assurance and compliance activities throughout engagements.



Business continuity & disaster management

Continuity of service is essential when delivering for the government. Our business continuity management complies with ISO 22301, ensuring we maintain robust plans and capabilities to sustain operations during disruptions and recover quickly from incidents. We regularly test and update these plans to keep them responsive to changing risks.

We conduct business impact assessments to identify critical functions and dependencies, enabling proportionate measures that safeguard essential government services. Our teams apply the NCSC's resilience principles, designing for failure and ensuring services degrade gracefully when issues arise.

Our delivery embeds resilience through DevOps practices and Infrastructure as Code (IaC). By codifying infrastructure in version-controlled repositories, we eliminate configuration drift and enable rapid environment rebuilds, reducing recovery times from days to minutes. Automated pipelines enable confident deployment and rapid rollback, minimising service disruption, while continuous monitoring provides early anomaly detection. We design services with disaster recovery integrated from the start, including redundancy, failover mechanisms, and backup strategies aligned with agreed Recovery Time Objectives and Recovery Point Objectives. This guarantees services can endure infrastructure failures, cyber incidents, or other disruptions while safeguarding data integrity.



Social value

We are dedicated to delivering meaningful social value in line with the government's five missions outlined in PPN 002 and the National Procurement Policy Statement.



Supporting economic growth

We create high-quality employment opportunities with fair wages and good working conditions, emphasising apprenticeships, graduate placements, and career development pathways. We actively collaborate with SMEs and VCSEs within our supply chain, building capacity and resilience across the sector.



Contributing to Britain as a clean energy superpower

We adopt sustainable practices throughout our operations, reducing our carbon impact by using remote-first delivery models, cutting travel, and making environmentally responsible procurement choices that accelerate progress towards net zero.



Advancing safer streets

We promote ethical employment practices and maintain rigorous modern slavery due diligence throughout our supply chain, ensuring our operations contribute positively to community safety and wellbeing.



Social value



Breaking down barriers to opportunity

We invest in skills development to address digital skills gaps, support underrepresented groups entering the technology sector, and ensure our recruitment practices promote diversity and inclusion. Through our dedicated Not for Profit department, we deliver digital transformation programmes to third-sector organisations at significantly reduced rates, enabling charities to access the same quality expertise as government departments and maximise their impact within communities.



Supporting a future-ready NHS

We prioritise employee wellbeing through flexible working, private healthcare, mental health support, and healthy workplace practices.

We provide clear, quantifiable, and time-sensitive commitments in every contract, with transparent reporting based on standard metrics.

Our dedicated Not-for-Profit practice offers preferential rates to third-sector organisations, enabling charities and social enterprises to access transformational capabilities to maximise their operational efficiency and service delivery while directly advancing the government's Social Value agenda.



 bids@esynergy.co.uk

 0207 444 4080

 esynergy.co.uk

Contact details