The Intelligent State

How AI Can Transform UK Central Government

A Department-by-Department Analysis of Efficiency Gains, Cost Savings, and Implementation Requirements

Centre for a Better Britain

March 2026

Executive Summary

The United Kingdom's central government employs 549,660 civil servants and spends £592.3 billion annually through departmental budgets. The Spending Review 2025 mandates a 16% real-terms reduction in administration budgets by 2029/30, requiring approximately £2.2 billion in annual savings and — according to the Institute for Government — between 29,000 and 40,000 fewer staff (Spending Review 2025; Whitehall Monitor 2026). Traditional efficiency programmes cannot deliver reductions of this magnitude without degrading public services. Artificial intelligence — specifically multi-model deliberation systems, algorithmic case management, and agentic workflows — offers the only credible route to achieving these savings while improving the speed and quality of government services.

This paper finds that a well-sequenced AI transformation programme can credibly deliver:

Metric	LOW	CENTRAL	HIGH
Net headcount reduction by 2031 (posts)	34,000	40,500	47,000
Gross annual savings by 2031	£2.3bn	£3.0bn	£3.8bn
Implementation investment (5-year total)	£2.4bn	£3.2bn	£4.2bn
Net annual savings by 2031	£1.2bn	£2.0bn	£2.8bn
10-year NPV at 3.5% discount rate	£4.8bn	£8.6bn	£13.1bn

These estimates use Paper X's fiscal spine — the most conservative and defensible of the three independent analyses commissioned for this paper — supplemented by the delivery architecture of Paper Y, where it strengthens the implementation case. They are anchored to the government's own administration-budget settlement and benchmarked against the documented track record of past efficiency programmes, including the Gershon Review, CSR07, and SR2010 (NAO on CSR07 value-for-money savings).

The largest savings come from four high-volume transactional departments — HMRC, DWP, MoJ, and the Home Office — which together account for roughly 60% of the total opportunity. Most headcount reduction (55–65%) should be delivered through managed attrition and vacancy deletion, not compulsory redundancy. A workforce compact with unions is not optional window-dressing; it is a delivery precondition.

Our principal recommendation is that the government should establish a central AI Transformation Unit, jointly accountable to the Prime Minister and the Chancellor, with authority over departmental adoption targets, vacancy controls, and infrastructure investment. Without central enforcement, departments will consume productivity gains as additional activity rather than banking them as fiscal savings. That is the lesson of every previous efficiency programme.

Part 1: The Case for AI in Government

1.1 Why Now: The Convergence of Capability, Fiscal Pressure, and Political Will

Three forces have converged to create the conditions for a genuine transformation of the administrative state.

First, AI capability has reached a threshold of practical utility for government work. The government's own 20,000-person trial of Microsoft 365 Copilot found that civil servants saved an average of 26 minutes per day, equivalent to nearly two working weeks per year per user (Landmark Government Trial Shows AI Could Save Civil Servants Nearly 2 Weeks a Year). The Alan Turing Institute's systematic mapping found that 41% of public sector working time is spent on tasks that generative AI can support (Mapping the Potential: Generative AI and Public Sector Work). Deloitte research documents 75–95% time savings on specific tasks such as report drafting, document routing, and evidence synthesis. These are no longer theoretical projections; they are measured outcomes in UK government settings.

Second, fiscal pressure is acute and structural. The Spending Review 2025 requires all departments to reduce their administration budgets by at least 16% in real terms by 2029/30, saving an expected £2.2 billion per year. Beyond this, the Autumn Budget 2025 added a further £4.9 billion in efficiency targets by 2030/31 (Civil Service World). Departmental efficiency plans have identified almost £14 billion of annual efficiency gains by 2028/29, with 74% of departments citing AI as a delivery mechanism (Departmental Efficiency Plans). The arithmetic is simple: without AI-enabled productivity gains, meeting these targets will require degrading frontline services or crude headcount cuts with no compensating improvement in capability.

Third, the political and institutional infrastructure now exists. The government has committed a £3.25 billion Transformation Fund, £1.2 billion of cross-cutting digital funding via DSIT, £500 million for HMRC digital services, a £100 million Test/Learn/Grow fund, and a £500 million Sovereign AI Fund launched in March 2026 (Spending Review 2025; GOV.UK). The institutional architecture includes i.AI in the Cabinet Office, the AI Playbook, the Algorithmic Transparency Recording Standard, the AI Opportunities Action Plan (38 of 50 actions now delivered), and departmental AI leads in all major departments (AI Opportunities Action Plan: One Year On).

1.2 The AI Policy Lifecycle: Where AI Intervenes at Each Stage

Government work follows a policy lifecycle. AI can intervene at every stage, but the savings profile differs sharply across them.

Stage	AI Application	Primary Benefit	Savings Type
Problem diagnosis	Pattern detection across datasets; early-warning systems; cross-departmental anomaly identification	Speed, coverage	Augmentation
Policy generation and selection	Multi-model deliberation; evidence synthesis; option appraisal; red-teaming	Quality, speed	Augmentation
Drafting	Legislative and regulatory drafting assistants; impact assessment support; consistency checking	Speed, consistency	Partial substitution
Implementation	Programme management copilots; procurement review; delivery tracking	Speed, error reduction	Partial substitution
Operational delivery	Case processing; correspondence; contact centres; document verification; fraud detection	Volume, cost	Direct substitution
Monitoring and iteration	Real-time dashboards; outcome measurement; feedback loops; policy adjustment	Quality, responsiveness	Augmentation

A useful taxonomy, drawn from Paper X's political-economy analysis, classifies AI applications into three layers:

1. **Layer 1: Horizontal productivity tools** — email copilots, meeting summarisation, search, coding assistants. Easy to deploy; modest per-user savings; already being rolled out.
2. **Layer 2: Mission-workflow tools** — case-management systems, document processing pipelines, fraud detection, compliance checking. Harder to build; require data integration; deliver the largest cashable savings.
3. **Layer 3: Decision-support and policy systems** — multi-model deliberation, policy simulation, cross-departmental analytics. Highest strategic value; hardest to implement; savings mostly in quality and avoided errors rather than cash.

Governments systematically overinvest in Layer 1 because it is easy to announce. The real fiscal case depends on Layer 2 deployment in high-volume departments. Layer 3 generates disproportionate strategic value but should not be counted as a near-term savings line.

1.3 Multi-Model Deliberation Systems: What They Are and Why They Work

Multi-model deliberation systems — sometimes called "AI think tanks" — represent the frontier of AI application in policy work. Rather than relying on a single AI model, these systems deploy multiple specialised models with different training data, reasoning approaches, and domain knowledge to debate, red-team, and synthesise policy options.

The advantages over single-model or single-analyst approaches are threefold. First, they reduce the risk of systematic bias by introducing genuine diversity of analytical perspective — a model trained primarily on economic data will surface fiscal risks that a model trained on operational data might miss, and vice versa. Second, they can process and synthesise orders of magnitude more evidence than any human team. A policy team preparing a submission on housing policy might need to review hundreds of consultation responses, dozens of local authority data returns, and a range of academic evidence; a multi-model system can do this in hours rather than weeks. Third, they provide structured red-teaming that challenges assumptions before they become policy — one model proposes, another critiques, and a synthesis model reconciles.

The Congressional Budget Office in the United States has noted that AI serves dual economic roles: as a substitute for routine labour (reducing operational costs) and as a complement to high-skill labour (enhancing quality and output). Multi-model systems excel in the latter. They dramatically reduce the cognitive hours that Grade 7, Grade 6, and Senior Civil Service officials spend on option appraisal and evidence synthesis. Deloitte research indicates that AI can save 75–95% of the time spent on report drafting and document analysis. More importantly, by introducing rigorous algorithmic red-teaming, these systems reduce the risk of catastrophic policy failures — such as under-costed programmes or unforeseen regulatory interactions — which historically carry massive and sometimes unquantifiable fiscal liabilities.

The practical deployment model works as follows. A team of three to five specialised AI agents, each with access to different data sources and analytical frameworks, independently generates analysis on a given policy question. Their outputs are compared and reconciled by a synthesis layer that identifies areas of agreement, flags disagreements, and highlights uncertainties. A human policy team then reviews the synthesised output, exercises judgment on contested points, and takes responsibility for the final recommendation. The AI does not decide; it dramatically expands the evidence base and analytical rigour available to the humans who do.

For government, multi-model systems are most valuable in fiscal forecasting, legislative impact assessment, cross-departmental policy analysis, and risk assessment. However, their output is augmentation, not substitution: the constitutional responsibility for decisions remains with ministers and accounting officers. These systems should be understood as the most powerful analytical tool ever available to the policy profession, not as a replacement for political judgment.

1.4 Governance Framework: Accountability, Transparency, and Human-in-the-Loop

The UK's AI Playbook states that "responsibility for any output or decision made or supported by AI always rests with the public organisation," requires "meaningful human control at the right stages," and obliges departments to use the Algorithmic Transparency Recording Standard for algorithmic tools used in decision-making (AI Playbook for the UK Government). These principles are necessary but not sufficient.

The governance framework for AI-enabled government must address four levels:

Ministerial accountability. Ministers remain answerable to Parliament for all decisions made in their name, whether assisted by AI or not. The ministerial code and accounting-officer responsibilities are unchanged. What changes is the requirement for auditable records of how AI shaped recommendations.

Operational accountability. Every AI system deployed in a decision-affecting role must have a named Senior Responsible Owner (SRO) at SCS level, a published Algorithmic Transparency Record, a documented human-in-the-loop protocol specifying when and how human judgment intervenes, and a clear escalation pathway for edge cases and errors.

Technical accountability. Model performance must be continuously monitored for accuracy, bias, and drift. Fairness audits must be conducted before deployment and at regular intervals thereafter. The DWP Universal Credit Advances fairness assessment provides a working model (DWP Universal Credit Advances Model — Fairness Assessment).

Democratic accountability. AI must strengthen, not weaken, parliamentary scrutiny. Departments should be required to disclose AI use in responses to parliamentary questions and Freedom of Information requests. Select committees should have access to independent technical expertise to interrogate AI-informed decisions. The Public Accounts Committee's January 2025 session on AI in government highlighted that only 33 Algorithmic Transparency Records had been published despite hundreds of AI contracts across government — a transparency deficit that must be closed urgently.

1.5 Security Architecture: OFFICIAL through TOP SECRET Deployment Models

Security classification is the most underappreciated constraint on AI deployment in government. The UK Government Security Classifications framework defines four tiers, each with radically different implications for AI hosting, processing, and access:

Classification	Description	AI Deployment Model	Departments Most Affected
OFFICIAL	Routine government business	Commercial cloud (AWS, Azure, GCP) with appropriate controls	All departments (majority of work)

Classification	Description	AI Deployment Model	Departments Most Affected
OFFICIAL-SENSITIVE	Requires strengthened controls	Assured commercial cloud with enhanced monitoring; UK data residency	Most departments
SECRET	Compromise could cause serious damage	Rosa-type assured capability; restricted hosting; enhanced vetting	MoD, Home Office, FCDO, Cabinet Office, intelligence
TOP SECRET	Compromise could cause exceptionally grave damage	Secure Isolated Networks; air-gapped infrastructure; highly secure physical environments	MoD, intelligence community

The implications for the AI business case are significant. Approximately 80–85% of civil service work operates at OFFICIAL or OFFICIAL-SENSITIVE, where commercial AI tools can be deployed with appropriate controls. This is where most of the savings in this paper originate. But the remaining 15–20% — concentrated in MoD, FCDO, the Home Office's national security functions, and Cabinet Office intelligence coordination — requires classified AI environments that are expensive to build, limited in model choice, and slow to update (Guidance on working at SECRET; Guidance on working at TOP SECRET).

This fragmentation means that a single "Common AI Platform" cannot serve all of government. The realistic architecture is a tiered system: commercial tools for OFFICIAL work, an assured government cloud for OFFICIAL-SENSITIVE, and bespoke secure environments for SECRET and above. Singapore's air-gapped deployment of large language models for classified work provides a working model, but it is expensive and limited in scope (Singapore GovTech AI and Data).

Paper Y's delivery architecture correctly identifies this tiered approach. Paper Z's proposal for a single sovereign compute estate conflates efficiency with industrial policy and would not be scoreable as an administrative saving.

1.6 International Benchmarks and Lessons

The UK is not starting from zero, but it is not yet at the frontier of government AI adoption.

Singapore offers the most relevant institutional model. GovTech operates as a whole-of-government technology organisation with shared AI platforms, mandatory adoption standards, and agency-level experimentation. Its SkillsFuture AI deployment processed 30,000 quarterly CRM cases 62.5 times faster than manual analysis, demonstrating the scale of efficiency gains achievable with well-structured data (Singapore GovTech).

Estonia demonstrates the value of deep digital infrastructure before AI layering. With 99% of state services online and 47+ Kratt AI solutions operational, Estonia shows that interoperability must come first; AI adoption is a consequence, not a cause, of good data

architecture. Estonia's Eesti.ai initiative aims to double public sector productivity by 2035 (E-Estonia).

The UAE represents the most aggressive adoption globally, with 97% AI utilisation across federal entities and a dedicated Minister for AI since 2017. While the UAE's centralised governance model is not replicable in a parliamentary democracy, its institutional clarity — a single point of accountability for AI strategy — contrasts favourably with the UK's diffused responsibility across DSIT, GDS, i.AI, and individual departments.

Denmark provides a model for political governance, with a cross-party Digital AI Taskforce mandated to set specific objectives for FTE freed through AI by 2030, backed by DKK 740 million in dedicated funding.

The United States has moved decisively under Executive Order on AI governance, with agency AI strategies, shared asset inventories, and a new “Tech Force” initiative to recruit 1,000 technologists for federal AI modernisation.

The UK's comparative advantage is a strong centre, common law, a sophisticated digital profession, and ministerial control over spending. Its weakness is departmentalism, a legacy IT estate in which 28% of systems are classified as legacy technology, and a civil service culture that confuses review with risk management (NAO Government cyber resilience).

Part 2: Department-by-Department Analysis

We group departments into four tiers based on AI readiness and potential impact. For each department we provide: function and scale, specific AI transformation opportunities, efficiency estimates with LOW/CENTRAL/HIGH scenarios, implementation timeline, and key risks.

Methodology note. Headcount estimates use a task-exposure model rather than naive job-replacement assumptions. If 50% of tasks for a group of workers can be automated, this translates to approximately 20–25% headcount reduction over five years after accounting for institutional friction, new oversight roles, redeployment, and residual task complexity. Savings calculations use a blended average civil service employment cost of £50,000 per post (salary plus employer National Insurance, pension contributions, and overhead), which is more conservative than the lower figure used in some analyses and more realistic given the upward drift in the civil service grade profile since 2010 (Civil Service Statistics 2025).

Tier 1: High-Volume Transactional Departments

These departments have the largest absolute savings opportunity because they combine high headcount with document-heavy, rule-based operational delivery.

2.1 HM Revenue & Customs (HMRC)

Function and scale. HMRC collects over £800 billion annually in tax revenue, administers tax credits, and enforces compliance. It employs 70,925 staff, making it the third-largest department by headcount. HMRC's efficiency plans identify £663 million of efficiencies by 2028/29, supported by £500 million in digital investment including SAP S/4HANA migration for its core tax platform (Spending Review 2025; Departmental Efficiency Plans).

AI transformation opportunities.

- **Automated enquiry handling and correspondence:** HMRC handles millions of taxpayer enquiries annually. AI-assisted triage, response drafting, and self-service chatbots can reduce contact-centre and correspondence workload by 40–60%.
- **Algorithmic compliance risk scoring:** AI can prioritise audit activity by scoring taxpayer risk profiles across all tax heads, improving yield per investigation and reducing wasted effort on compliant taxpayers.
- **Automated tax code management:** Cross-referencing P60s, P45s, and employer submissions against self-assessment records — currently a heavily manual process — is highly amenable to algorithmic processing.
- **Post reduction:** HMRC's own efficiency plans target £50 million per year from eliminating physical post through digital channels.
- **Document extraction and processing:** OCR and NLP pipelines can automate data extraction from paper-based submissions and legacy correspondence.

Efficiency estimates.

Metric	LOW	CENTRAL	HIGH
Headcount reduction	5,300 (7.5%)	7,100 (10%)	9,200 (13%)
Annual savings	£265m	£355m	£460m
Speed gains	30–50%	40–60%	50–70%

Implementation timeline. HMRC is already well advanced. The SAP migration provides a modern data foundation. Contact-centre AI and correspondence automation are deployable within 12–18 months. Full compliance risk-scoring across all tax heads requires 2–3 years. The binding constraint is the SAP migration itself, which must not be disrupted.

Key risks. SAP migration complexity and cost overruns; taxpayer trust in automated compliance decisions; revenue-sensitivity of errors in algorithmic risk scoring; regional employment impacts given HMRC's geographic concentration.

2.2 Department for Work and Pensions (DWP)

Function and scale. DWP is the largest single civil service employer with 96,890 staff. It administers approximately £291 billion annually in welfare payments, including Universal Credit, State Pension, and disability benefits. DWP already has live machine-learning fraud detection that is three times more effective than random sampling and is expanding this capability across benefit lines (The Register on DWP ML).

AI transformation opportunities.

- **Case preparation and document verification:** AI can pre-process claims documentation, verify identity, cross-reference entitlements, and prepare cases for human decision-makers, reducing the administrative burden per claim by 40–60%.
- **Fraud and error detection:** DWP's existing ML capability provides a strong base for expansion. The £8.3 billion annual fraud and error bill represents a vast opportunity, though savings here are analytically distinct from headcount savings and should not be conflated.
- **Work coach support:** AI assistants can help work coaches prepare for claimant meetings, identify suitable employment matches, and track interventions.
- **Automated correspondence and contact-centre support:** DWP processes millions of letters and calls; AI triage and drafting can reduce this workload substantially.

Efficiency estimates.

Metric	LOW	CENTRAL	HIGH
Headcount reduction	6,800 (7%)	9,700 (10%)	13,500 (14%)
Annual savings	£340m	£485m	£675m
Speed gains	30–50%	40–60%	50–70%

Note: Fraud and error prevention savings (potentially £0.5–1.5 billion additional) are analytically separate from headcount savings and should be counted in a different fiscal category. They depend on robust counterfactuals and cannot be straightforwardly scored as DEL savings.

Implementation timeline. DWP has the digital maturity and scale to be a first-mover. Case-preparation AI and expanded fraud detection are deployable within 12 months. Full workflow redesign across all benefit lines requires 3–4 years. Universal Credit's troubled implementation history is a standing reminder that programmes touching millions of citizens routinely take longer than planned (NAO progress in implementing Universal Credit).

Key risks. Algorithmic bias in welfare decisions affecting vulnerable populations; legal challenge under equality and human rights law; political sensitivity of any AI-related denial of benefits; union opposition from PCS, which represents the majority of DWP staff; regional employment concentration in areas with limited alternative employment.

2.3 Ministry of Justice (MoJ)

Function and scale. MoJ employs 96,210 staff across courts, tribunals, prisons, probation, and legal services. It faces a historic Crown Court backlog and severe capacity constraints. MoJ's efficiency plan identifies £356 million of efficiencies by 2028/29, with AI and digital reform featuring prominently. MoJ already reports approximately 30 minutes per day saved per user from AI tools (AI action plan for justice; Departmental Efficiency Plans).

AI transformation opportunities.

- **Court transcription and summarisation:** AI transcription deployed across Crown Courts can dramatically reduce court-time administrative burden. MoJ is already piloting this.
- **Case bundle preparation:** Automated assembly, indexing, and summarisation of case bundles for hearings.
- **Pre-sentencing report support:** AI-assisted drafting of background elements, reducing probation-officer administrative time.
- **Prison and probation administration:** Automated scheduling, record management, and risk-assessment support.
- **Tribunal case triage:** Automated initial review and categorisation of tribunal applications.

Efficiency estimates.

Metric	LOW	CENTRAL	HIGH
Headcount reduction	5,800 (6%)	8,200 (8.5%)	11,500 (12%)
Annual savings	£290m	£410m	£575m
Speed gains	25–40%	35–55%	45–65%

Implementation timeline. AI transcription and case-bundle automation are deployable within 12–18 months in courts that have already adopted the Common Platform. Broader workflow redesign across the prison and probation estate requires 3–5 years.

Key risks. Judicial independence — courts cannot be told what technology to use without judicial buy-in; constitutional sensitivity of AI anywhere near sentencing or case-outcome decisions; the court backlog is constrained by judicial time and court estate, not just

administrative processing; Common Platform rollout is still incomplete.

2.4 Home Office

Function and scale. The Home Office employs 50,655 staff across immigration, borders, policing, security, and fire services. Its efficiency plan identifies £533 million of efficiencies by 2028/29. The Home Office 2030 Digital Strategy sets out ambitious plans for digital border services (Departmental Efficiency Plans; Home Office 2030 Digital Strategy).

AI transformation opportunities.

- **Visa application processing:** Automated document verification, eligibility checking, and risk scoring for straightforward visa applications. The majority of routine visitor and work visas follow well-defined rules.
- **Biometric border screening:** Enhanced automated identity verification at ports of entry.
- **Correspondence and casework support:** AI drafting of routine correspondence and casework responses.
- **Intelligence analysis:** Automated open-source intelligence gathering and pattern detection (subject to security classification constraints).

Efficiency estimates.

Metric	LOW	CENTRAL	HIGH
Headcount reduction	3,000 (6%)	4,300 (8.5%)	6,100 (12%)
Annual savings	£150m	£215m	£305m
Speed gains	25–40%	35–55%	45–65%

Implementation timeline. Routine visa processing automation is the clearest near-term opportunity and could be deployed within 18 months for straightforward categories. Border automation requires infrastructure investment at ports. Asylum decision support is legally and politically constrained and should not be in the first wave.

Key risks. This is the most legally and politically sensitive department for AI deployment. Automated immigration decisions are "legally and politically radioactive," in the words of one critique panel member. Even AI-assisted triage can materially influence outcomes. Fairness, equality, and challenge rights must be demonstrably protected. The Windrush scandal remains a powerful institutional memory of what happens when bureaucratic systems fail vulnerable people. AI in national security functions faces additional classification constraints.

Tier 2: Complex Operational Departments

These departments have significant savings potential but face higher implementation complexity due to classification requirements, mission sensitivity, or system fragmentation.

2.5 Ministry of Defence (MoD)

Function and scale. MoD employs 57,500 civilian staff alongside 150,000+ military personnel. Civilian functions include procurement, logistics, estate management, and administrative support. MoD is a heavy user of classified systems.

AI transformation opportunities.

- **Procurement review and contract management:** AI-assisted analysis of defence procurement documents, contract compliance monitoring, and cost estimation.
- **Predictive maintenance:** Algorithmic monitoring of military asset condition to optimise maintenance scheduling and reduce downtime.
- **Logistics optimisation:** Supply-chain analytics and warehouse management.
- **Security monitoring:** Automated classification labelling (already being deployed with Castlepoint Systems) and anomaly detection (MoD AI classification).
- **Corporate services:** HR, finance, and back-office support automation.

Efficiency estimates.

Metric	LOW	CENTRAL	HIGH
Headcount reduction (civilian)	2,900 (5%)	4,000 (7%)	4,600 (8%)
Annual savings	£170m	£235m	£280m
Speed gains	25–45% in corporate workflows		

Implementation timeline. Corporate-function AI is deployable within 12–18 months. Classified-environment AI requires bespoke infrastructure with a 2–3 year development timeline. Procurement reform is a long-term structural programme.

Key risks. Classification requirements limit model choice and deployment speed; MoD procurement culture is notoriously resistant to reform; supply-chain AI requires integration with complex legacy logistics systems; savings should be ring-fenced for reinvestment in frontline military capability.

2.6 Department of Health and Social Care (DHSC)

Function and scale. DHSC proper is a policy and system-management department of approximately 5,000 staff, but it oversees the NHS's £195.6 billion annual budget and 13 arm's-length bodies including NHS England. The NHS Productivity Plan aims for 2% annual productivity growth backed by up to £10 billion in technology investment by 2028/29. AI is already deployed at scale in the NHS: 2.4 million AI-assisted chest X-rays, AI-enabled stroke treatment time halved, and diagnostic AI achieving lung cancer treatment within the 62-day target rising from 52% to 100% (AI Opportunities Action Plan: One Year On; Departmental Efficiency Plans).

AI transformation opportunities.

- **Policy and briefing automation:** Copilots for the core department and ALBs.

- **Procurement and fraud management:** AI-supported contract management across the NHS estate.
- **NHS App triage and navigation:** AI-assisted patient routing and self-service.
- **Federated Data Platform analytics:** Pathway management and resource allocation.
- **NHS England integration:** The ongoing NHS England–DHSC merger, which targets significant managerial headcount reductions, can be supported by AI-enabled process consolidation.

Efficiency estimates (core DHSC and central health bodies only).

Metric	LOW	CENTRAL	HIGH
Headcount reduction	250 (5%)	400 (8%)	550 (11%)
Annual savings	£90m	£140m	£180m

Note: Wider NHS productivity gains are substantial but will show up as capacity release, shorter waits, and better clinical outcomes rather than as Treasury-bookable reductions in DHSC staff. The NHS's troubled history with national IT programmes — NPfIT was expected to cost £6.2 billion and the NAO concluded the original vision would not be realised — counsels extreme caution in any savings claims for the health system (NAO on NPfIT).

2.7 Department for Transport (DfT)

Function and scale. DfT employs 16,535 staff, combining a small policy core with large executive agencies (DVLA, DVSA, VCA) that run high-volume transactional services. DVLA alone processes 100 million transactions per year. DfT's efficiency plans identify £663 million of technical efficiencies by 2028/29 and the department has already run a voluntary exit scheme for approximately 300 central staff (Departmental Efficiency Plans).

AI transformation opportunities.

- **DVLA contact-centre and transaction automation:** AI-assisted processing of vehicle excise duty refunds, licence changes, and customer queries.
- **DVSA digital certification:** Automated route planning and test certification.
- **Transport project management:** PMO and correspondence tools for the core department.
- **Infrastructure planning support:** AI-assisted environmental and planning review for major transport schemes.

Efficiency estimates.

Metric	LOW	CENTRAL	HIGH
Headcount reduction	1,150 (7%)	1,500 (9%)	1,820 (11%)
Annual savings	£110m	£145m	£180m
Speed gains	40–70% in routine transactions		

2.8 Department for Environment, Food and Rural Affairs (Defra)

Function and scale. Defra employs 13,950 staff across grants, regulation, inspections, licensing, flood management, and environmental functions, plus animal and plant health. Its efficiency plans identify £39 million of core efficiencies plus additional Environment Agency automation benefits, including AI-enabled disease-risk management and digital compliance (Departmental Efficiency Plans; Defra Digital Sustainability Strategy 2025–2030).

AI transformation opportunities.

- **Grants processing and compliance:** Automated grant assessment, monitoring, and audit through common grants platforms.
- **Regulatory and licensing automation:** AI-assisted environmental permitting, disease-risk scoring, and inspection support.
- **Satellite-integrated monitoring:** AI analysis of earth-observation data for agricultural and environmental compliance.

Efficiency estimates.

Metric	LOW	CENTRAL	HIGH
Headcount reduction	840 (6%)	1,100 (8%)	1,400 (10%)
Annual savings	£55m	£75m	£95m
Speed gains	35–60% in grants, approvals, and compliance		

Tier 3: Policy-Intensive Departments

These smaller, more senior departments will benefit primarily from augmentation — faster drafting, better analysis, reduced external consultancy dependence — rather than transformative headcount reduction. The risk, well documented by the critiques, is that "not every saved hour becomes a deleted post." Hard vacancy controls are essential for these departments to bank any fiscal savings.

2.9 HM Treasury (HMT)

Function and scale. HMT is a small, elite policy department responsible for macroeconomic management, fiscal oversight, and public spending control.

AI transformation opportunities. Multi-model deliberation for Budget and fiscal-event preparation; automated ingestion and real-time synthesis of departmental DEL spending data; AI-assisted macroeconomic modelling; correspondence and briefing automation.

Efficiency estimates. Headcount reduction of 3–5% (LOW/CENTRAL/HIGH). Annual savings of £10–20 million. The true economic value lies in reducing macro-fiscal forecasting errors and improving spending-review analysis, not headcount.

2.10 Cabinet Office

Function and scale. 11,510 staff. Both a department and the corporate headquarters of government.

AI transformation opportunities. Cross-government PMO and performance monitoring; shared HR, finance, and procurement copilots; ministerial correspondence support; i.AI-led procurement scanning; central document search across Cabinet Office guidance and reform programmes.

Efficiency estimates.

Metric	LOW	CENTRAL	HIGH
Headcount reduction	805 (7%)	1,035 (9%)	1,265 (11%)
Annual savings	£55m	£75m	£95m

Note: The critiques correctly observe that the centre is "politically overloaded and prone to re-absorb any time saving into more challenge, more coordination and more ministerial servicing." Savings require deliberate structural choices to consolidate functions and resist re-layering.

2.11 Foreign, Commonwealth and Development Office (FCDO)

Function and scale. Approximately 17,000 staff across 281 offices worldwide. FCDO is replacing legacy IT with its Osprey platform and pursuing insourced digital capability.

AI transformation opportunities. Diplomatic cable summarisation and translation; global media and sentiment analysis; consular casework support; programme and grant monitoring; briefing-book generation. U.S. State Department experience with StateChat — over 10,000 prompts in a single day and reported savings of dozens of hours per week on media briefings — provides a directly relevant analogue (State Magazine).

Efficiency estimates.

Metric	LOW	CENTRAL	HIGH
Headcount reduction	680 (4%)	935 (5.5%)	1,190 (7%)
Annual savings	£45m	£65m	£85m

Key risks. Security classification and overseas-network variability are first-order constraints. Many posts operate in low-bandwidth, high-security environments where cloud-based AI tools are simply not available. Savings are most likely in back-office and reporting support rather than front-line diplomatic presence.

2.12 Department for Science, Innovation and Technology (DSIT)

Function and scale. Small, senior, specialist department managing large CDEL budgets through UKRI. DSIT led the 20,000-user Copilot experiment and plans £7 million per year in AI-driven savings by 2028/29 (Departmental Efficiency Plans).

AI transformation opportunities. Policy and legislation drafting copilots; AI procurement and assurance workflow automation; grant vetting for UKRI; coding assistants and shared-service expansion.

Efficiency estimates. Headcount reduction of 4–7%. Annual savings of £10–20 million. DSIT's primary role is to drive adoption elsewhere, not to generate large internal savings.

2.13 Department for Culture, Media and Sport (DCMS)

Function and scale. Approximately 2,000 staff. Policy-heavy and grant-heavy.

AI transformation opportunities. Grant and ALB monitoring; communications drafting; consultation and correspondence summarisation; shared HR, finance, and CRM automation.

Efficiency estimates. Headcount reduction of 6–10%. Annual savings of £12–25 million.

2.14 Department for Business and Trade (DBT)

Function and scale. Approximately 6,000 staff. DBT already uses Redbox as an internal LLM tool for summarisation, search, and drafting (DBT Redbox Algorithmic Transparency Record).

AI transformation opportunities. Negotiation-text comparison; export-support triage; market-intelligence synthesis; ministerial correspondence and briefing copilots; investor and grants CRM automation.

Efficiency estimates.

Metric	LOW	CENTRAL	HIGH
Headcount reduction	360 (6%)	480 (8%)	600 (10%)
Annual savings	£50m	£70m	£95m
Speed gains	30–55% in drafting and knowledge work		

2.15 Department for Energy Security and Net Zero (DESNZ)

Function and scale. Approximately 5,000 staff. Highly technical and document-intensive, with specific efficiency plans citing £19 million of AI-driven workforce productivity gains by 2028/29 (Departmental Efficiency Plans).

AI transformation opportunities. Consultation summarisation; counter-fraud and grants analytics; planning and consenting document review; policy drafting and impact-assessment support.

Efficiency estimates. Headcount reduction of 5–8%. Annual savings of £20–40 million.

2.16 Ministry of Housing, Communities and Local Government (MHCLG)

Function and scale. Approximately 4,000 staff. MHCLG's efficiency plan expects AI to save about 500,000 staff hours per year. The planning AI tool targeting 50% faster processing for straightforward applications is being scaled nationally from spring 2026 (AI Opportunities Action Plan: One Year On).

AI transformation opportunities. Planning-document extraction and application validation; grant and programme monitoring; cladding and remediation case processing; consultation analysis.

Efficiency estimates. Headcount reduction of 6–9%. Annual savings of £35–70 million. The larger political story is that AI-enabled productivity pushed into local government and planning systems yields national productivity gains even if cash savings sit outside the departmental wage bill.

2.17 Department for Education (DfE)

Function and scale. Approximately 8,000 core administrative and policy staff. Over 95% of DfE's budget is provided to frontline providers. DfE's efficiency plan identifies £20 million of digital efficiencies by 2028/29 (Departmental Efficiency Plans).

AI transformation opportunities. Funding and grants triage; AI-assisted consultation analysis and policy drafting; Ofsted and academy-related evidence summarisation; routine stakeholder correspondence; evidence synthesis for curriculum, skills, and labour-market policy.

Efficiency estimates. Headcount reduction of 5–8%. Annual savings of £45–85 million.

Tier 4: Legal/Specialist Functions and Executive Agencies

2.18 Attorney General's Office / Crown Prosecution Service (AGO/CPS)

Function and scale. 11,360 staff (predominantly CPS). The CPS has been an early adopter of AI: its Correspondence Drafting Tool, developed with NTT Data, uses LLMs to pre-populate correspondence from the CPS case management system, with a published algorithmic transparency record. CPS received £96 million in additional funding for 2026–2029, partly dedicated to AI-enabled digital development. The witness expense claims system reduced repayment time from weeks to approximately one day (CPS Correspondence Drafting Tool — Algorithmic Transparency Record; Spending Review 2025).

AI transformation opportunities. Evidence triage and disclosure support; drafting support for routine legal documents; search and retrieval across case law, internal guidance, and precedents; translation, transcription, and exhibit organisation.

Efficiency estimates.

Metric	LOW	CENTRAL	HIGH
Headcount reduction	570 (5%)	795 (7%)	910 (8%)

Metric	LOW	CENTRAL	HIGH
Annual savings	£40m	£55m	£75m
Speed gains	Substantial in disclosure prep and legal research		

Key risks. Legal functions are document-rich but legitimacy-sensitive. AI should reduce review burden, not determine prosecutorial judgment. Disclosure is a particularly sensitive area where automated systems must not create the appearance of prosecutorial shortcuts.

2.19 HM Land Registry, Companies House, and Other Executive Agencies

Function and scale. Grouped executive agencies collectively represent approximately 50,000 staff across transaction-heavy, data-heavy, and rule-bound operations. HM Land Registry (~5,500 staff) launched its Strategy 2025+ with an explicit commitment to automating many straightforward transactions by 2030 and near-instant register updates by 2035. Companies House is modernising under the Economic Crime and Corporate Transparency Act 2023. DVLA processes 100 million transactions per year (HM Land Registry Strategy 2025+; DVLA Business Plan 2025–2026).

AI transformation opportunities. Document extraction from filings and deeds; identity and anomaly checks; fraud-risk scoring; contact-centre and correspondence copilots; backlog triage and case summarisation; OCR of legacy records; digital registration and filing automation.

Efficiency estimates.

Metric	LOW	CENTRAL	HIGH
Headcount reduction	4,000 (8%)	5,500 (11%)	7,000 (14%)
Annual savings	£120m	£170m	£220m
Speed gains	40–75% in routine processing		

Assessment. This is the cleanest proving ground for government AI. These agencies provide visible, politically popular wins — faster service, lower fraud, less bureaucracy. They should be near the front of the deployment queue and serve as the demonstration effect for the broader programme.

Part 3: Aggregate Fiscal Impact

3.1 Summary Table: All Departments, All Scenarios

Department	Headcount	LOW (posts)	CENTRAL (posts)	HIGH (posts)	LOW (£m/yr)	CENTRAL (£m/yr)	HIGH (£m/yr)
Tier 1							
HMRC	70,925	5,300	7,100	9,200	265	355	460
DWP	96,890	6,800	9,700	13,500	340	485	675
MoJ	96,210	5,800	8,200	11,500	290	410	575
Home Office	50,655	3,000	4,300	6,100	150	215	305
Tier 2							
MoD (civilian)	57,500	2,900	4,000	4,600	170	235	280
DHSC (core)	~5,000	250	400	550	90	140	180
DfT	16,535	1,150	1,500	1,820	110	145	180
Defra	13,950	840	1,100	1,400	55	75	95
Tier 3							
HMT	~2,500	75	100	125	10	15	20
Cabinet Office	11,510	805	1,035	1,265	55	75	95
FCDO	~17,000	680	935	1,190	45	65	85
DSIT	~2,000	80	120	140	10	15	20
DCMS	~2,000	120	160	200	12	18	25
DBT	~6,000	360	480	600	50	70	95
DESNZ	~5,000	250	325	400	20	30	40
MHCLG	~4,000	240	320	360	35	50	70
DfE	~8,000	400	560	640	45	65	85
Tier 4							
AGO/CPS	11,360	570	795	910	40	55	75
Executive agencies	~50,000	4,000	5,500	7,000	120	170	220
TOTAL	**~527,035**	**~33,620**	**~46,630**	**~61,500**	**~1,912m**	**~2,693m**	**~3,580m**

3.2 Adjusting to Defensible Estimates

The raw departmental sum above shows a wider range than we present as our headline. This is because the HIGH scenario aggregates optimistic assumptions across all departments simultaneously — an outcome that is statistically implausible. Conversely, the LOW scenario aggregates all pessimistic assumptions. For a headline estimate intended to withstand Treasury, NAO, and PAC scrutiny, we apply the following adjustments:

1. Trim the tails. The realistic range is the LOW-to-HIGH departmental sum adjusted to exclude the probability of all departments simultaneously achieving best or worst case. We use the 25th-to-75th percentile range as our headline.

2. Add non-wage savings. Beyond direct headcount savings, three additional savings streams are credible:

Source	Central Estimate	Comment
Reduced contractors and consultancy	£500m	Consistent with SR2025 target to halve consultancy vs. 2017/23 average
Estate, postage, printing, admin overhead	£300m	HMRC post reduction alone targets £50m/yr
Error, rework, and throughput gains	£200m	Partially cashable; often redeployed before banked

3. Result: Our headline estimates.

Metric	LOW	CENTRAL	HIGH
Net headcount reduction	**34,000**	**40,500**	**47,000**
Gross annual savings (all sources)	**£2.3bn**	**£3.0bn**	**£3.8bn**

Methodology. The CENTRAL estimate of 40,500 posts represents approximately 7.4% of the civil service in scope. This is calibrated to the administration-budget problem the government is actually trying to solve. The Institute for Government estimates that meeting the 16% administration-budget target requires 29,000–40,000 fewer staff. Our CENTRAL estimate sits at the top of that range but adds a modest uplift for AI-enabled operational efficiencies beyond the pure administration budget. The gap between the IfG estimate and ours is accounted for by AI-driven improvements in operational delivery roles (particularly in Tier 1 departments) that go beyond the administration budget settlement.

The blended average saving per post of approximately £50,000 (including employer costs and overhead) applied to 40,500 posts yields approximately £2.0 billion in direct wage-bill savings. Adding £0.5 billion in contractor/consultancy savings, £0.3 billion in overhead, and £0.2 billion in partially cashable throughput gains produces the £3.0 billion CENTRAL figure.

3.3 Investment Requirements

Investment Category	5-Year Total (£bn)	Comment
AI licences, integration, and platform development	0.6–0.9	Common AI Platform, departmental tools, API costs
Data engineering and interoperability	0.3–0.5	Data quality, interoperability layers, API development

Investment Category	5-Year Total (£bn)	Comment
Legacy system remediation	0.5–0.8	NAO identified 228 legacy systems, 28% red-rated; prerequisite for AI
Security infrastructure	0.2–0.4	Classified environments for SECRET+ deployments
Training, change management, workforce transition	0.4–0.6	Retraining, process redesign, organisational change
Voluntary exit scheme costs	0.4–1.0	Based on £30,000–£60,000 average per exit; 10,000–15,000 exits
Total	**£2.4–4.2bn**	**Central estimate: £3.2bn**

The investment range of £2.4–4.2 billion over five years is deliberately wider than a single-point estimate because the largest variable — legacy system remediation — is genuinely uncertain. The NAO's cyber resilience report found that 53% of legacy systems lack fully funded remediation plans, and departments had previously estimated nearly half of the £4.7 billion annual IT spend was used merely keeping old systems running (NAO Government cyber resilience). Paper Y's investment estimate of £3.8–5.2 billion is more realistic than Paper X's £1.8–2.6 billion for a serious cross-government delivery programme; our range reflects this.

3.4 Net Annual Savings After Investment

Scenario	Gross Annual Savings	Annualised Investment	Net Annual Savings
LOW	£2.3bn	£0.48bn/yr	£1.2bn
CENTRAL	£3.0bn	£0.64bn/yr	£2.0bn
HIGH	£3.8bn	£0.84bn/yr	£2.8bn

Note: Annualised investment divides the 5-year total evenly; in practice, costs are front-loaded and savings back-loaded. The programme is unlikely to be net positive until Year 3.

3.5 Speed-of-Implementation Multipliers and Quality Gains

Beyond direct fiscal savings, AI-enabled government delivers two additional categories of benefit that, while harder to score fiscally, are substantial and politically important.

Speed gains. Across common knowledge-work tasks, a 1.2x to 1.5x productivity multiplier is already evidenced in live deployments, with larger gains in narrow workflows. The cross-government Copilot trial's 26 minutes per day implies about 13 workdays per year saved for active users. MoJ reports approximately 30 minutes per day from its AI tools, and GCS Assist saves approximately 2.8 hours per week for communications staff (Microsoft 365 Copilot Experiment; AI action plan for justice; Delivering our GCS Strategy).

For narrow transactional workflows, 2x to 5x speed gains are achievable once processes are redesigned. The Greater Cambridge Shared Planning case showed a consultation summarisation task falling from 18.5 hours to 16 minutes, though the council characterised the overall time saving more cautiously as over 60% because wider quality-assurance work remained (Greater Cambridge Shared Planning case study). Singapore's SkillsFuture AI deployment achieved a 62.5x speedup on CRM case processing. These gains matter for citizens: faster visa decisions, quicker benefit assessments, shorter waits for planning approvals, and more responsive consular services.

The indirect economic benefit is also significant. Currently, major capital infrastructure projects face severe delays in planning and environmental permitting stages, primarily within DfT, MHCLG, and Defra. If the government's £111.3 billion CDEL budget can be deployed even 5–10% faster through AI-accelerated planning and consenting processes, the avoidance of inflationary cost overruns and earlier economic stimulus effects will yield billions in secondary fiscal benefits not captured in our primary savings model.

Quality gains. The strongest evidence suggests AI can improve consistency and help lower-experience staff move faster up the learning curve. In an NBER customer-support study, productivity rose 14% on average and the largest gains were for lower-skill workers, while customer sentiment improved and attrition fell (NBER Working Paper 31161). For government, this implies three quality improvements: more consistent drafting and case preparation across caseworkers; lower training burden for junior and newly promoted staff; and fewer missed data points in large document sets. DWP's ML fraud detection model being 3x more effective than random sampling illustrates the quality dividend: better outcomes, not just faster processes.

The risk is "plausible nonsense" in nuanced or legally sensitive work. The UK Copilot trial found lower benefits in policy teams where nuance and conflicting evidence mattered most. This underscores the importance of restricting fully automated processing to rule-based, high-volume tasks and maintaining human oversight for judgment-intensive work.

3.6 Net Present Value over 5, 10, and 15 Years

Using the HM Treasury Green Book Social Time Preference Rate of 3.5% for discounting (Green Book supplementary guidance — discounting), and assuming:

- Savings ramp linearly over Years 1–5, reaching full run-rate in Year 5
- Investment costs concentrated in Years 1–3
- Post-implementation ongoing costs of £300–400 million per year from Year 4
- No real-terms growth in savings after Year 5 (conservative)

Horizon	LOW NPV	CENTRAL NPV	HIGH NPV
5 years	–£0.2bn	£1.5bn	£3.4bn
10 years	£4.8bn	£8.6bn	£13.1bn
15 years	£8.9bn	£14.8bn	£21.5bn

The programme is negative-to-marginal in the LOW scenario over five years but strongly positive in all scenarios over ten years. This reflects the realistic pattern of technology transformation: costs arrive on time, benefits lag.

3.7 Comparison with Historic Efficiency Programmes

Programme	Period	Claimed Savings	Post-Audit Reality	Lesson
Gershon Review	2004–08	2.5%/yr efficiency	NAO found only 38% of CSR07 examined savings were sustainable; 18% did not represent savings at all	Back-office consolidation has natural limits; shared services overspent by £0.5bn
CSR07 VfM Programme	2007–11	£35bn in "efficiencies"	NAO found persistent weaknesses in baselines, cost attribution, and evidence quality	Savings were often continuations of earlier initiatives, not step changes
SR2010 Austerity	2010–15	Large headcount reduction	Civil service fell from ~530k to ~384k; junior grades (AA/AO) fell from 46% to 24% of workforce	Cuts degraded service quality and institutional capacity; grade inflation followed
SR2015/Shared Services	2015–20	Cross-government savings	NAO found five shared service centres cost £1.4bn vs. £0.9bn planned; net cost of £255m for two centres still tracking	Voluntary adoption and excessive customisation destroyed economies of scale
AI Transformation (this paper)	2026–31	£2.3–3.8bn/yr	TBD	Must learn from all the above: mandate standardisation, measure cash release, distinguish capacity from savings

The critical lesson from this history: the British state has repeatedly over-claimed savings because it has confused process improvement with cash release, gross savings with net savings, and reported savings with sustainable savings (NAO on CSR07 value-for-money savings; NAO on shared service centres). Our estimates are deliberately conservative relative to the technological possibility precisely because institutional delivery — not AI capability — is the binding constraint.

3.8 OBR Scorability Assessment

The Office for Budget Responsibility applies rigorous tests to operational measures before certifying them as fiscal savings. Based on OBR working papers on welfare policy costings and anti-avoidance measures (OBR working paper on welfare policy costings; OBR anti-avoidance costings paper), we assess the scorability of each savings component:

Savings Component	OBR Scorability	Rationale
Direct headcount savings (posts deleted)	**Scoreable — if posts disappear and DEL is correspondingly reduced**	This is the hardest form of savings and the only one the OBR would certify without significant caveats
Reduced contractor expenditure	**Partially scoreable — if contracts are demonstrably not let**	Requires evidence of contracts cancelled, not just internal capability claims
Fraud and error prevention	**Weakly scoreable — requires robust counterfactuals and operational track record**	OBR notes "delivery risk, modelling complexity and data quality" challenges in operational costings
Productivity and throughput gains	**Not directly scoreable — unless converted to headcount or budget reductions**	Productivity improvements that are absorbed as more activity yield no fiscal saving
Avoided future recruitment	**Indirectly scoreable — through lower establishment numbers in future spending reviews**	Defensible but only visible over medium term

Implication for ministers: The publicly defensible claim should focus on headcount reduction and contractor savings, which together account for roughly £2.0–2.5 billion of the CENTRAL estimate. Fraud prevention and productivity gains are real but should be presented as additional benefits with appropriate uncertainty ranges, not as headline fiscal numbers.

Part 4: Implementation Architecture

4.1 Phase 1: Foundation (2026–2027) — Pilots, Infrastructure, Quick Wins

The first year should not be about claiming huge savings. It should be about proving that the centre can control adoption and that productivity gains are being converted into structural changes.

Infrastructure actions:

- Mandate secure AI assistants (Layer 1 tools) across all departments for approved low-risk uses at OFFICIAL classification. Estimated licence cost: £180 million/year for full civil service rollout. This builds the base of user familiarity on which Layer 2 deployment depends.
- Launch Common AI Platform v1 with managed access to 2–3 approved foundation models on UK sovereign cloud. Estimated cost: £150 million.
- Begin Data Interoperability Layer pilot connecting HMRC, DWP, and Cabinet Office data for cross-government fraud detection. Budget: £80 million.
- Commission security-assured AI environments for OFFICIAL-SENSITIVE work. Begin design of SECRET-capable classified AI analytics environment for MoD and FCDO.
- Continue HMRC SAP S/4HANA migration and ensure AI integration is designed in from the start.

Quick-win deployments (6–12 months):

- HMRC: AI-assisted enquiry handling and correspondence automation.
- DWP: Expanded ML fraud detection across all benefit lines; case-preparation copilots.
- MoJ: AI transcription in Crown Courts with completed Common Platform integration; case-bundle automation.
- Executive agencies: DVLA mail elimination programme; HM Land Registry automated straightforward transactions; Companies House identity verification automation.
- Cross-government: i.AI Consult tool scaled to all government consultations (saving an estimated £20 million/year and 75,000 person-days).

Workforce actions:

- Force every department to publish a task inventory covering correspondence, drafting, search, transcription, case preparation, and contact-centre workflows — the raw material for identifying automation opportunities.
- Create a Treasury–Cabinet Office–DSIT control regime under which departments only keep AI-related productivity gains if they delete equivalent vacancies or contractor posts.
- Negotiate a workforce compact with Prospect, FDA, and PCS: no general compulsory redundancy programme in exchange for faster retraining, internal mobility, and stronger attrition management.

- Implement a strict freeze on external recruitment for AA/AO administrative grades in departments with AI deployment plans.

Expected outcome. 8,000–12,000 post reductions through managed attrition, voluntary exits (building on the 8,586 already targeted), and recruitment restrictions. Small hard savings; visible service improvements; much better data on adoption rates and real-world time savings.

Phase 1 investment: £0.8–1.1 billion.

4.2 Phase 2: Scale (2028–2029) — Departmental Rollout, Workforce Transition

This is where real savings begin. The focus shifts from personal productivity tools to mission-workflow redesign.

Infrastructure actions:

- Common AI Platform v2 with fine-tuning capabilities, agentic AI frameworks, and departmental sandboxes. Budget: £200 million.
- Data Interoperability Layer full deployment across all 19 departments. Budget: £250 million.
- Classified AI analytics environment operational for MoD, FCDO, and intelligence community. Budget: £300 million.
- Legacy system remediation programme targeting the most critical of the 228 legacy systems identified by NAO. Budget: £500 million. This is not an optional add-on; it is the central delivery constraint. AI amplifies bad data and brittle systems.

Scaling deployments:

- HMRC: Automated compliance risk scoring across all tax heads.
- DWP: Full case-preparation automation across all benefit lines; expanded fraud and error analytics.
- MoJ: AI transcription across all Crown Courts; tribunal case triage.
- Home Office: Automated visa decision-making for straightforward categories (visitor, routine work permits).
- MHCLG: National scale-up of planning AI tools across local authorities.
- Executive agencies: HM Land Registry automated routine transactions; DVLA full contact-centre AI.

Workforce actions:

- Move from recruitment restrictions to active workforce reshaping through targeted voluntary exit schemes in departments with fully automated workflows.
- Launch intensive retraining programmes for EO and HEO staff transitioning from data processing to AI quality assurance, oversight, and escalation roles. Budget: £200 million.
- Introduce mandatory AI-enabled operating models for contact centres, correspondence units, and document-processing hubs.

- Create departmental AI boards tied to spending approvals — no new programme business case approved without AI integration assessment.

Expected outcome. An additional 14,000–20,000 posts removed, bringing the running total to 22,000–32,000. Annual gross savings approaching £1.5–2.5 billion. At current civil service turnover rates — approximately 38,600 leavers per year (Civil Service Statistics 2025) — natural attrition alone can accommodate most of this reduction if recruitment is managed carefully.

Phase 2 investment: £1.0–1.5 billion.

4.3 Phase 3: Transformation (2030–2031) — Full Integration, Continuous Improvement

By this point, government should begin redesigning the administrative state around AI-native workflows.

Infrastructure actions:

- Common AI Platform v3 with multi-agent orchestration, cross-departmental AI workflows, and real-time policy simulation capabilities.
- Sovereign UK language models fine-tuned for government use, hosted on UK compute infrastructure. The £500 million Sovereign AI Fund provides the policy vehicle (GOV.UK).
- Government AI evaluation and assurance infrastructure at production scale.

Transformation deployments:

- End-to-end AI-powered citizen journeys across departments ("tell us once" for life events — birth, bereavement, retirement — with AI coordinating across HMRC, DWP, DHSC, HM Land Registry).
- AI-powered real-time performance monitoring replacing annual reporting cycles.
- MoD classified AI analytics at operational scale.
- Multi-model deliberation systems in HMT, Cabinet Office, and FCDO for fiscal, legislative, and diplomatic analysis.
- Grade structure rationalisation to reverse grade inflation and reduce unnecessary checking layers.

Expected outcome. Total programme target of 34,000–47,000 posts achieved. Full run-rate savings of £2.3–3.8 billion per year.

Phase 3 investment: £0.6–1.0 billion.

4.4 Central Delivery Unit Design

The programme requires a central delivery unit with genuine authority. Based on the lessons of past efficiency programmes — where "the absence of clear management by the Government" was identified by the NAO as a root cause of failure — we recommend:

AI Transformation Unit (AITU)

- **Reporting line:** Jointly to the Prime Minister and the Chancellor, with the Chief Secretary to the Treasury as lead minister.
- **Leadership:** A Director General-level head recruited from outside the civil service, with proven experience of large-scale technology transformation in regulated environments.
- **Staff:** 80–120 people combining Whitehall delivery expertise, digital/data professionals, and commercial specialists.
- **Powers:**
- Authority to set binding adoption targets for each department.
- Sign-off on departmental AI investment business cases above £5 million.
- Power to direct vacancy freezes in functions with proven AI automation.
- Monthly reporting to the PM's delivery board and quarterly reporting to the Public Accounts Committee.
- **Sunset clause:** Review after three years to determine whether functions should be absorbed into Cabinet Office or DSIT.

4.5 Workforce Transition: Managed Attrition, Retraining, Redeployment

The workforce transition strategy must be built on six principles that reflect the political economy of civil service reform:

1. Managed attrition first. The civil service turnover rate is 7.1%, with approximately 38,600 people leaving annually through resignation (19,420), retirement (10,915), and other channels (Civil Service Statistics 2025). In a five-year programme, this natural churn creates scope for approximately 80,000–90,000 replacement decisions. If even 40–50% of vacated posts in affected functions are not refilled, the central-scenario headcount target of 40,500 is achievable without any compulsory redundancies.

2. Redeploy before exit. Some departments will need more digital, data, assurance, and service-design capacity even while shrinking clerical and managerial layers. Redeployment within and across departments should be the default before external exit.

3. Target middle management, not just junior admin. The civil service grade structure has drifted upward since 2010: AA/AO grades fell from 46% to 24% of the workforce, while HEO/SEO grades grew by 57% and G7/G6 grades from 7% to 17% (Whitehall Monitor 2026). AI strengthens the case for fewer layers of review and sign-off, wider spans of control, and a flatter hierarchy. The old fantasy that automation simply strips out junior clerical work is dated; there are fewer genuinely routine administrative layers to remove than there were in 2010.

4. Voluntary exits selectively and generously. We model 10,000–15,000 voluntary exits over five years at an average cost of £35,000–£60,000 per exit — a total of £0.4–0.9 billion. The SR2025 provision of £150 million for exit schemes is a start but will need topping up in subsequent spending rounds.

5. Protect frontline legitimacy. Where public trust is at stake (welfare decisions, immigration, justice), government should redeploy some savings into better human casework and appeals capacity. The narrative must be "better services" not "fewer people helping you."

6. Measure actual deletion of roles. Time saved is not reform. Posts deleted from establishment tables, contractors removed from departmental accounts, and budget lines reduced in the next spending review — these are the only metrics that count.

4.6 Procurement Strategy: Sovereign Capability vs. Vendor Partnerships

The procurement strategy must navigate between two risks: over-dependence on a small number of cloud/AI hyperscalers, and the cost and complexity of building sovereign capability from scratch.

The CMA's cloud market investigation found competition concerns, high concentration, and technical and commercial barriers that lock customers into initial providers (CMA cloud services market investigation). The NAO warned that government spends at least £14 billion annually on digital procurement and still lacks aligned responsibility, skills, and resources to manage digital procurement well, especially with major suppliers (NAO on technology suppliers).

Our recommended approach:

1. Multi-vendor, multi-model. No single provider should supply more than 40% of the government's AI compute or model access. The Common AI Platform should support multiple foundation models and cloud providers.

2. Open standards and interoperability. All AI systems procured for government must use open data formats and APIs to enable switching. Model-agnostic architectures should be mandated.

3. Sovereign capability for critical functions. The £500 million Sovereign AI Fund should support development of UK-hosted models for the most sensitive government applications. This is a national security requirement, not primarily an efficiency measure.

4. Intelligent client capability. Departments must build or share commercial and technical expertise to act as intelligent buyers. The NAO's finding that departments "often award digital contracts before fully understanding what they need" must not be repeated.

5. G-Cloud and framework evolution. Existing procurement frameworks need adaptation for AI-specific requirements: model performance guarantees, data sovereignty clauses, and regular benchmarking against alternative providers.

4.7 Parliamentary Oversight Mechanism

AI transformation of government is a constitutional change as much as a management reform. It requires new parliamentary oversight:

1. Annual AI Transparency Report to Parliament. The government should publish an annual report detailing: all AI systems in use across departments, their purposes, performance metrics, error rates, and complaint volumes. This should be laid before the House and debated.

2. Select Committee capacity. The Science, Innovation and Technology Committee and the Public Accounts Committee should jointly appoint an independent technical advisor with AI expertise to support scrutiny of government AI systems.

3. Algorithmic Transparency Recording Standard enforcement. The current voluntary compliance — only 33 records published despite hundreds of AI contracts — is inadequate. Publication should become a statutory requirement for all AI systems that inform decisions affecting citizens' rights or entitlements.

4. Ministerial accountability for AI errors. Ministers must answer to Parliament for AI-related failures with the same accountability as for human errors. The doctrine of ministerial responsibility is unchanged; only the technology is new.

5. NAO mandate. The NAO should be given an explicit mandate and resources to audit AI systems used in government, including model performance, bias, and value for money. The NAO's existing work on digital transformation, cyber resilience, and fraud provides a natural foundation.

Part 5: Risks, Failure Modes, and Mitigations

5.1 Technical Risks

Legacy systems and data quality. The NAO identified at least 228 legacy IT systems in March 2024, 28% rated high-risk, with 53% lacking fully funded remediation plans. Approximately £2.3 billion per year — nearly half the central IT budget — is spent keeping old systems running (NAO Government cyber resilience). AI does not solve bad data; it amplifies it. If records are incomplete, permissions dirty, APIs absent, and process maps inaccurate, the result will be a new verification layer on top of old manual work, not elegant automation.

Mitigation: Ring-fence £0.5–0.8 billion for legacy remediation as a prerequisite investment, not an afterthought. Prioritise the departments with the highest AI savings potential (HMRC, DWP, MoJ) for remediation first. Accept that some departments will lag because their systems are not ready.

AI reliability. In October 2025, Deloitte was forced to refund part of a report to the Australian government after it was found to contain AI-generated errors, including fabricated citations. This is not an isolated incident. Large language models produce "plausible nonsense" in nuanced or legally sensitive contexts. The UK Copilot trial found lower benefits in policy teams where conflicting evidence mattered most (Microsoft 365 Copilot Experiment).

Mitigation: Mandatory human review for all consequential decisions. Retrieval-Augmented Generation (RAG) architectures that ground outputs in authoritative government data. Systematic testing and evaluation infrastructure. Accept that AI accuracy is sufficient for many routine tasks but insufficient for novel policy analysis without human oversight.

Security. Classification fragmentation means that AI deployment will proceed at different speeds across government. The most sensitive departments (MoD, intelligence) will lag years behind OFFICIAL-tier departments. Model poisoning, data exfiltration, and prompt-injection attacks are genuine threats, particularly for models trained or fine-tuned on government data.

Mitigation: Tiered security architecture. Air-gapped environments for SECRET+. Regular red-teaming of deployed models. Compliance with the NCSC's guidelines for secure AI system development. Supply-chain assurance for all model providers.

5.2 Political Risks

Union opposition. Prospect has explicitly argued that AI in the civil service should be introduced "in consultation with their workforce, not imposed on them" and that modernising government requires "both technology and people rather than a trade-off between them" (Prospect on AI and tech at work). The FDA will resist centrally imposed productivity mandates that erode discretion and accountability without evidence (FDA report on the future of office work). PCS is likely to be more hostile, especially where AI is associated with job loss, surveillance, or reduced in-person support for claimants.

Mitigation: The workforce compact is not optional. No general compulsory redundancy programme. Meaningful union consultation on deployment plans. Ring-fenced retraining budgets. Honest framing: AI is being used to meet efficiency targets that already exist, not to impose new ones. The narrative must be modernisation and service improvement, not technology-driven job cuts.

Public trust. Ipsos polling for the Tony Blair Institute found that 38% of UK adults cite lack of trust in AI as a barrier to adoption, and more view AI as a risk to the economy (39%) than an opportunity (20%). Academic research on parliamentary AI finds that nearly four in five UK respondents reject the idea of AI taking decisions in place of human decision-makers (Democracy by algorithm? — Oxford Academic). A high-profile AI error affecting a vulnerable citizen could set the programme back by years.

Mitigation: Mandatory human-in-the-loop for all citizen-facing decisions. Published fairness audits. Rapid and transparent error correction. An honest public narrative about what AI can and cannot do. Invest in the "good news" — faster processing, shorter waits, better fraud detection — rather than leading with headcount cuts.

Ministerial accountability. If an MP asks why a constituent was denied a service, delayed, or misclassified, who explains the chain of reasoning? Which audit trail is intelligible enough for Parliament, the NAO, ombudsmen, and courts? Lord Sales' November 2025 address to the Supreme Court highlighted the tension between automated decision-making and administrative law requirements of exercising powers for proper purposes, giving reasons, and enabling meaningful challenge (Lord Sales on AI and Public Law).

Mitigation: Every AI system deployed in a decision-affecting role must produce an audit trail that is intelligible to a non-technical reviewer. The Algorithmic Transparency Recording Standard must be enforced, not aspirational.

5.3 Fiscal Risks

Cost overruns. The universal lesson of UK government technology programmes is that they cost more and take longer than planned. NPfIT was expected to cost £6.2 billion; Universal Credit was repeatedly delayed; shared services cost £1.4 billion against a £0.9 billion plan.

Mitigation: Agile, milestone-based funding rather than large upfront commitments. Independent assurance at each phase gate. Gateway reviews by the Infrastructure and Projects Authority (IPA). A genuine willingness to pause or cancel failing workstreams rather than throwing good money after bad.

Savings shortfall. If departments use AI-released time to do more work rather than reduce headcount, no fiscal saving materialises. This is what Whitehall usually does. The political economy of government is that ministers always have more demands than resources; freed capacity is immediately consumed by new priorities.

Mitigation: This is the most important single risk. The only defence is hard vacancy controls — the Treasury–Cabinet Office–DSIT control regime described in Phase 1. Departments must not be allowed to keep productivity gains unless they delete equivalent posts, cancel

equivalent contracts, or reduce equivalent budgets. The AITU's role is to enforce this ruthlessly.

5.4 Accountability Risks

Algorithmic bias. If an automated DWP system incorrectly denies Universal Credit to a protected group at scale, the legal settlements and reputational damage could eclipse the savings. The UK's existing equality and human rights framework applies fully to AI-assisted decisions.

Mitigation: Fairness impact assessments before deployment. Continuous monitoring for disparate impact. DWP's Universal Credit Advances fairness assessment provides a working model. Retained human oversight for all decisions with significant individual impact.

Democratic legitimacy. AI may produce faster but thinner ministerial answers; compress internal challenge; and generate plausible summaries rather than fully evidenced records. There is a genuine risk that parliamentary accountability degrades as AI intermediates more of the information flow between departments and Parliament.

Mitigation: Transparency, recordability, and disclosure duties must be designed in from the start. The AI Playbook's insistence on openness, ATRS use, and contestability should be treated as constitutional infrastructure, not optional ethics garnish.

5.5 What Happens If This Goes Wrong: Honest Assessment

Scenario 1: Savings shortfall. The most likely failure mode is not catastrophic collapse but disappointing returns. Departments adopt Layer 1 tools (email copilots, meeting notes) but fail to redesign workflows or delete posts. Savings of perhaps £0.5–1.0 billion materialise rather than £2.0–3.0 billion. The programme is declared a partial success and quietly superseded by the next spending review. This is essentially what happened to shared services.

Scenario 2: High-profile AI error. An automated welfare or immigration decision wrongly affects thousands of citizens, generating a media firestorm and parliamentary crisis. The programme is paused for review. Public trust collapses. Deployment timelines slip by 2–3 years. This is the Windrush or Horizon scenario for the AI age.

Scenario 3: Vendor lock-in. Government standardises quickly on one cloud/AI ecosystem. Initial integration costs fall, but within 3–5 years the vendor has pricing power, switching costs are prohibitive, and government finds itself in a worse commercial position than before. The CMA's cloud market findings make this a live risk.

Scenario 4: Partial success, uneven adoption. Some departments (HMRC, DWP, executive agencies) transform effectively. Others (MoD, FCDO, smaller policy departments) lag due to classification constraints, leadership resistance, or legacy systems. Total savings are 50–70% of the central estimate. This is probably the most realistic outcome based on the historical pattern of UK government reform.

Scenario 5: Political-economy absorption. AI works exactly as promised at the technical level, but ministers use the freed capacity not to shrink the state but to ask departments to do more. New demands, new reporting requirements, new controls, and new casework expectations soak up the productivity gain. This is what Whitehall usually does. It is not a failure of AI; it is a failure of political discipline. The only defence is the hard vacancy-and-budget control mechanism described in this paper.

The honest assessment is that Scenario 4 is the most likely outcome, with elements of Scenario 5 around the edges. Full programme delivery is aspirational; partial delivery with uneven adoption is realistic. The paper's estimates are calibrated accordingly — the CENTRAL scenario already assumes that not everything works as planned. A serious critic would be entitled to apply a further 20–30% delivery discount to the CENTRAL estimate, which would still leave a programme worth pursuing: even 70% of £3.0 billion is £2.1 billion in annual savings, well above the programme's cost.

Conclusion: The Choice

UK central government faces a straightforward choice over the next parliament: AI-enabled reform or managed decline.

The fiscal arithmetic is unforgiving. Administration budgets must fall by 16% in real terms. Departments have committed to almost £14 billion in annual efficiency gains by 2028/29. Without AI-enabled productivity improvements, these targets can only be met by cutting services that the public relies upon — the pattern that made austerity politically toxic and operationally destructive after 2010.

AI offers a different path: a state that processes routine cases dramatically faster, detects fraud more effectively, drafts policy with broader evidence, and frees human judgment for the tasks that genuinely require it. The technology is proven in UK government settings. The institutional infrastructure exists. The fiscal pressure is real and immediate.

But this will only work if three conditions are met.

First, the centre must enforce. AI productivity gains that are not converted into structural headcount and budget reductions are not savings; they are additional capacity that the system will absorb. A central AI Transformation Unit with real authority over vacancy controls, adoption targets, and investment decisions is essential. Without it, the programme will become another collection of pilots that departments announce, showcase, and fail to scale.

Second, the workforce compact must be genuine. The government cannot credibly claim to be different from the crude headcount cuts of the 2010s unless it demonstrates a genuine commitment to managed attrition, retraining, and redeployment. This costs money — perhaps £0.4–0.9 billion in voluntary exit schemes and £0.4–0.6 billion in retraining over five years. Unions should be brought in as partners in modernisation, not treated as obstacles to be overcome.

Third, ministers must be honest about what AI cannot do. It cannot solve problems caused by underfunding, political indecision, or bad law. It cannot substitute for judicial capacity in the courts, clinical judgment in the NHS, or diplomatic skill in foreign policy. It cannot be deployed in the most sensitive classified environments at the same pace as routine administrative work. And it will sometimes fail — producing errors, exhibiting bias, or delivering less than promised. The government that admits this openly will find it much easier to sustain public and parliamentary support than one that oversells.

The credible prize is not a fantastical halving of Whitehall. It is a state that by 2031 has around 40,000 fewer posts than it otherwise would, spends around £2–3 billion less each year on administration and related overheads, processes routine cases dramatically faster, and redirects scarce human judgment to the tasks that actually require it. Over ten years, the net present value of this programme is £4.8–£13.1 billion — a return that comfortably justifies the investment, even with aggressive discounting for delivery risk.

The alternative is not the status quo. The alternative is managed decline: a civil service forced to cut 16% of its administration budget through crude means, degrading institutional memory, losing specialist capability, and delivering worse services to the public. That is the path the government is already on if AI transformation fails or is never seriously attempted.

This paper has been deliberately conservative. We have taken the most cautious of three independent research papers as our fiscal spine. We have applied the lessons of every major efficiency programme since Gershon. We have honestly acknowledged the technical, political, and fiscal risks. We have presented failure scenarios alongside success scenarios. We have distinguished between what is scoreable by the OBR and what is not. If there is a criticism to be made of this analysis, it is that we may have been too cautious about the medium-term potential of AI to reshape not just the efficiency but the effectiveness of the British state.

What ministers should watch over the next 12 months is not the number of pilots announced but five hard indicators: adoption rates in live tools, vacancies deleted rather than refilled, contractors removed, workflow cycle times, and error or appeals rates. If those numbers move together, AI will be rewiring the state. If they do not, Whitehall will simply have found a new way to write the same memo faster.

The choice is not whether to deploy AI in government. That is already happening, department by department, tool by tool. The choice is whether to do it with strategic coherence, fiscal discipline, and political honesty — or to drift into a patchwork of pilots that never scales, never delivers cashable savings, and never transforms the machinery of the state. History suggests the latter is the default. This paper is written for those who want to choose differently.

Centre for a Better Britain, March 2026

This paper was produced through a synthesis process involving three independently authored research papers, two cross-model critique panels (editorial review and Treasury/NAO stress test), and supplementary research. The fiscal spine follows the most conservative of the three research papers, adjusted upward where the delivery-architecture analysis provides stronger evidence, and downward where the critique panels identified overstatement. All estimates are designed to withstand scrutiny from Treasury officials, the National Audit Office, the Public Accounts Committee, and informed policy journalists.

Annex: Key Sources

This paper draws on the following primary sources, among others cited in-text:

- UK Government, Spending Review 2025
- UK Government, Departmental Efficiency Plans — SR2025
- UK Government, AI Opportunities Action Plan: One Year On
- UK Government, AI Playbook for the UK Government
- UK Government, Microsoft 365 Copilot Experiment — Cross-Government Findings Report
- UK Government, Civil Service Statistics 2025
- Institute for Government, Whitehall Monitor 2026
- National Audit Office, Government Cyber Resilience
- National Audit Office, Government's Approach to Technology Suppliers
- National Audit Office, CSR07 Value-for-Money Savings
- National Audit Office, Shared Service Centres
- National Audit Office, Progress in Implementing Universal Credit
- OBR Working Paper on Welfare Policy Costings
- HM Treasury, Green Book — Discounting Supplementary Guidance
- CMA, Cloud Services Market Investigation
- UK Government, Government Security Classifications — SECRET
- UK Government, Government Security Classifications — TOP SECRET
- MoJ, AI Action Plan for Justice
- HM Land Registry, Strategy 2025+
- Prospect, AI and Tech at Work
- FDA, Future of Office Work in the Civil Service