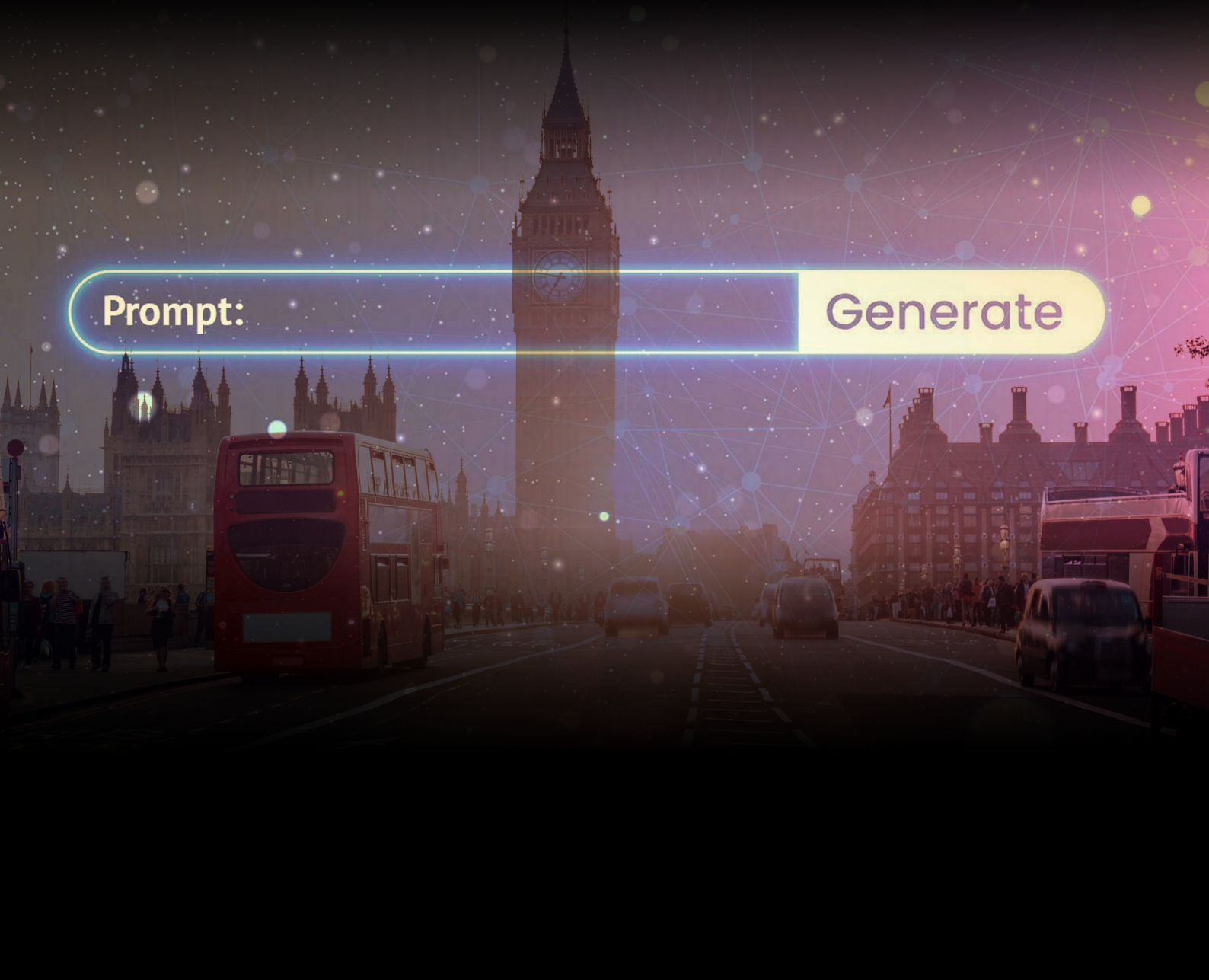
UNIVERSITY OF CAMBRIDGE

Bennett Institute for Public Policy Cambridge

Using Large Language Models responsibly in the civil service: a guide to implementation

Aleksei Turobov



Author
Aleksei Turobov, Research Associate of the AI and Geopolitics Project (AIxGEO)

Suggested citation
Turobov, A. (2025). 'Using Large Language Models responsibly in the UK civil service: a guide to implementation.' Bennett Institute for Public Policy, University of Cambridge.

Acknowledgement
We are grateful to members of the UK Civil Service from different departments for the incredibly detailed and thoughtful comments and recommendations on earlier versions of this policy resource.

About the Bennett Institute for Public Policy

The Bennett Institute for Public Policy at the University of Cambridge is committed to interdisciplinary academic and policy research into the major challenges facing the world, and to high-quality teaching of the knowledge and skills required in public service. Our research connects the world-leading work in technology and science at Cambridge with the economic and political dimensions of policymaking. We are committed to outstanding teaching, policy engagement, and to devising sustainable and long-lasting solutions.
bennettinstitute.cam.ac.uk

Contents

1. Introduction

UK civil servants, along with officials at other levels of government and in other countries, face a potentially transformative moment in the adoption of digital tools with the availability of Large Language Models (LLMs). As these powerful Artificial Intelligence (AI) systems reshape how organisations process information and deliver services, civil servants need to navigate unprecedented opportunities for enhanced public service delivery and complex challenges of responsible implementation. Given their capabilities, the use of LLMs can enable efficiencies including speeding up some tasks such as evidence synthesis or summarising very large numbers of documents. The integration of LLMs into civil service operations occurs within an established framework of accountability, data protection, and service standards. Civil servants face the challenge of harnessing these powerful new tools while maintaining their high standards or reliability and accountability. This challenge is particularly acute given the pressure to improve efficiency and effectiveness in public service delivery while ensuring robust governance and maintaining public trust.

This paper complements the Generative AI Framework for HMG[1], providing detailed implementation approaches specifically for LLM integration. While the HMG Framework establishes overarching principles for generative AI use across government, this guidance serves as a practical framework for understanding and implementing LLMs within civil service, aiming to bridge the gap between technological potential and practical implementation, providing civil servants with clear, actionable insights for responsible LLM integration.

This guidance primarily focuses on civil service policy development, analysis, and administrative functions in the UK, while acknowledging distinct applications in frontline service delivery. Different civil service functions – from policy development to public-facing services – may require varied approaches to LLM implementation. While examples here primarily address policy and analytical work, the principles can be adapted for service delivery contexts, and indeed for contexts at other levels of government and outside the UK.

Key objectives of this guidance:

- Enable informed decision-making about the use of LLMs
- Ensure alignment with civil service values and standards
- Provide practical frameworks for responsible deployment
- Support effective risk management and governance
- Guide the development of institutional capacity for LLM adoption

This guidance balances theoretical understanding with practical implementation. The initial sections establish crucial foundations about LLM capabilities and limitations, directly informing the practical frameworks and tools provided in later sections. This structure ensures that implementation guidance is grounded in a solid understanding of the technology's potential and constraints.

This implementation guide builds upon established government frameworks such as AI guide in the public sector[2], implementation planning[3], AI assessment framework[4], AI ethics and safety[5] and managing AI projects[6] while addressing specific LLM integration needs.

1. Generative AI Framework for HMG
   URL: https://www.gov.uk/government/publications/generative-ai-framework-for-hmg
2. A guide to using artificial intelligence in the public sector.
   URL: https://www.gov.uk/government/publications/understanding-artificial-intelligence/a-guide-to-using-artificial-intelligence-in-the-public-sector
3. Planning and preparing for artificial intelligence implementation.
   URL: https://www.gov.uk/guidance/planning-and-preparing-for-artificial-intelligence-implementation
4. Assessing if artificial intelligence is the right solution.
   URL: https://www.gov.uk/guidance/assessing-if-artificial-intelligence-is-the-right-solution.
5. Understanding artificial intelligence ethics and safety.
   URL: https://www.gov.uk/guidance/understanding-artificial-intelligence-ethics-and-safety
6. Managing your artificial intelligence project.
   URL: https://www.gov.uk/guidance/managing-your-artificial-intelligence-project

LLM implementation should support and enhance civil servants' ability to uphold the Civil Service Code's core values[7].

1.1 Understanding LLMs: core principles for use

LLMs represent a sophisticated convergence of pattern recognition, probabilistic prediction, and contextual processing. Understanding these fundamental mechanisms enables civil servants to make informed decisions about their implementation and use.

At their heart, LLMs are probabilistic systems. Every output is generated through a series of probability calculations. The model assesses the likelihood of different words or phrases appearing in a given context, making choices based on patterns learned during training. This probabilistic foundation has crucial implications for government use:

- Outputs represent statistical predictions and not logical deductions
- Results can vary even with identical inputs
- Confidence levels in the outputs differ across tasks and contexts

Understanding this probabilistic nature is essential for appropriate task selection, risk assessment, quality control measures and output validation protocols.

LLMs operate through a *transformer architecture*, which enables them to process language in a way that captures both meaning and context. This foundation allows the models to process text by converting words into mathematical representations (*embeddings*) that capture their meaning and relationships. These representations help the model understand language in a way that enables sophisticated analysis and generation tasks.

The transformer architecture employs attention mechanisms that allow the model to:

- Consider relationships between all words simultaneously
- Weigh the importance of different contextual elements
- Maintain consistency across long passages of text

LLMs can be adjusted to better suit different tasks through various settings, much like adjusting the settings on any other tool. Two key types of adjustments are available:

1. Response style settings - control how the LLM generates its responses:

- More focused and consistent outputs (like when writing formal documents)
- More varied and creative outputs (useful for brainstorming or exploring options)

Think of this as adjusting the tone of a conversation - sometimes, you need formal and precise language and other times, you need more creative and exploratory approaches.

2. Technical settings - control practical aspects of how the LLM works:

- How much information it can process at once
- How detailed its responses should be
- How it structures its outputs

These adjustments (technically known as 'temperature control' and 'model parameters' using an Application Programming Interface - API) help ensure the LLM's outputs match your specific needs.

7. The Civil Service code.
URL: https://www.gov.uk/government/publications/civil-service-code/the-civil-service-code

Temperature control

Lower settings (0–0.3)	Higher settings (0.7–1.0)
Produce consistent, predictable responses	Generate more diverse outputs
Ideal for fact-based tasks and formal documentation	Suitable for exploratory tasks
Reduce creative variation	Increase creative variation

Model parameters:

- *Context window size* affects document processing capacity
- *Token limits* influence response length
- Sampling methods affect output diversity
- Response formatting options enable structured outputs

1.2 Considerations for use

LLMs can be applied across various civil service functions, though implementation approaches may differ between policy development, administrative work, and service delivery contexts. LLMs operate within specific boundaries, which directly affects the scope for their practical application in official functions. Most fundamentally, they work with fixed knowledge cutoff dates and training parameters that define their expertise boundaries. This limitation requires careful consideration in government contexts, where current and accurate information is crucial for decision-making/political processes.

The resource implications of LLM deployment present another critical consideration for civil service implementation. Processing demands increase proportionally with input volume, while response times vary based on task complexity. These factors significantly influence both system efficiency and operational capabilities. For instance, while analysing lengthy policy documents or managing multiple departmental queries, civil servants should consider these computational scaling relationships to optimize resource utilisation and ensure efficient task completion. This requires balancing processing demands against the practical value of outputs - larger inputs require more computational resources and time, necessitating careful assessment of the cost-benefit relationship for each task.

To effectively implement LLMs within government operations, departments should adopt a structured approach to deployment and integration. This begins with a comprehensive pre-deployment assessment examining task suitability and resource requirements within specific departmental contexts. Security considerations are paramount, requiring careful alignment with existing government protocols and the establishment of robust quality control mechanisms[8].

8. An additional useful framework could be the Responsible Innovation Model.
URL: https://www.gov.uk/government/publications/the-model-for-responsible-innovation

Successful integration depends on developing clear procedures and guidelines that reflect technical capabilities and civil service requirements. This includes establishing validation protocols, maintaining detailed documentation of usage patterns, and implementing regular performance monitoring systems. The integration process should incorporate feedback mechanisms that enable continuous improvement while maintaining high standards of public service delivery.

These considerations and implementation frameworks enable civil servants to make informed decisions about LLM deployment while ensuring responsible and effective use. By understanding these fundamental aspects, departments can:

- Develop appropriate deployment strategies aligned with their specific needs
- Implement effective control measures that maintain civil service standards
- Create robust protocols for ongoing monitoring and evaluation
- Ensure sustainable and responsible integration across government functions

The following sections build upon these foundational principles, providing detailed guidance for specific applications, risk management strategies, and implementation best practices. This guidance will help civil servants leverage LLM capabilities effectively while maintaining the high standards expected of government operations.

2. Framework for implementation

The implementation framework provided here serves as a practical roadmap for civil servants integrating LLMs into existing work processes. It ensures alignment with civil service standards while maintaining institutional requirements. The integration of LLMs into civil service work represents a journey rather than a fixed destination. While the detailed framework that follows provides comprehensive guidance, its application should be viewed through the lens of your specific departmental context and needs.

Note: While this framework primarily addresses policy and analytical functions, departments can adapt it for service delivery contexts, considering specific requirements of public-facing services.

The framework operates through three complementary layers that together create a comprehensive approach to LLM adoption:

1. Established civil service process (upper layer - grey blocks) - represents standard departmental workflows, from initial review through to final validation. This ensures LLM implementation aligns with existing practices and maintains institutional standards.

2. Governance requirements (middle layer - red blocks) - embeds essential civil service obligations:
- Information security protocols
- Data protection requirements
- Classification standards
- Compliance mechanisms

3. LLM implementation guide (lower layer - green blocks with prompt examples in yellow blocks) - provides practical instruction for:
- Task-specific prompting
- Quality assurance steps
- Validation requirements
- Error detection procedures

LLM implementation occurs within three interconnected layers of civil service work:
First, there is the lifecycle of your work - the established processes, methodologies, and workflows that define how your department delivers its responsibilities. This forms the foundation upon which any LLM implementation must build.

Second, there is the domain knowledge specific to your area - whether that's policy development, analysis, research, or specialised departmental functions. This expertise determines how LLMs can best support and enhance your work rather than replace professional judgment.

Third are the public sector constraints and requirements that ensure accountability, maintain standards, and protect public trust. These shape how we implement any new technology while upholding civil service values.

Implementation considerations:
- Test LLM capabilities for specific tasks before full deployment
- Validate outputs against domain expertise
- Document limitations and workarounds
- Maintain flexibility in the implementation approach
- Regular capability assessment and adjustment

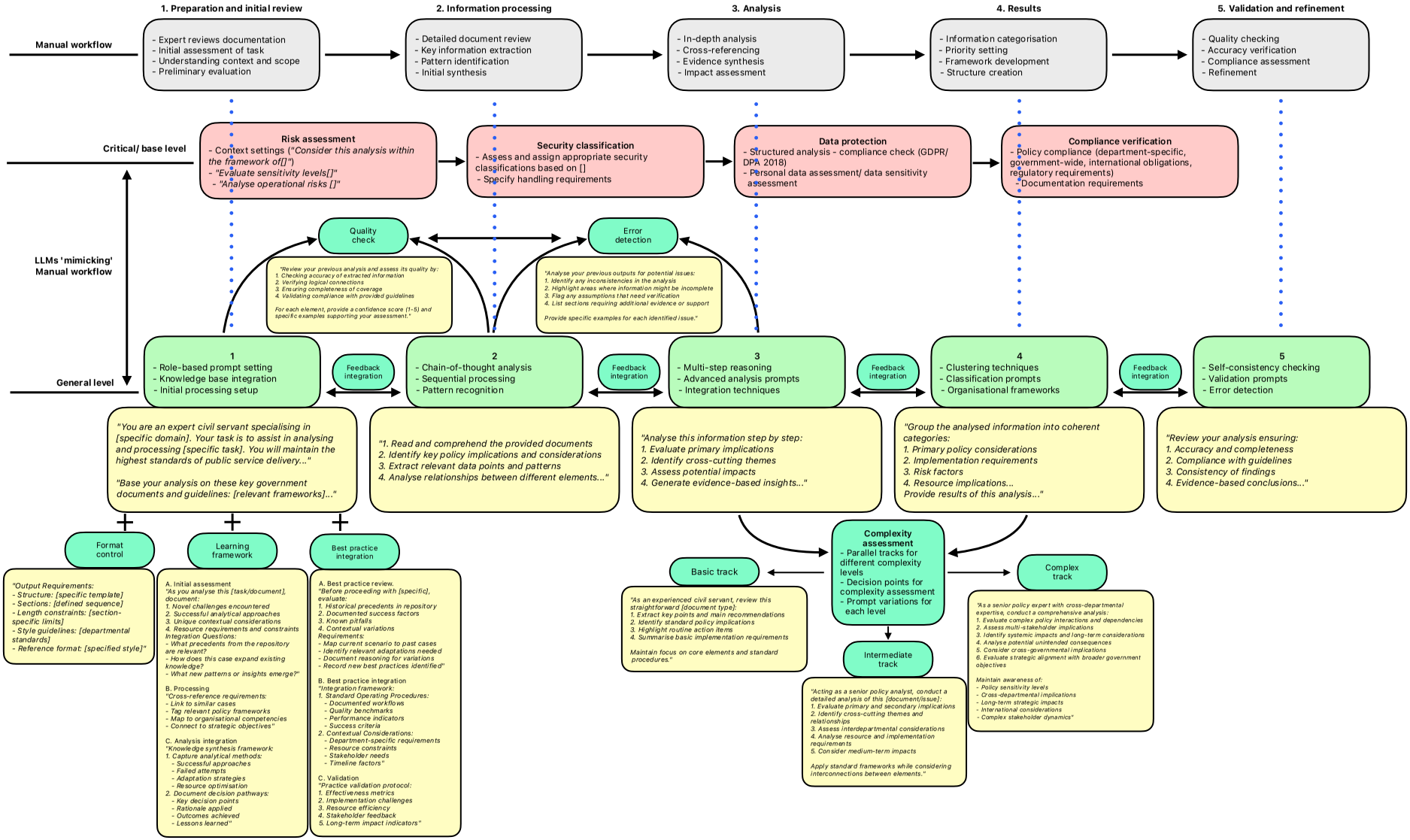
Each component includes specific checkpoints and validation steps, ensuring LLM use aligns with civil service requirements while enhancing operational efficiency.

The detailed framework that follows demonstrates how these layers interact in practice. It provides a structured approach that you can adapt based on:

- Your team's specific functional requirements
- The LLM capabilities available to you
- Your domain's particular expertise needs
- Your resource and operational constraints

Think of this framework not as a rigid prescription but as a guided exploration. Each element can be adjusted to match your context while maintaining the core principles of responsible LLM implementation. As you review the detailed process flow, consider how each element might adapt to your specific needs while preserving the essential governance and quality requirements of civil service.

Framework for LLM implementation



2.1 Operationalising the framework: prompt examples

The integration of LLMs into civil service workflows requires clear, structured approaches to common tasks. While the framework provides the overall implementation structure, specific task templates ensure consistent, governance-compliant applications across different scenarios.

Task category	Prompt structure	Key components	Governance requirements
Policy analysis & evaluation	As a policy specialist in [domain], conduct analysis following [methodology]: 1. Define scope and objectives 2. Gather and assess evidence 3. Develop and compare options 4. Form recommendations Consider: - Policy context - Implementation pathways - Resource implications - Risk factors	• Clear objectives • Evidence base • Impact assessment • Stakeholder mapping • Risk evaluation	• Policy compliance check • Classification review • Impact assessment • Data protection review
Research	Acting as a research analyst specialising in [specific area]: Conduct [analysis] following systematic review principles: 1. Quality assessment [specified criteria] 2. Data/evidence analysis 3. Gap analysis 4. Results/outcomes formulations Apply: - Quality framework [specified] - Confidence metrics - Evidence hierarchies	• Methodology • Quality criteria • Evidence assessment • Analysis framework • Recommendations	• Source verification • Methodology compliance • Quality assurance • Peer review requirements
Briefing development[9]	Prepare ministerial briefing on [topic]: Structure: 1. Executive summary (150 words) 2. Key points (prioritised) 3. Analysis 4. Options 5. Recommendations Ensure: - Clear narrative flow - Evidence-based analysis - Actionable recommendations - Risk considerations	• Clear format • Key messages • Target audience • Supporting evidence • Recommendations • Risk assessment	• Security/sensitivity classification • Information handling • Stakeholders • Accuracy verification

Task category	Prompt structure	Key components	Governance requirements
Document review	Conduct document review as [specialist role]: 1. Initial assessment 2. Content analysis 3. Compliance check 4. Recommendations Focus areas: - Policy alignment - Implementation implications - Resource requirements - Risk factors	• Review framework • Assessment criteria • Compliance checklist • Recommendations format	• Document handling • Classification check • Compliance verification • Quality assurance
Strategic assessment[10]	As a strategic advisor, assess [topic]: 1. Context analysis 2. Trend identification 3. Impact evaluation 4. Option development Consider: - Long-term implications - Cross-departmental impacts - Resource requirements - Strategic alignment	• Strategic framework • Impact metrics • Option analysis • Implementation pathway	• Strategic alignment • Cross-department review • Impact assessment • Risk evaluation

Developing effective prompts:

1. Define purpose
- Identify specific task requirements
- Clarify desired outcomes
- Consider governance constraints

2. Select structure
- Match prompt type to task
- Consider complexity level
- Ensure compliance with: departmental decision-making protocols, documentation requirements, data standards, security classifications

3. Test and refine
- Validate outputs
- Check for compliance
- Iterate based on results

9. While policy teams (and briefing managers) direct briefing content, analytical teams provide specialist LLM support for evidence synthesis and analysis. This creates a collaborative workflow where policy teams define requirements and maintain editorial control, while analytical teams deploy LLM capabilities to enhance evidence base and support analysis.
10. Strategic assessments combine policy expertise with analytical capability, where analytical teams provide LLM support to enhance evidence analysis and pattern recognition.

3. Implementing principles

LLMs serve as powerful analytical tools to support, not replace, civil service expertise. The key to successful integration lies in thoughtful alignment with existing processes and clear maintenance of human oversight. However, focusing on intended outcomes rather than just on processes, departments can leverage LLMs to not only support current operations but also identify opportunities for service improvement and innovation.

When integrating LLMs into your workflow, consider them as collaborative tools that enhance your analytical capabilities. For instance, in policy document analysis, this partnership takes shape through clear role division:

LLMs support initial processing by:
- Generating comprehensive document summaries
- Identifying recurring themes and patterns
- Highlighting potential policy implications

Meanwhile, civil servants focus on critical tasks:
- Evaluating strategic implications
- Making informed policy decisions
- Ensuring alignment with departmental objectives

3.1 Risk-based approach and usage levels

Risk assessment forms the foundation of responsible LLM implementation. Consider your task's complexity and potential impact when determining appropriate LLM use - different tasks carry different levels of risk.

Level 1: Foundation tasks - involve basic information processing with minimal risk.
- *What*: Information gathering and initial analysis
- *Examples*: Document summaries, background research
- *Controls*: Basic review and verification
- *When to use*: Regular information processing tasks

Level 2: Analytical support - encompasses pattern identification and trend analysis.
- *What*: Pattern identification and trend analysis
- *Examples*: Policy research, data analysis
- *Controls*: Regular validation and peer review
- *When to use*: Analytical and research tasks

Level 3: Policy development - requires more rigorous oversight.
- *What*: Recommendations and assessments
- *Examples*: Impact analysis, policy evaluation
- *Controls*: Rigorous review process at each stage
- *When to use*: Policy development/evaluation support

Level 4: Critical decisions - LLM use in design-making support and policy implementation should include extensive safeguards and multilayer oversight.
- *What*: Decision support and implementation
- *Examples*: Policy implementation guidance
- *Controls*: Comprehensive oversight
- *When to use*: With extensive safeguards

While LLMs offer valuable support, specific scenarios require explicit exclusion from LLM use. Critical decision-making contexts such as crisis briefings, emergency response, or situations directly affecting individual rights - demand more nuanced human expertise and established protocols. This reflects both technical limitations and civil service responsibilities. Judgment and direct accountability should be human-centric when stakes are high, time-critical, or outcomes directly impact public safety or individual rights. Departments should clearly document these boundaries, maintaining flexibility to review and adjust as capabilities evolve.

3.2 Core principles

Successful LLM implementation rests on three fundamental principles that guide daily operations.

1. Evidence drives implementation through documented capabilities and measured outcomes (success indicators, measurement frameworks, evaluation protocols). Regular evaluation ensures LLM use aligns with departmental needs while maintaining effectiveness.
2. Transparency ensures accountability through clear documentation and traceable processes (documentation requirements, communication protocols, audit mechanisms). This creates confidence in LLM-supported work while maintaining public trust.
3. Continuous learning enables improvement through shared experiences and adapted practices (feedback loops, iteration processes, knowledge sharing frameworks). This collective knowledge strengthens departmental capability while maintaining consistent standards.

3.3 Judgment and responsibility

Officials' judgment remains paramount when using LLMs. While these tools offer powerful analytical support, they complement rather than replace human expertise.

Consider LLMs as analytical assistants that inform, not determine, decisions. Your professional judgment provides crucial context interpretation and ensures alignment with civil service values.

Maintain clear responsibility through:

- Documented decision rationales
- Transparent review processes
- Regular effectiveness evaluation

4. Practical implementation

4.1 Prompt engineering and technical approaches

The effectiveness of LLM use largely depends on how we communicate with these systems. Understanding prompt engineering techniques enables civil servants to achieve consistent, reliable results.

Prompt technique	Description	Example wording
Zero-shot prompting	Direct instruction without examples, relying on the model's inherent capabilities. Best for straightforward tasks with clear parameters.	"Analyse this policy proposal and provide three key recommendations."
Few-Shot Learning	Providing specific examples of desired input-output pairs before the main task. Improves accuracy through demonstration.	"Task 1: Policy A → Impact X; Task 2: Policy B → Impact Y; Now analyse: Policy C"
Chain-of-thought	Explicit reasoning process broken into sequential steps. Enhances logical progression and reduces errors.	"Think through this step-by-step: 1) First, consider..., 2) Then analyse..., 3) Finally, conclude..."
Self-consistency	Multiple independent attempts at the same task, then synthesising results. Improves reliability through consensus.	"Approach this analysis from three different angles, then synthesise the common findings."
Role prompting	Assigning specific professional or expert perspectives to guide analysis.	"As an experienced policy researcher specialising in social welfare, analyse..."
Task decomposition	Breaking complex tasks into smaller, manageable components.	"Let's break this analysis into parts: 1. Current situation, 2. Proposed changes, 3. Impact assessment..."
Format specification	Explicit structure requirements for output organisation.	"Present your analysis as follows: SUMMARY: []; KEY FINDINGS:[]; RECOMMENDATIONS:[]"
Maieutic prompting	Socratic questioning approach to elicit deeper analysis.	"What are the key factors? Why are they important? How do they interact?"
Recursive refinement	Iterative improvement through sequential prompting.	"Review the initial analysis. What aspects need deeper examination? Let's focus on those."
Constrained generation	Setting specific parameters or limitations for the response.	"Analyse this within these constraints: budget limit, six-month timeline, existing staff resources."
Comparative framework	Structured comparison across multiple dimensions.	"Compare options across: cost, feasibility, impact, and timeline."
Knowledge integration	Explicitly requesting synthesis of multiple knowledge domains.	"Combine insights from economic analysis and social impact research to evaluate..."
Tree of thoughts	Branching decision paths with multiple possible outcomes.	"Consider three possible implementation paths: A) If we..., B) Alternatively..., C) Or we could..."

Prompt technique	Description	Example wording
Reflexive analysis	Self-evaluation and criticism of initial outputs.	"Review your analysis. What assumptions were made? What could be challenged?"
Context enhancement	Providing comprehensive background before the main task.	"Given this background: [context], and considering these factors: [factors], analyse..."
Output validation	Building in verification steps for accuracy and completeness.	"After your analysis, verify: 1) Evidence supports conclusions, 2) All stakeholders considered..."
Multi-perspective analysis	Incorporating different viewpoints systematically.	"Analyse from perspectives of: 1. Service users, 2. Providers, 3. Administrators"
Template prompting	Using standardised formats for consistent outputs.	"SITUATION:[]; CHALLENGE:[]; OPTIONS:[]; RECOMMENDATION:[]"
Evidence integration	Explicit requirements for supporting evidence.	"For each conclusion, cite specific evidence from the provided materials."

Additional quality assurance (verification) prompts approaches:

- “Explain your approach to this analysis”
- “What assumptions have you made?”
- “What are the limitations of your analysis?”
- “What alternative approaches did you consider?”
- “How confident are you in these conclusions?” or any other ‘self-assessment’ techniques

4.2 Best practices and effective use

Effective LLM implementation benefits from thoughtful consideration of processes and practices[11].

1. Structured engagement

Clear communication with LLMs enhances reliability. Begin with precise objectives and structured approaches. When analysing policy documents, for instance, specify:

- Analysis scope and purpose
- Clear objective definition
- Output format specification
- Validation criteria
- Quality benchmarks

This structured approach ensures consistent, verifiable results while maintaining departmental standards.

2. Quality control integration

Build quality assurance into your workflow rather than treating it as a final step. Implement regular checkpoints for:

- Output validation
- Accuracy verification
- Compliance confirmation
- Stakeholder review

11. An additional useful framework could be the Theory of Change approach.
URL: https://analysisfunction.civilservice.gov.uk/policy-store/the-analysis-function-theory-of-change-toolkit/

Stage	Control measures	Validation points
Input	• Data quality check • Format validation	• Source verification • Completeness check
Processing	• Output monitoring • Error detection	• Accuracy assessment • Consistency review
Final review	• Compliance check • Impact assessment	• Policy alignment • Stakeholders review

Quality assurance approaches align with established government frameworks, including the Aqua Book[12] guidance and Government Functional Standard[13]. These frameworks provide additional context for implementing robust quality controls.

3. Security and governance

Maintain robust security practices throughout LLM implementation:

- Assess information sensitivity before processing
- Apply appropriate security classifications
- Maintain clear audit trails
- Document decision rationales

Remember that every interaction with LLMs should align with civil service standards for information handling.

Understanding and verifying LLM outputs - when receiving LLM-generated content, consider:

- Request explanation of approach: "Explain your methodology for this analysis"
- Verify logical progression: "Walk through your reasoning step-by-step"
- Check information sources: "What information did you use to reach these conclusions?"
- Test consistency: request the same analysis multiple times to check for variation

The fundamental challenge stems from the architecture of public and corporate LLMs - these systems operate on infrastructure outside government control, with implications for data protection and information security. Even with robust provider safeguards, departments cannot guarantee against unintended data retention, training usage, or potential information leakage.

Critical security restrictions:

- All classified documents, regardless of level
- Information subject to the Official Secrets Act
- Sensitive personal data
- Restricted operational information and financial records
- Diplomatic communications
- Critical infrastructure details

Understanding that not all government work requires the same level of security protection, departments can implement a tiered approach to LLM usage:

12. The Aqua Book: guidance on producing quality analysis.
URL: https://www.gov.uk/government/publications/the-aqua-book-guidance-on-producing-quality-analysis-for-government
13. Government Functional Standard. GovS 002: Project delivery
URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1002673/1195-APS-CCS0521656700-001-Project-Delivery-standard_Web.pdf

Prohibition	Restricted usage	Controlled usage
Information that should never be processed through public or corporate LLMs: • Materials classified SECRET and above • Personal data requiring special protection • Critical national infrastructure information • Sensitive international/diplomatic communications	For official but non-sensitive work, requiring: • Security assessment • Government-approved platforms only • Full audit capability • Regular security evaluation	Permitted for: • Public domain information • General administrative tasks • Policy research using public sources • Non-sensitive documentation

Local models (that run directly on a user's device or local network with full access and control) offer greater security management but present their own challenges in resource requirements, maintenance complexity, update management and potential cybersecurity vulnerabilities. While local models provide maximum control over data security and processing, their implementation demands careful cost-benefit analysis, weighing enhanced security against substantial resource requirements and operational complexities – making them suitable primarily for highly sensitive operations.

To mitigate security risks while maximising the benefits of LLM implementation, departments can establish and maintain data protection safeguards:

- Use approved, secure LLM platforms
- Maintain comprehensive audit trails
- Apply data minimisation/anonymisation principles
- Enforce access controls
- Regular security assessments
- Monitor for potential data leaks

The security landscape for LLM usage continues to evolve. While local models or hybrid deployment approaches may offer enhanced security control, they present their own challenges regarding resources, maintenance, and vulnerability management. Departments should maintain active engagement with developments in:

- Government-specific LLM infrastructure
- Enhanced security protocols
- Private deployment options
- Hybrid security models

4.3 Quality assurance mechanisms

Quality assurance in LLM implementation requires systematic attention to validation and documentation. Effective validation ensures reliable outputs while maintaining civil service standards:

1. Initial assessment review outputs for:
- Accuracy and completeness
- Policy alignment
- Practical applicability
- Evidence base

2. Expert validation - engage appropriate expertise for:
- Technical accuracy
- Policy implications
- Implementation feasibility
- Risk assessment

Documentation standards. Maintain comprehensive records of:
- Standard templates
- Review protocols
- Decision processes
- Validation steps
- Quality checks
- Outcome evaluations

This documentation ensures transparency while supporting continuous improvement.

Continuous improvement. Learn from experience to enhance implementation:
- Regular assessments
- Feedback collection
- Document successful approaches
- Share best practices
- Address identified challenges
- Update procedures accordingly

Output identification and transparency - ensure LLM-generated content is clearly identified:
- Include generation metadata (date, model, version)
- Add clear attribution statements
- Implement consistent marking system (example format: [LLM-Generated Content], Generated: [Date/Time], Model: [LLM Type], Purpose: [Task Description], Review Status: [Verification Level])

5. Risk management and control

Risk management in LLM implementation requires a balanced approach that protects public sector integrity while enabling innovation. This section provides practical tools and frameworks for:

- Identifying and assessing risks
- Implementing appropriate controls
- Monitoring effectiveness
- Ensuring continuous improvement

5.1 Security and sensitivity

Civil servants should approach LLM implementation with a clear understanding of their department's security requirements. This involves balancing innovation with the protection of sensitive information and public trust.

The Government Security Classifications Policy[14] and data protection requirements establish clear boundaries for LLM implementation. All civil service documents carry specific classification levels that determine their handling requirements. For LLM use, only documents classified as 'Official' may be considered for processing, subject to departmental security controls and monitoring protocols. Documents classified as secret or top-secret are explicitly excluded from LLM processing to maintain security integrity.

Regarding personal data, current Civil Service guidance and data protection regulations (GDPR) prohibit any processing of personal information through LLMs. This applies comprehensively to all personal data, whether of staff or the public, regardless of perceived sensitivity level. No exceptions exist for minimal or apparently non-sensitive personal information.

These requirements ensure both security compliance and data protection while supporting appropriate LLM implementation. Teams must verify document classification and screen for personal data before any LLM processing, maintaining clear audit trails throughout.

Security considerations fall into three key areas:

1. Data protection

Protecting sensitive information requires careful attention to both input and output processes. When using LLMs, consider:
"*How sensitive is the information being processed? What controls are needed to protect it? What would the implications be if this information were compromised?*"

2. Process security

Secure processes ensure consistent protection throughout the LLM implementation. Each step in the LLM workflow needs appropriate controls. Think about *who* needs access, *when* they need it, and *how* to maintain security without impeding necessary work.

3. Output management

Managing LLM outputs requires careful consideration of both immediate and long-term implications:
"*Consider how outputs will be used, stored, and shared. What controls are needed to protect sensitive information while ensuring necessary access?*"

14. Government Security Classifications Policy.
URL: https://www.gov.uk/government/publications/government-security-classifications/government-security-classifications-policy-html

5.2. Monitoring and risk management

Risk management in LLM implementation is not about eliminating all risks but managing them effectively. Consider three key principles:

1. Proportionality - match your controls to your risks. Not every use of LLMs requires the highest level of security, but every use needs appropriate protection.
2. Practicality - controls should enable work, not prevent it. Look for solutions that protect while supporting efficient operations.
3. Adaptability - as LLM technology evolves, so must our approach to risk management. Build flexibility into your controls.

Effective risk management integrates seamlessly into daily operations.

Regular review cycle - build regular reviews into your workflow. This isn't about creating extra work but ensuring security remains effective.

- *What is working well?*
- *What needs adjustment?*
- *What new risks have emerged?*

Stage	**Key activities**	**Tools/Methods**
Identification	• Risk mapping • Impact evaluation	• Assessment templates • Expert consultation
Analysis	• Risk quantification • Control assessment	• Scoring matrices • Gap analysis
Response	• Control design • Implementation	• Mitigation strategies • Action plans
Monitoring	• Performance tracking • Effectiveness review	• Metrics dashboard • Regular reporting

Practical monitoring - rather than complex monitoring systems, focus on key indicators that matter for your department: monitor what matters – not everything that moves. Focus on indicators that tell you about real risks to your operations.

LLM-specific risk monitoring	
Output reliability • Monitor for hallucination instances • Track consistency across similar queries • Document edge cases and limitations Establish output validation protocols	Technical boundaries • Monitor context window usage limits • Track token consumption patterns • Assess prompt engineering effectiveness Document model version impacts
Response patterns • Monitor for unexpected output variations • Track response quality trends	Performance Metrics • Track success rates for different task types • Monitor processing time variations

LLM-specific risk monitoring	
• Assess prompt injection vulnerabilities Document behavioral changes across updates	• Assess resource utilisation patterns Document effectiveness across departments

Remember that risk management is an ongoing journey, not a destination. Start small, learn from experience, and gradually build confidence in managing LLM-related risks. The key is to focus on the department's specific needs and context.

6. Implementation roadmap

Implementing LLMs in government requires careful consideration of both opportunities and responsibilities. This section provides practical guidance for civil servants looking to integrate these tools effectively into their work:

- Pilot programme development
- Policy framework integration
- Risk mitigation strategies
- Progressive scaling

6.1 Initial steps and piloting

Start by identifying a specific challenge (e.g. analyse large volumes of consultation responses). Focusing on specific needs makes it easy to measure success and learn valuable lessons for broader implementation.

- Current challenges that LLMs might help address
- Available resources and capabilities
- Potential risks and mitigation strategies

The practical approach will consist of mapping current processes, identifying where LLMs could add value and determining what is an efficient way to implement it.

Readiness assessment

- Technical infrastructure evaluation
- Staff capability assessment
- Resource availability review
- Stakeholder mapping

Pilot project success depends on the following:

- Meaningful but contained task selection
- A low-risk but valuable outcome
- Clear in their objectives
- Easy to measure

Pilot programme development

Project type	Scope	Success metrics
Document analysis	• Single department • Limited dataset	• Accuracy rates • Time savings
Process automation	• Specific workflow • Controlled environment	• Error reduction • Efficiency gains
Research support	• Defined topic area • Clear boundaries	• Quality improvement • Resource optimisation

6.2 Ongoing management

Regularly reviewing points could help to understand what is working and what needs to be adjusted. The learning point from piloting is that small, frequent adjustments are better than major changes.

Effective ongoing management means focusing on practical value.

"*Are these tools making our work better? Are they helping us serve the public more effectively*?"
Such questions can guide decisions about continued use and expansion.

Successful LLM integration into civil service operations requires a structured yet adaptable approach across three interconnected areas:

- Process alignment - begin with a systematic review of your current workflows to identify where LLM integration can add most value. Map existing processes in detail, paying particular attention to information flow patterns, decision points, quality control stages. Based on this, update standard operating procedures to reflect new LLM-enhanced workflows.
- Capability building - successful integration depends on building both individual and organisational capabilities. Develop a comprehensive training program that addresses technical skills for LLM interaction, output evaluation competencies, risk awareness and management, best practice application. Document and share best practices as they emerge, including successful prompt strategies, effective quality control methods, efficient workflow integrations, problem-solving approaches. Thus, such knowledge sharing mechanisms should capture learning from early adopters, facilitate peer-to-peer learning, enable cross-department collaboration and support continuous improvement.
- Performance monitoring - implement regular assessment cycles to track integration effectiveness. develop feedback collection systems that capture user experiences and challenges, process efficiency gains, quality improvement opportunities, resource utilisation patterns. Measure impact across multiple dimensions, e.g. operational/functional efficiency, output quality, user capability development, resource optimisation.

6.3 Integration strategy

Integration principles:

1. Workflow enhancement
- Map existing processes
- Identify integration points
- Define enhancement opportunities
- Maintain operational continuity

2. Implementation checkpoints

Stage	Key activities	Validation points
Initial setup	• Process mapping • Integration planning	• Workflow alignment • Stakeholders approval
Deployment	• Controlled rollout • User training	• Performance monitoring • Feedback collection
Optimisation	• Process refinement • Capability building	• Efficiency assessment • Impact evaluation

3. Success metrics framework:
- Performance indicators
- Quality measures
- Efficiency gains
- User satisfaction

6.4 Scaling framework

Scaling up LLM use should follow demonstrated success: expanding and scaling LLM usage depends on clear evidence of value and strong controls in place. Developing mechanisms and unified approaches through the expansion of areas of application helps build support across departments/units/teams.

Think about scaling in terms of build capability before expanding use. Understanding both the potential and limitations of these tools is crucial.

Scaling principles:
1. Value first - every expansion should ensure clear benefits to public service delivery/team workflow.
2. Learn continuously - create opportunities for teams to share experiences and insights based on tests and experiments with LLMs.
3. Stay focused - keep sight of your core mission and expected outcomes.

Implementation success and long-term value lie in focusing on practical value rather than technology.

Remember that implementation isn't about perfect execution - it's about continuous improvement in service delivery. Start where you are, use what you have, and build on what you learn.

From the beginning:
1. Understand your context - "*What specific challenges could LLMs help address in your work?*"
2. Start small - "*Which project could provide valuable learning?*"
3. Build capability - "*How can you help your team/unit develop the necessary skills?*"

Successful scaling of LLM implementation across departments requires formal governance structures rather than relying solely on organic growth. A centralised approach in operating through a dedicated centre (something like a Centre of Excellence/ Expertise) ensures consistent standards, comprehensive risk management, flexibility and adaptability to changes and, at the same time, transparently accumulates and disseminates best practices tailored to specific departmental needs. Such a structural approach can centralise expertise and implementation guidance, systematise knowledge transfer, use a risk assessment approach and monitoring, and centralise the best-practise library and AI/LLMs risk register, among others.

By establishing these governance mechanisms before widespread adoption, departments can prevent fragmented implementation while providing local initiative and experiments to flourish, manage aggregate risks effectively, and ensure LLM usage aligns with strategic objectives while maintaining appropriate oversight.

7. Resources

7.1 LLMs

Large language models and generative AI: House of Lords Communications and Digital Committee report (15 November 2024) https://lordslibrary.parliament.uk/large-language-models-and-generative-ai-house-of-lords-communications-and-digital-committee-report/

Generative AI framework for HM Government. Created by the Central Digital and Data Office (18 January 2024) https://www.gov.uk/government/publications/generative-ai-framework-for-hmg

A guide to using artificial intelligence in the public sector. https://www.gov.uk/government/publications/understanding-artificial-intelligence/a-guide-to-using-artificial-intelligence-in-the-public-sector

Planning and preparing for artificial intelligence implementation. https://www.gov.uk/guidance/planning-and-preparing-for-artificial-intelligence-implementation

Assessing if artificial intelligence is the right solution. https://www.gov.uk/guidance/assessing-if-artificial-intelligence-is-the-right-solution

Understanding artificial intelligence ethics and safety. https://www.gov.uk/guidance/understanding-artificial-intelligence-ethics-and-safety

Managing your artificial intelligence project. https://www.gov.uk/guidance/managing-your-artificial-intelligence-project

National Audit Office report-Use of artificial intelligence in government. https://www.nao.org.uk/reports/use-of-artificial-intelligence-in-government/

Digital transformation in government: addressing the barriers to efficiency. House of Commons Committee report. https://publications.parliament.uk/pa/cm5803/cmselect/cmpubacc/1229/report.html

The governance of artificial intelligence: interim report. House of Commons. https://committees.parliament.uk/publications/41130/documents/205611/default/

Communications and Digital Committee Large language models and generative AI. 1st Report (February 2024). https://publications.parliament.uk/pa/ld5804/ldselect/ldcomm/54/5402.htm

Guidance on the Impact Evaluation of AI Interventions https://www.gov.uk/government/publications/the-magenta-book/guidance-on-the-impact-evaluation-of-ai-interventions-html

Hackenburg, K., & Margetts, H. (2024). Evaluating the persuasive influence of political microtargeting with large language models. Proceedings of the National Academy of Sciences, 121(24), https://www.pnas.org/doi/10.1073/pnas.2403116121

Myers, D., Mohawesh, R., Chellaboina, V. I., Sathvik, A. L., Venkatesh, P., Ho, Y. H., ... & Jararweh, Y. (2024). Foundation and large language models: fundamentals, challenges, opportunities, and social impacts. Cluster Computing, 27(1), 1-26. https://link.springer.com/article/10.1007/s10586-023-04203-7

Naveed, H., Khan, A. U., Qiu, S., Saqib, M., Anwar, S., Usman, M., ... & Mian, A. (2023). A comprehensive overview of large language models. arXiv preprint, https://arxiv.org/pdf/2307.06435

Van Noordt, C., & Misuraca, G. (2022). Artificial intelligence for the public sector: results of landscaping the use of AI in government across the European Union. Government information quarterly, 39(3), 101714. https://www.sciencedirect.com/science/article/pii/S0740624X22000478

Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2019). Artificial intelligence and the public sector—applications and challenges. International Journal of Public Administration, 42(7), 596-615. https://www.tandfonline.com/doi/full/10.1080/01900692.2018.1498103

Heseltine, M., & Clemm von Hohenberg, B. (2024). Large language models as a substitute for human experts in annotating political text. Research & Politics, 11(1), https://journals.sagepub.com/doi/full/10.1177/20531680241236239

Mellon, J., Bailey, J., Scott, R., Breckwoldt, J., Miori, M., & Schmedeman, P. (2024). Do AIs know what the most important issue is? Using language models to code open-text social survey responses at scale. Research & Politics, 11(1), https://journals.sagepub.com/doi/10.1177/20531680241231468

7.2 Prompt engineering

Prompt engineering. Guide. OpenAI, https://platform.openai.com/docs/guides/prompt-engineering

Prompt engineering overview. Guide. Anthropic, https://docs.anthropic.com/en/docs/build-with-claude/prompt-engineering/overview

Schulhoff, S., Ilie, M., Balepur, N., Kahadze, K., Liu, A., Si, C., ... & Resnik, P. (2024). The Prompt Report: A Systematic Survey of Prompting Techniques. *arXiv preprint,* https://arxiv.org/pdf/2406.06608v1

Gao, A. (2023). Prompt engineering for large language models. *Available at SSRN 4504303*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4504303

Li, Y. (2023). A practical survey on zero-shot prompt design for in-context learning. *arXiv preprint*, https://arxiv.org/pdf/2309.13205

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., ... & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, *35*, 24824-24837. https://arxiv.org/abs/2201.11903

Zhao, Z., Wallace, E., Feng, S., Klein, D., & Singh, S. (2021, July). Calibrate before use: Improving few-shot performance of language models. In *International conference on machine learning* (pp. 12697-12706). PMLR. https://arxiv.org/abs/2102.09690

7.3 Additional resources

Introduction to AI assurance. Department for Science, Innovation & Technology (February 2024) https://assets.publishing.service.gov.uk/media/65ccf508c96cf3000c6a37a1/Introduction_to_AI_Assurance.pdf

AI Management Essentials. Department for Science, Innovation & Technology. https://assets.publishing.service.gov.uk/media/672a5706094e4e60c466d19f/AI_Management_Essentials_tool_Self-Assessment.pdf

Zhang, Y., Cai, Y., Zuo, X., Luan, X., Wang, K., Hou, Z., ... & Dong, J. S. (2024). The Fusion of Large Language Models and Formal Methods for Trustworthy AI Agents: A Roadmap. *arXiv preprint,* https://arxiv.org/pdf/2412.06512v1 https://arxiv.org/pdf/2412.06512v1

AI, N. (2024). Artificial intelligence risk management framework: Generative artificial intelligence profile. https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM computing surveys (CSUR)*, *54*(6), 1-35. https://arxiv.org/pdf/1908.09635

Bisbee, J., Clinton, J. D., Dorff, C., Kenkel, B., & Larson, J. M. (2023). Synthetic replacements for human survey data? The perils of large language models. *Political Analysis*, 1-16. https://www.cambridge.org/core/journals/political-analysis/article/synthetic-replacements-for-human-survey-data-the-perils-of-large-language-models/B92267DC26195C7F36E63EA04A47D2FE

Zhang, H., Wu, C., Xie, J., Lyu, Y., Cai, J., & Carroll, J. M. (2023). Redefining qualitative analysis in the AI era: Utilizing ChatGPT for efficient thematic analysis. *arXiv preprint*, https://arxiv.org/abs/2309.10771

Borger, J. G., Ng, A. P., Anderton, H., Ashdown, G. W., Auld, M., Blewitt, M. E., ... & Naik, S. H. (2023). Artificial intelligence takes center stage: exploring the capabilities and implications of ChatGPT and other AI-assisted technologies in scientific research and education. *Immunology and cell biology*, *101*(10), 923-935. https://pubmed.ncbi.nlm.nih.gov/37721869/

Christou, P. A. (2023). How to use artificial intelligence (AI) as a resource, methodological and analysis tool in qualitative research? *Qualitative Report*, *28*(7). https://nsuworks.nova.edu/tqr/vol28/iss7/9/

Fiannaca, A. J., Kulkarni, C., Cai, C. J., & Terry, M. (2023, April). Programming without a programming language: Challenges and opportunities for designing developer tools for prompt programming. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (pp. 1-7). https://research.google/pubs/programming-with-a-programming-language-challenges-and-opportunities-for-designing-developer-tools-for-prompt-programming/

Gao, J., Choo, K. T. W., Cao, J., Lee, R. K. W., & Perrault, S. (2023). CoAIcoder: Examining the effectiveness of AI-assisted human-to-human collaboration in qualitative analysis. *ACM Transactions on Computer-Human Interaction*, *31*(1), 1-38. https://arxiv.org/abs/2304.05560

Nguyen-Trung, K. (2024). ChatGPT in Thematic Analysis: Can AI become a research assistant in qualitative research? *OSF Preprint*. https://osf.io/preprints/osf/vefwc

Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems*, *3*, 121-154. https://www.sciencedirect.com/science/article/pii/S266734522300024X

Turobov, A., Coyle, D., & Harding, V. (2024). Using ChatGPT for thematic analysis. *arXiv preprint*, https://arxiv.org/abs/2405.08828

Zhang, H., Wu, C., Xie, J., Lyu, Y., Cai, J., & Carroll, J. M. (2023). Redefining qualitative analysis in the AI era: Utilizing ChatGPT for efficient thematic analysis. *arXiv preprint*, https://arxiv.org/abs/2309.10771

Kulal, A., Rahiman, H. U., Suvarna, H., Abhishek, N., & Dinesh, S. (2024). Enhancing public service delivery efficiency: Exploring the impact of AI. Journal of Open Innovation: Technology, Market, and Complexity, 10(3). https://www.sciencedirect.com/science/article/pii/S2199853124001239

Agrawal, A., Gans, J., & Goldfarb, A. (2022). Power and prediction: The disruptive economics of artificial intelligence. Harvard Business Press. https://store.hbr.org/product/power-and-prediction-the-disruptive-economics-of-artificial-intelligence/10580?srsltid=AfmBOorMCSHRT7sana5ONVSql4klBWhFSXuTYPhCO-ZIYer8FeZn2sDC

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature machine intelligence, 1(5), 206-215. https://www.nature.com/articles/s42256-019-0048-x

Leslie, D. (2019). Understanding artificial intelligence ethics and safety. arXiv preprint. https://www.turing.ac.uk/sites/default/files/2019-08/understanding_artificial_intelligence_ethics_and_safety.pdf

UNIVERSITY OF CAMBRIDGE

Bennett Institute for Public Policy Cambridge

bennettinstitute.cam.ac.uk