

# CMS

# Artificial Intelligence

# Playbook

Prepared by the CMS AI Explorers

Version 4



This document contains links to non-United States Government websites and mentions of specific businesses, devices, or materials to clearly explain a concept or use case. We are providing these links and context because they contain additional information relevant to the topic(s) discussed in this document or that otherwise may be useful to the reader. We cannot attest to the accuracy of information provided on the cited third-party websites or any other linked third-party site. We are providing these links for reference only; linking to a non-United States Government website does not constitute an endorsement by CMS, HHS, or any of their employees of the sponsors or the information and/or any products presented on the website. Also, please be aware that the privacy protections generally provided by United States Government websites do not apply to third-party sites.

The CMS Artificial Intelligence Playbook (AI Playbook) discusses a dynamically evolving subject, and continuous reviews and updates are planned. Feedback is welcome via email to [ai@cms.hhs.gov](mailto:ai@cms.hhs.gov). Additional opportunities for CMS employees to contribute to this document will be posted in the CMS Slack channel [#ai-community](#).

You may access the latest version of this publication at: <https://ai.cms.gov/>

## Acknowledgments

This AI Playbook Version 4 was developed through extensive collaboration and thoughtful input from contributors across CMS. We are grateful for the valuable contributions from representatives across many CMS components and initiatives. The playbook benefited from several rounds of rigorous review, with each phase bringing unique perspectives and expertise. Our colleagues provided critical feedback that shaped the playbook's content, structure, and accessibility. Their insights helped ensure this resource effectively serves all intended audiences while maintaining technical accuracy and practical applicability. Special recognition goes to the Office of Information Technology (OIT) AI Explorers team for their dedication to writing, managing stakeholder engagement, and shepherding this document through the entire development process. The contributors below (by role and alphabetical order of Last Name) were the key to developing such a comprehensive endeavor.

**Executive Sponsors:**

A. Colon

**Principal Authors:**

M. Artz; C. Rutherford

**Subject Matter Experts & Contributing Authors:**

C. Lam; S. Rego; W. Rubin; A. Said; M. Whittington; I. Vailikit

**Managing Editors:**

M. Artz; P. Lohani; I. Vailikit

**Design & Production Team:**

C. Bechara

**Reviewers & Advisory:**

E. Adjakwah; A. Arnold; S. Bluher; D. Bobrosky; W. Gordon; R. Hurlbut; A. Mason-Elbert; D. Quinn; J. Slade; C. Stoltz; T. Thompson; Z. Withers; E. Wood

## Version Information

Updates are organized through a versioning system where major changes will be indicated by the first number (e.g., from 1.0 to 2.0), while minor alterations will be indicated by adding a decimal (e.g., from 1.0 to 1.1). All updates are recorded in the Version Control Table below, showing the version, change date, and change details.

Version	Change Date	Editor Control	Details
1.0	2021-09-15	X. Wu	Initial Publication via AI Pilot
2.0	2022-10-18	X. Wu, T. Ahmad	Reorganization, Use Case, Expansion into Responsible AI (RAI)
3.0	2024-05-15	X. Wu, C. Rutherford	Complete rewrite focused on CMS application of AI/ML and Human-Centered AI Development
4.0	2025-09-12	M. Artz, C. Rutherford	Reorganization and improvements to foster direct application within CMS. Updates to ensure alignment to organizational strategic goals.

## Cite As

Centers for Medicare & Medicaid Services. *CMS AI Playbook (Version 4)*. 2025.  
<https://ai.cms.gov/playbook>.

## Content

Version Information .....	ii
Cite As .....	ii
Executive Summary.....	1
1. Introduction .....	2
1.1. Context.....	2
1.2. Contents .....	2
1.3. Intended Audience .....	3
2. AI Primer .....	4
2.1. Technical View of AI .....	4
2.1.1. Elements of an AI Solution .....	4
2.1.2. AI Model Mechanics.....	5
2.2. Organizational View of AI.....	6
2.2.1. Types of AI Efforts.....	7
2.2.2. The Organizational Ecosystem of AI .....	9
Key Takeaways - AI Primer .....	10
3. AI at CMS.....	11
3.1. Current AI Portfolio .....	11
3.1.1. AI Use Cases by AI Capability .....	11
3.1.2. AI Use Cases by Stage of Development.....	12
3.2. AI Maturity at CMS.....	13
3.2.1. AI Organizational Maturity Model.....	13
3.2.2. CMS Maturity Assessment .....	15
3.3. Guiding Principles for AI at CMS .....	17
3.3.1. Organizational AI Enablement.....	19
3.3.2. AI Innovation .....	20
3.3.3. Human-Centered AI .....	21
3.3.4. AI Performance Drivers .....	23
Key Takeaways - AI at CMS .....	24
4. Governance .....	26
4.1. Governance Roles and Structure .....	26
4.2. The Governance Process .....	29
4.2.1. The Two-Axis Approach: Balancing Opportunity and Risk .....	30
4.2.2. Graduated Evaluation.....	32
4.2.3. Scheduled Reviews by Risk/Opportunity Category.....	33
4.3. Registries and Dashboards .....	34
4.3.1. Registries .....	34
4.3.2. Dashboards and Reports .....	34
Key Takeaways - Governance .....	35
5. Conducting an AI Project.....	37
5.1. Starting a Project.....	37
5.1.1. Stages .....	37

5.1.2. AI Decision Framework .....	38
5.1.3. Team Skillsets .....	39
5.1.4. Stakeholder Collaboration.....	40
5.1.5. Case Study Example - Starting a Project.....	40
5.2. Research and Approach .....	41
5.2.1. Identify the Business Problem .....	41
5.2.2. Establishing Requirements.....	43
5.2.3. Implementing the AI Decision Framework.....	44
5.2.4. Designing AI with the Human in Mind .....	47
5.2.5. Case Study Example - Research and Approach .....	50
5.3. Design and Development.....	50
5.3.1. Designing Human-AI Interactions .....	50
5.3.2. Planning for Versioning .....	51
5.3.3. Preparing Data .....	52
5.3.4. Selecting the Right Models .....	54
5.3.5. Developing and Testing Models .....	56
5.3.6. Case Study Example - Design and Development.....	57
5.4. Deployment and Integration .....	58
5.4.1. Deploying an AI Product.....	58
5.4.2. Acceptance and Adoption.....	60
5.4.3. Scaling .....	61
5.4.4. Case Study Example - Deployment and Integration.....	62
Key Takeaways - Conducting an AI Project.....	63
6. Looking Ahead.....	65
6.1. A Glimpse into Future Technologies .....	65
6.1.1. Near Future Technology .....	66
6.1.2. Distant Future Technology .....	67
6.2. Organizational Preparation .....	68
6.2.1. Evolving Policies and Governance.....	68
6.2.2. Continuing Workforce Transformation.....	69
6.2.3. Continuing Interdisciplinary Collaboration .....	69
6.2.4. Managing Organizational Change for AI Adoption.....	70
Key Takeaways - Looking Ahead.....	74
References.....	75
Appendices.....	A-1
Appendix A. External Resources .....	A-1
Appendix B. Internal Resources.....	B-1
Appendix C. Acronyms.....	C-1

## List of Figures

Figure 1. Relationships Between AI Solution Elements.....	4
Figure 2. Types of AI Efforts.....	7
Figure 3. AI Organizational Alignment and Interdisciplinary Collaboration .....	9
Figure 4. AI Use Cases by AI Capability .....	12
Figure 5. AI Use Cases by Stage of Development.....	12
Figure 6. CMS AI Organizational Maturity Model .....	13
Figure 7. Guiding Principles for AI.....	18
Figure 8. Governance Approach.....	30
Figure 9. Two-Axis Impact Assessment Approach.....	31
Figure 10. Graduated Documentation Concept .....	32
Figure 11. Stages of an AI Project .....	37
Figure 12. AI Decision Framework Overview .....	38
Figure 13. Starting a Project for CMS Chat.....	41
Figure 14. Feasibility, Desirability, Viability .....	42
Figure 15. AI Decision Framework Process Flow.....	44
Figure 16. Human-Centered AI Matrix Guide .....	48
Figure 17. Research and Approach for CMS Chat .....	50
Figure 18. Practical Threat Modeling for AI .....	54
Figure 19. Evaluation Labels for Llama 3.1 8b and Claude 3.5 Sonnet .....	55
Figure 20. Design and Development for CMS Chat.....	57
Figure 21. Deployment and Integration for CMS Chat.....	63
Figure 22. Technology Timeline .....	65

## List of Tables

Table 1. Playbook Audience Groups and Relevant Chapters.....	3
Table 2. AI Elements and Definitions .....	5
Table 3. Comparison of ML AI and Symbolic AI.....	6
Table 4. Types of AI Efforts .....	8
Table 5. AI Organizational Maturity Model Applications .....	14
Table 6. CMS AI Maturity Needs .....	15
Table 7. CMS AI Maturity Efforts .....	16
Table 8. Organizational AI Enablement Domains .....	19
Table 9. AI Innovation Domains .....	21
Table 10. Human-Centered AI Domains .....	22
Table 11. AI Performance Drivers Domains.....	23
Table 12. Governance Roles and Structure .....	27
Table 13. Example Two-Axis (Risk x Opportunity) Evaluation Factors.....	31
Table 14. Example Information Collection for AI Governance .....	32
Table 15. Example Reevaluation Cadence.....	33
Table 16. Team Roles and Responsibilities .....	39
Table 17. Roles of Agency Groups .....	40
Table 18. Requirement Types.....	43
Table 19. Indicators For and Against AI Suitability .....	45
Table 20. Indicators for Buying vs. Building AI .....	45
Table 21. Indicators for Enhancing an Existing System vs. Transitioning to a New System.....	46
Table 22. Algorithmic Risk and Impact Assessment Framework.....	49
Table 23. Advantages of Versioning for AI Projects.....	51
Table 24. Common Tools for Version Control Approaches.....	52
Table 25. Data Preparation Tasks for AI Projects.....	52
Table 26. Model Selection Decision Guide.....	54
Table 27. Key Considerations for LLM Implementation .....	56
Table 28. Model Development and Testing.....	56
Table 29. Deploying Your AI System .....	58
Table 30. AI Technology Acceptance .....	60
Table 31. AI Technology Adoption .....	61
Table 32. Scaling an AI Product .....	61
Table 33. Integrated Assistants .....	66
Table 34. Agentic AI.....	66
Table 35. Integrated Multimodal AI .....	67
Table 36. Digital Twins.....	67
Table 37. Distant Future Technologies .....	67
Table 38. Change Management Frameworks and Concepts.....	73

## CMS Artificial Intelligence Playbook

---

Table 39. External Resources .....	A-1
Table 40. General AI Tools and Trainings.....	B-1
Table 41. AI Development Resources.....	B-1
Table 42. CMS Tools and White Papers from the AI Explorers Program .....	B-2

# Executive Summary

The Centers for Medicare & Medicaid (CMS) Artificial Intelligence (AI) Playbook documents the agency's current AI maturity and adoption efforts while providing resources for teams to accelerate future progress. Over the past few years, CMS has progressed from the AI Organizational Maturity Model's *Stage 1: Exploratory* into the later stages of *Stage 2: Foundation Building*. Each updated version of this Playbook reflects CMS' evolving relationship with AI technologies and practices.

With general information about AI now widely accessible online, Version 4 concentrates on providing CMS-specific context, guidance, and tools for leadership, project teams, and information technology (IT) and security roles supporting AI initiatives within the agency. The content of this Playbook is guided by the April 2025 Office of Management and Budget (OMB) memos ([M-25-21](#) and [M-25-22](#)) and informed by research and feedback gathered from CMS staff across multiple components and disciplines. It reflects ongoing work supporting CMS' AI maturity, including efforts related to the AI Cross-Cutting Initiative (AI CCI), infrastructure development, CMS Chat, and increased training through Workforce Resilience (WR) and other initiatives. These efforts are actively in progress.

The Playbook emphasizes that advancing AI maturity is a shared responsibility across roles and teams. It serves as a practical reference and call to action for CMS staff to participate in shaping how the agency adopts and uses AI to serve the public.

The Playbook has six chapters as outlined in the table below:

Chapter	Description
<b>1. Introduction</b>	Outlines the purpose of the Playbook. It describes the Playbook's contents and intended audiences, including leadership and managers, AI project teams, IT and security teams, and CMS staff.
<b>2. AI Primer</b>	Defines key technical terms and organizational concepts to help CMS leaders and AI project team members establish a shared understanding before pursuing an AI initiative, project, product, pilot, or proof of concept.
<b>3. AI at CMS</b>	Summarizes the status of current AI use cases across the agency from the Health and Human Services (HHS) FY 2024 Use Case Inventory. This chapter also introduces the CMS AI Maturity Model and outlines Four Guiding Principles for AI at CMS: Organizational AI Enablement, AI Innovation, Human-Centered AI, and AI Performance Drivers.
<b>4. Governance</b>	Outlines recommendations for current frameworks, policies, and approval processes guiding AI development and use at CMS. Includes descriptions of necessary roles and approaches to review AI projects that present varying levels of risk and opportunity.
<b>5. Conducting an AI Project</b>	Provides a step-by-step overview of the AI project lifecycle while emphasizing iterative design and development via proofs of concept and pilots. Topics include determining whether AI is the right solution, deciding whether to buy or build AI, identifying the business problem, designing with the human in mind, iterative development, deployment, and integration.
<b>6. Looking Ahead</b>	Describes emerging technologies and trends and provides recommendations for how CMS can prepare for the future by aligning administrative structures, policies, and resources to effectively integrate AI's evolving capabilities within the agency.

Feedback is welcome via email to [ai@cms.hhs.gov](mailto:ai@cms.hhs.gov).

## 1. Introduction

The Centers for Medicare & Medicaid Services (CMS) Artificial Intelligence (AI) Playbook, Version 4, advances AI maturity and innovation by providing leaders, business managers, AI project teams and information technology (IT) and security teams with resources for AI development, deployment, governance, and risk mitigation.

### 1.1. Context

AI offers opportunities to reduce administrative burdens, enhance operational efficiency and optimize healthcare spending while improving service delivery. As the steward of health coverage for over 160 million Americans, CMS is strategically integrating AI to streamline processes and accelerate healthcare innovation. The agency aims to ensure AI adoption supports CMS' mission and long-term goals.

Aligned with Office of Management and Budget (OMB) memos [M-25-21](#) and [M-25-22](#), the [2020 National Artificial Intelligence Initiative Act](#), and the [AI Action Plan](#), this AI Playbook Version 4 was created by the OIT AI Explorers program. AI Explorers builds cross-agency AI communities, fuels research, development, and pilot projects, and has been honored with both a 2025 CMS Honor Award and a FORUM 2025 Disruptive Technology Award. CMS is already using AI to combat fraud, waste, and abuse, detect anomalies in provider data, and enhance employee efficiency, such as through the CMS Chat. By streamlining analytical and administrative tasks, AI will enable CMS teams to focus on higher-value work, improving efficiency and service delivery for the American public.

### 1.2. Contents

The CMS AI Playbook is regularly updated to reflect the current AI needs at CMS. Previous versions of the CMS AI Playbook were written for a workforce newly introduced to AI and provided overviews of AI technology and general guidance on AI-based software development.

With general information about AI now widely accessible online and in CMS' AI Community Slack channel (link to channel can be found in Appendix B), Version 4 concentrates on providing CMS-specific context, guidance, and tools for teams pursuing AI initiatives within the agency. It is intended for CMS staff who are planning, building, overseeing, or supporting AI projects. It is also a resource for leaders responsible for making decisions about AI investments and the organizational processes needed to support them.

Advancing CMS AI maturity and accurately capturing these efforts is an ongoing process and is driven by collaboration across the agency. The AI Explorers team conducted interviews, discussion forums, and feedback sessions with CMS staff members to capture these efforts and produce a Playbook that would best meet the agency's needs at this time.

The result is a version that aims to specifically support CMS leadership and technical teams by providing CMS-specific guidance on AI organizational maturity (Chapter 3), governance (Chapter 4), practical tools for AI adoption, collaboration, and product development (Chapter 5), and organizational change management and next steps (Chapter 6). Each chapter ends with a Key Takeaways section that highlights specific actions and considerations relevant to the reader's role.

### 1.3. Intended Audience

This Playbook prioritizes the needs of leadership and managers, AI project teams, and IT and security teams in pursuing AI maturity efforts and developing AI products. It provides tailored guidance aligned with each group's AI-related responsibilities as they plan and execute AI initiatives and projects.

Table 1 below and the corresponding icons help readers identify their audience group, understand their AI-related responsibilities, and determine which Playbook chapters are most relevant to their needs. This allows readers to prioritize the content most applicable to them while gaining insight into how other audience groups may engage with AI.

*Table 1. Playbook Audience Groups and Relevant Chapters*

Audience Group	Example Roles	Responsibilities	Relevant Chapters
<b>Leadership and Managers</b> 	<ul style="list-style-type: none"> <li>Executives</li> <li>Division director</li> <li>Program manager</li> <li>Policy advisor</li> <li>AI Cross-Cutting Initiative (CCI) representative</li> </ul>	<ul style="list-style-type: none"> <li>Drive AI initiatives that improve operational efficiency.</li> <li>Identify strategic opportunities and measure return on investment.</li> <li>Ensure AI solutions align with CMS' mission and evolving healthcare needs.</li> </ul>	Chapters 2, 3, 4, 5, and 6 offer strategic guidance for advancing AI initiatives agency-wide and overseeing adoption.
<b>AI Project Teams</b> 	<ul style="list-style-type: none"> <li>Product and project manager</li> <li>Data scientist</li> <li>AI/machine learning (ML) developers</li> <li>Product designer</li> <li>Human-Centered Design researcher</li> </ul>	<ul style="list-style-type: none"> <li>Research and design AI solutions</li> <li>Develop and deploy Human-Centered AI solutions.</li> <li>Align AI roadmaps with leadership objectives and user research.</li> </ul>	Chapters 3, 4, and 5 offer practical guidance and tools for the implementation of AI projects.
<b>IT and Security Teams</b> 	<ul style="list-style-type: none"> <li>DevOps engineer</li> <li>Cybersecurity analyst</li> <li>Cloud architect</li> <li>IT support specialists</li> </ul>	<ul style="list-style-type: none"> <li>Ensure infrastructure, cybersecurity, and compliance for AI systems.</li> <li>Implement risk mitigation techniques and maintain system integrity.</li> <li>Support secure AI integration into CMS' existing technology stack.</li> </ul>	Chapters 3, 4, and 5 offer insight into the role of IT and security in operationalizing and sustaining AI systems.
<b>CMS Staff</b> 	<ul style="list-style-type: none"> <li>Healthcare policy analyst</li> <li>Customer service representative</li> <li>Contract specialist</li> </ul>	<ul style="list-style-type: none"> <li>Provide subject matter expertise related to CMS programs and operations</li> <li>Apply AI-enabled tools to streamline workflows and improve efficiency.</li> <li>Provide real-world feedback to refine AI usability and impact.</li> <li>Adjust processes to use AI-driven insights when possible.</li> </ul>	While general CMS staff are not the primary audience for most guidance provided, Chapters 2, 3, and 6 offer a big picture view of AI at CMS.

Once readers have a sense of their AI-related responsibilities and the Playbook chapters most relevant to them, they will be better equipped to understand how guidance is structured throughout the Playbook.

## 2. AI Primer

As teams across CMS pursue AI activities, developing a shared language of AI technical terms and organizational efforts is essential for effective collaboration and AI adoption. This chapter begins by defining AI terms used throughout the Playbook and explaining core AI mechanics. It then describes types of AI efforts and explains how different teams and functions at CMS contribute to and are affected by AI. Aligning terminology and expectations—both within this Playbook and in practice—fosters clearer communication and more effective coordination across the interdisciplinary teams it emphasizes.

### 2.1. Technical View of AI

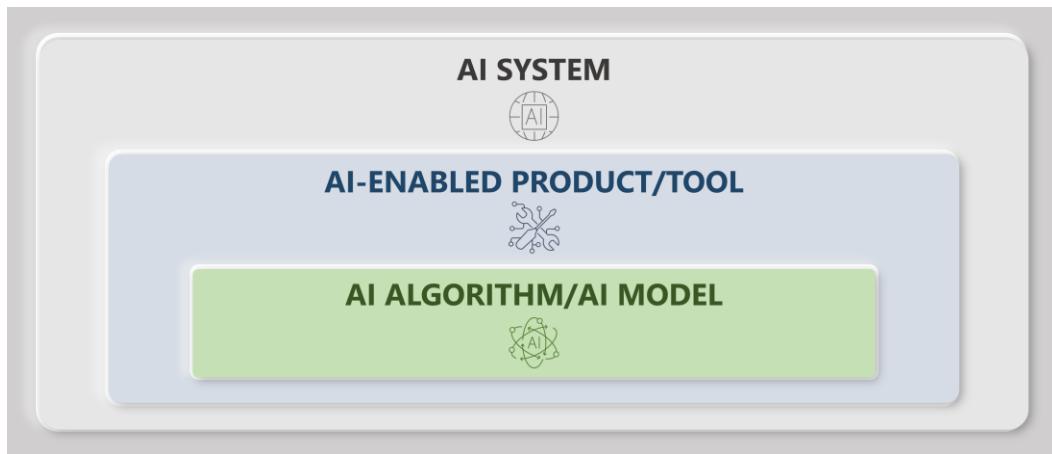
As described by the Assistant Secretary for Technology Policy and the Office of the National Coordinator for Health Information Technology, “Artificial Intelligence (AI) enables computer systems to perform tasks normally requiring human intelligence—for example, recognizing patterns, learning from experience, drawing conclusions, making predictions, etc.” (ASTP/ONC, 2025). In practical terms, AI includes both simple systems that follow programmed rules to make basic decisions and advanced systems that learn from data to recognize patterns, make predictions, and create new content, thereby performing tasks that once required human thinking.

#### 2.1.1. Elements of an AI Solution

The elements of an AI solution include:

- **AI systems:** setups that integrate various components
- **AI products and tools:** applications that use AI for specific functions
- **AI algorithms or models:** the basic computational methods

Figure 1 shows how an AI system encompasses an AI product or tool, and how AI algorithms and AI models form the basic computational method within a product or tool.



*Figure 1. Relationships Between AI Solution Elements*

Understanding the relationships among these elements and aligning on their definitions (provided in Table 2 below) ensures teams share a mutual understanding that is foundational for effective communication and coordination.

Table 2. AI Elements and Definitions

AI Element	Definition
<b>AI System</b>	A machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations or decisions influencing real or virtual environments. Artificial intelligence systems use machine and human-based inputs to: (A) perceive real and virtual environments; (B) abstract such perceptions into models through analysis in an automated manner; and (C) use model inference to formulate options for information or action (US Code, 2025).
<b>AI Tool or Product</b>	A software or application that incorporates AI-powered capabilities to enhance its functionality but does not operate as an independent AI system. These tools use AI models or algorithms to assist users in decision-making, automation or analysis, often as part of a broader workflow (Office of Management and Budget (OMB), 2025).
<b>AI Algorithm</b>	A set of mathematical or logical instructions that define how an AI system analyzes data, identifies patterns, and produces outputs. AI algorithms serve as the building blocks of AI models, ranging from rule-based logic to complex machine-learning techniques that improve over time (NIST, 2023).
<b>AI Model</b>	A specific component of an AI system that applies computational, statistical, or machine-learning techniques to process data and generate outputs. AI models are trained on datasets and can vary in complexity, from simple predictive models to advanced machine-learning architectures that evolve through continuous learning (DHS Artificial Intelligence Roadmap, 2024).

### 2.1.2. AI Model Mechanics

AI mechanics is a complex subject spanning fields from hardware accelerators to mathematical optimization (Vipra & Myers West, 2023). A *general* understanding of AI mechanics can support a team in evaluating key factors when selecting an AI approach, such as differences in cost, or why one approach may carry greater risks than another.

This section will highlight two of the most significant AI approaches. While this information is not all-encompassing, it is meant as a starting point to support teams having initial conversations about which AI approach might be best suited for their project. The first approach is ML AI, the approach behind most state-of-the-art models, including modern chatbots (Stanford University, 2021). The second is symbolic AI, an approach that once defined the field and historically received the most funding and research interest of any approach (Stanford University, 2021). While symbolic AI historically defined the field of AI and ML dominates the field today, a hybrid approach is gaining traction as researchers explore ways to combine symbolic AI with ML AI (Stanford University, 2021).

**In the ML approach, models “learn” by identifying patterns in training datasets.** The amount of data required to do this can be substantial, and the cost to train a model from scratch can be prohibitive. However, generically trained versions of these models can be fine-tuned to specific use cases for cost-effective solutions. These models are limited by the quality and relevance of their training data, and their performance can degrade over time as environments change, though retraining with new data may improve their performance. Additionally, interpreting why these models produce specific outputs can be challenging and uncertain. The ML approach has attracted most of the research and attention for more than a decade and is often the best starting point for AI projects today (Stanford University, 2021).

**Symbolic AI uses logic rules to manipulate pre-defined symbols that represent a specific problem.** The symbols can represent various elements such as concepts, objects, actions, and relationships. Effectively defining symbols for non-trivial tasks is difficult, and as a result a model's brittleness tends to scale with problem complexity. This approach is experiencing a revival by research teams attempting to use ML to define symbols that are then manipulated with logical rules (Stanford University, 2021). This approach may eventually offer advantages over the ML alternative, including lower costs from reduced computational needs and greater explainability of model outputs. For now, it generally remains the more brittle approach and is not recommended as the best starting point for most projects (Saad & Elson, 2025).

The following table outlines the characteristics of ML AI and symbolic AI, offering teams and leaders a comparison to help determine which approach best aligns with their project needs.

*Table 3. Comparison of ML AI and Symbolic AI*

Characteristic	ML AI	Symbolic AI
<b>Construction</b>	Initial version of the model is uncalibrated and cannot generate meaningful outputs. Model "learns" by determining how training data is distributed.	Symbols are manually created to capture relevant aspects of a domain, such as concepts, objects and actions.
<b>Risks</b>	Mismatch between training data and operational environment leads to limited model performance. Changes in operating environment relative to training data, can cause model performance to degrade.	Symbols do not adequately capture problem complexity or subtlety, leading to a brittle model.
<b>Adaptability</b>	Possible to periodically retrain the model to adapt it to a changing environment.	Symbols must be manually recalibrated.
<b>Explainability</b>	Very challenging to track how inputs lead to model outputs.	Relatively easy to track process used by the model to derive conclusion.
<b>Utility to AI Projects Today</b>	Generally, the best starting point for most AI projects at this time.	Unlikely to be a good starting point for majority of AI use cases today.

Both ML and symbolic AI present distinct implementation challenges. Effective use of either approach requires a clearly defined use case, access to appropriate data or knowledge sources, monitoring of the models, and refinement. While these complexities may limit immediate adoption, they should not deter teams from exploring AI solutions. With appropriate planning (see Section 5.1), iterative development (see Section 2.2.1), and established governance practices (see Chapter 4), teams can identify viable opportunities for effective AI development.

## 2.2. Organizational View of AI

With an initial shared technical understanding in place, team members can shift focus to align on how AI work is organized and managed within CMS.

### 2.2.1. Types of AI Efforts

Understanding the distinctions between AI initiatives, projects, products, pilots, and proofs of concept will help clarify what an AI effort encompasses, and which agency roles drive their execution. Alignment on these types of efforts, depicted in Figure 2, will support planning, execution, and scaling efforts within the agency.

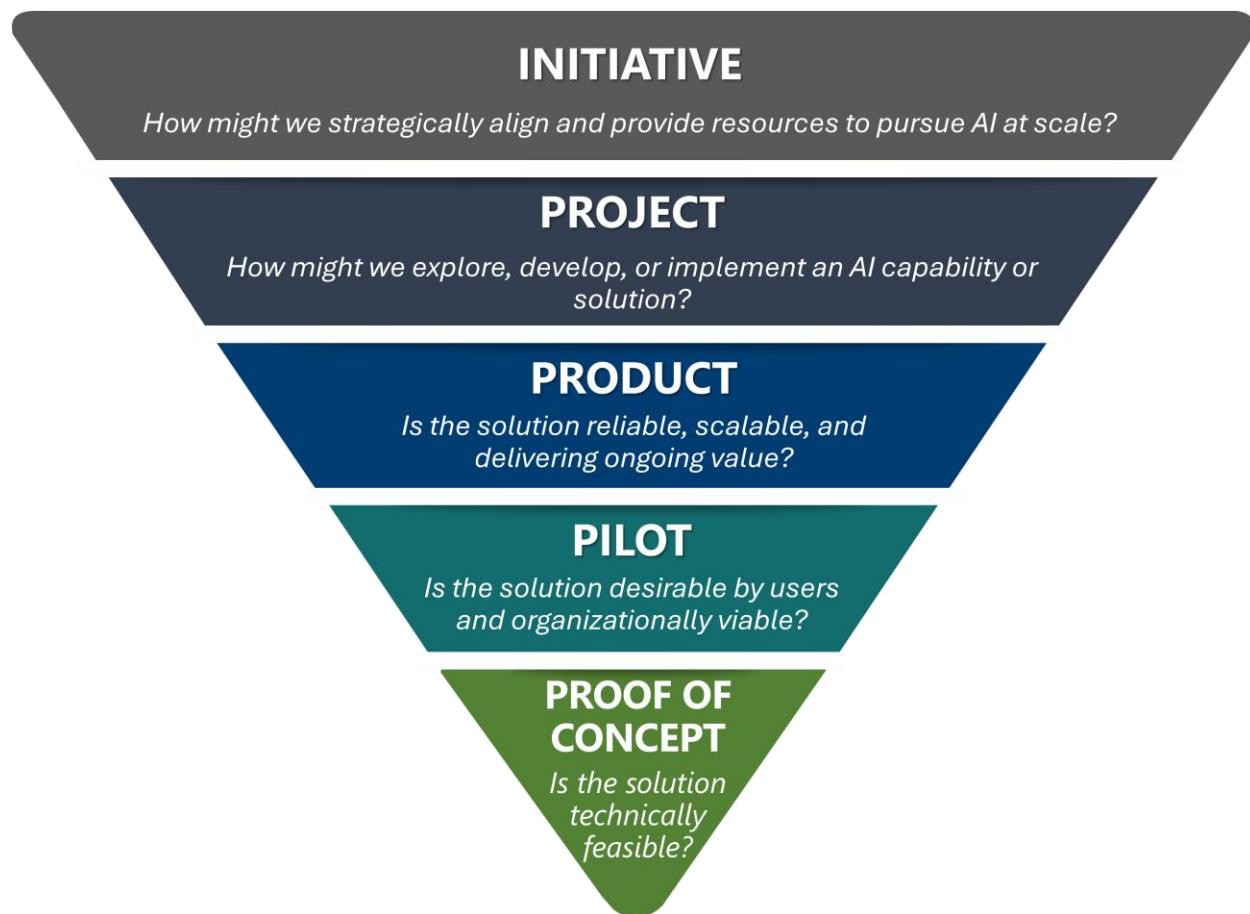


Figure 2. Types of AI Efforts

Overall, AI initiatives are broader in scope, driven by leadership, and intended to set organizational AI priorities. AI projects explore specific AI capabilities and are led by cross-functional teams. They can encompass products, pilots, and proofs of concept. Products, pilots, and proofs of concept are more technical in nature and focus on building, testing, and validating AI solutions. Table 4 provides the features of each of these AI efforts and an example of each at CMS.

Table 4. Types of AI Efforts

AI Term	Features	Example
AI Initiative	<ul style="list-style-type: none"> <li>A long-term, strategic effort within CMS to explore, adopt, or scale the use of AI.</li> <li>Sets priorities, provides resources, and establishes governance frameworks.</li> <li>Can include multiple AI projects, potential AI products, and various CMS components.</li> <li>Driven by leaders and managers as they collaborate with AI project teams, IT and security teams, and technical subject matter experts (SMEs).</li> <li>Answers the question: <i>How might we strategically align and provide resources to pursue AI at scale?</i></li> </ul>	The AI CCI at CMS facilitates AI infrastructure development, AI governance, and workforce training. It advises on policy recommendations for healthcare settings and promotes information sharing on relevant AI topics.
AI Project	<ul style="list-style-type: none"> <li>A focused, time-bound effort within an AI initiative to explore, develop, or implement a specific AI capability or solution.</li> <li>Follows a structured lifecycle with defined goals</li> <li>May result in an AI product.</li> <li>Led by a team that may include data scientists, data engineers, and human-centered design (HCD) researchers.</li> <li>Answers the question: <i>How might we explore, develop, or implement an AI capability or solution?</i></li> </ul>	OIT's development of the AI Workspace and CMS Chat, which created both infrastructure and products to meet diverse AI needs across different user groups at CMS.
AI Product	<ul style="list-style-type: none"> <li>A developed and deployed application or system that uses AI to deliver value to users</li> <li>Meets operational standards, serves defined use cases, and requires ongoing maintenance and support.</li> <li>Includes minimum viable products (MVPs) that are used to validate core functionality, gather user feedback, and guide future development through real-world use.</li> <li>Led by a team that may include product managers, data scientists, data engineers, and HCD researchers, user experience (UX) designers, and software developers.</li> <li>Answers the question: <i>Is the solution reliable, scalable, and delivering ongoing value?</i></li> </ul>	CMS Chat, a secure, custom generative AI tool that assists employees with their day-to-day work to increase productivity and efficiency.
AI Pilot	<ul style="list-style-type: none"> <li>A testable prototype with limited functionality that allows real users to test its value in a business setting.</li> <li>Validates business viability and user desirability before production.</li> <li>Typically runs for weeks to months.</li> <li>Led by a team that may include product managers, data scientists, data engineers, and HCD researchers, UX designers, and software developers.</li> <li>Answers the question: <i>Is the solution desirable by users?</i></li> </ul>	Limited release of CMS Chat to a select group of users to validate usefulness and gather feedback before wider deployment.
Proof of Concept (POC)	<ul style="list-style-type: none"> <li>An early-stage prototype focused on validating the technical feasibility of an AI approach.</li> <li>Not intended for end users</li> <li>Created in code/data science notebooks or development environments.</li> <li>Conducted by small technical teams that include product managers, data scientists or machine learning engineers.</li> <li>Answers the question: <i>Is the solution technically feasible?</i></li> </ul>	Initial testing of large language models within CMS' secure environment to validate technical feasibility for what would later become CMS Chat.

## CMS Artificial Intelligence Playbook

AI efforts across the spectrum—whether they are grassroots proofs of concept or top-down initiatives—are essential to advancing the agency's organizational AI maturity. Regardless of the effort type, collaboration across roles and teams is critical for any effort's success. The next section further explores which teams are responsible for domains affected by AI efforts.

### 2.2.2. The Organizational Ecosystem of AI

AI efforts at CMS are not confined to a single team or technical group. They require ongoing collaboration across diverse roles—from leadership and product development teams to IT, security, and enabling functions like procurement, legal, and human resources. Each plays a vital role in ensuring that AI efforts are not only innovative but also secure, ethical, scalable, and aligned with CMS' mission.

Figure 3 illustrates this cross-functional collaboration by mapping core organizational functions around AI. These functions, such as change management, cybersecurity, data management, and talent, are shown surrounding the AI center-point to emphasize that successful implementation depends on the active engagement of multiple teams. While the audience groups and functions shown are not exhaustive or mutually exclusive, they highlight the breadth of coordination required across CMS to effectively integrate AI into the agency's operations.

This holistic view reinforces that AI is not just a technical endeavor, but also an organizational one. Recognizing and fostering this collaboration is essential to ensuring alignment, security, and long-term viability in AI systems.

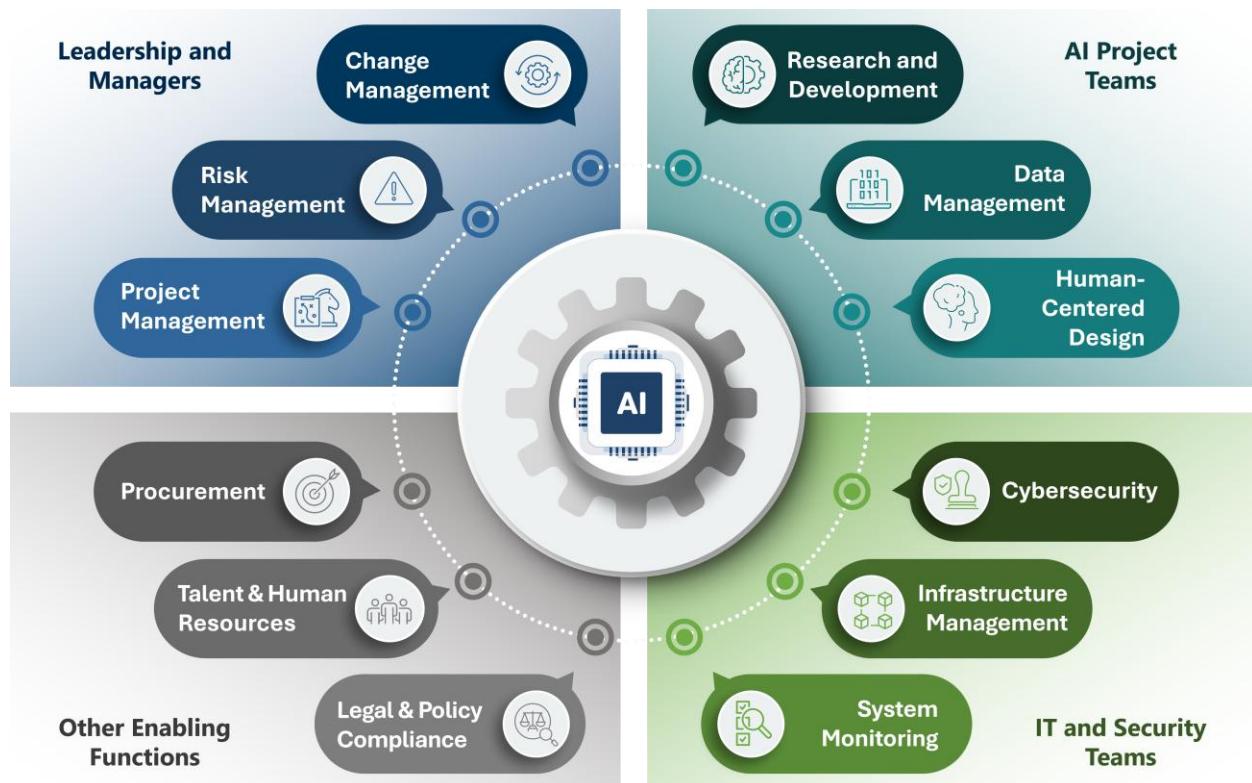


Figure 3. AI Organizational Alignment and Interdisciplinary Collaboration

## Key Takeaways - AI Primer

Chapter 2 provided a foundational overview of AI, including key definitions, technical approaches, and the organizational functions that enable effective implementation at CMS. While especially useful for leadership and managers, this chapter also offers valuable context for anyone looking to better engage with or contribute to AI efforts across the agency.



- Aligning on shared terminology and expectations when describing AI elements or efforts will enable more effective collaboration and AI adoption amongst interdisciplinary roles.
- AI systems encompass AI-enabled products and tools, and their underlying AI algorithms and models. Learning how these parts work, and how methods like ML AI and symbolic AI differ, helps teams make informed decisions when developing or selecting AI solutions.
- AI efforts at CMS require coordination across teams, departments, and functions to ensure alignment, security, and long-term success.

This foundation in AI concepts and organizational considerations establishes the groundwork for teams pursuing AI implementation across the agency. The next chapter provides a detailed look at CMS' current AI portfolio, maturity model, and guiding principles that can help shape AI initiatives and projects.

### 3. AI at CMS

At CMS, AI initiatives are uniquely shaped by the agency's mission and needs. As a healthcare agency, AI at CMS has the potential to influence the healthcare industry, policy, and the public. This chapter first examines the current state of AI implementation within CMS, then assesses the agency's organizational readiness and proposes a maturity model based on those findings. Finally, it proposes four guiding principles as a framework that teams can reference as they work toward advancing AI maturity.

#### 3.1. Current AI Portfolio

CMS' AI portfolio reflects a dynamic and evolving landscape with 66 use cases identified across 13 components in the fiscal year 2024 [HHS AI Use Case Inventory](#). These use cases draw on a range of primary AI capabilities (most notably machine learning and natural language processing), and covers all stages of development, from the *Initiation Stage* to the *Operation and Maintenance Stage*. This portfolio snapshot establishes CMS' current AI baseline, allowing insights into adoption patterns and strategic opportunities.

##### 3.1.1. AI Use Cases by AI Capability

In 2024, 66 AI use cases were reported within CMS components. To analyze the diverse range of AI usage across CMS, use cases were classified based on seven categories of AI capabilities: Natural Language Processing (NLP), Computer Vision, Generative AI, ML, Deep Learning, Large Language Models (LLMs), and Symbolic AI.

Modern AI implementations often integrate multiple capabilities to achieve their objectives. To simplify classification, each use case was assigned a primary capability based on its dominant function. When descriptions were technically ambiguous, classification relied on the best available information.

Of the seven categories, four are currently represented in CMS' use case portfolio as primary capabilities: NLP, ML, Generative AI, and LLMs, with one case utilizing Symbolic AI as non-primary.

ML and NLP form the foundation of CMS' AI portfolio, demonstrating their alignment with CMS' operational needs, such as fraud detection, claims processing, and patient care improvement.

Figure 4 shows a significant increase in NLP and ML applications from FY23 to FY24, suggesting growing confidence in these capabilities and their effectiveness in delivering positive organizational outcomes. The emerging presence of Generative AI and LLMs highlights CMS' readiness to explore advanced and transformative AI technologies. A detailed breakdown of AI use cases by capability is illustrated in Figure 4.

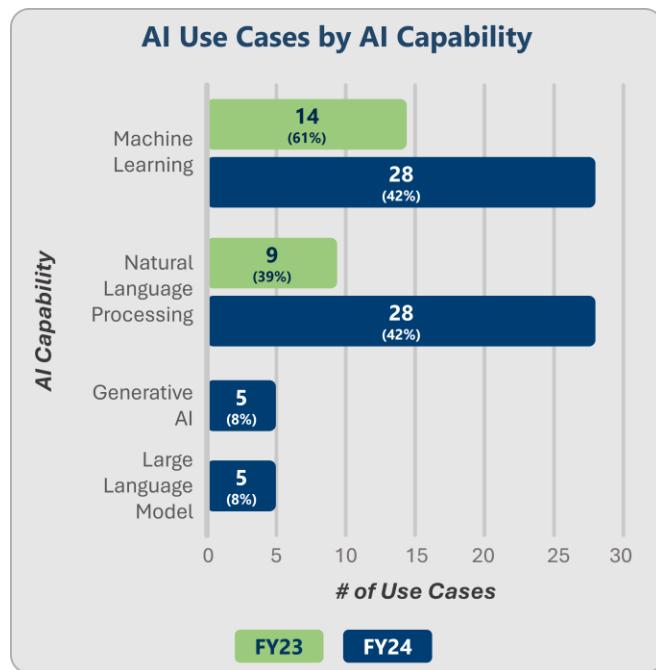


Figure 4. AI Use Cases by AI Capability

### 3.1.2. AI Use Cases by Stage of Development

Figure 5 below illustrates the distribution of use cases across stages, from *Initiation* to *Operation and Maintenance*. This distribution demonstrates that CMS is maintaining a consistent flow of AI projects and a maturing approach to AI-related efforts, with efforts spanning from initial exploration to deployment. Specifically, the large number of operational use cases indicates that CMS is successfully transitioning AI projects from innovation to stable, scalable implementations. Additionally, the presence of cancelled projects highlights CMS' willingness to experiment and learn, underscoring its commitment to AI innovation rather than indicating failure. Future AI projects can leverage the insights and lessons learned from these cancellations to improve feasibility assessments, resource planning, and implementation plans.

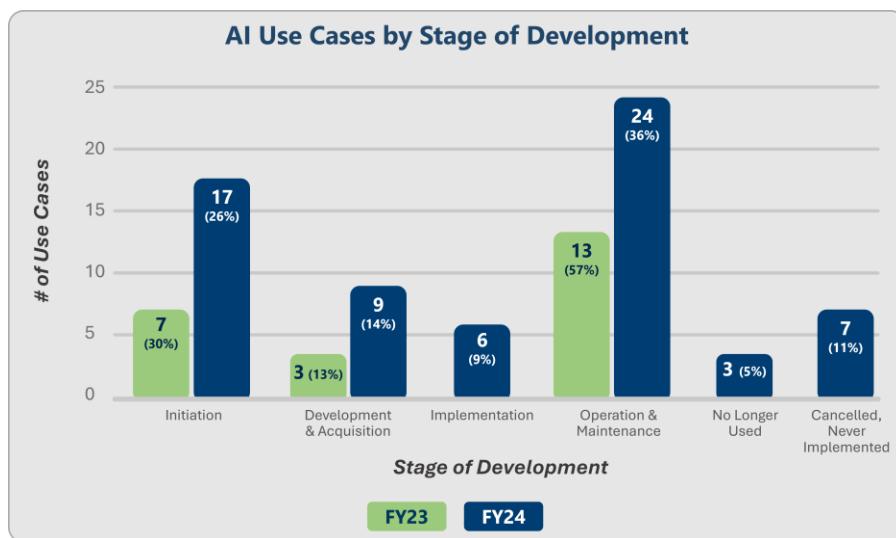


Figure 5. AI Use Cases by Stage of Development

## CMS Artificial Intelligence Playbook

By highlighting where and how AI is being explored and applied across the agency today, this portfolio overview helps illustrate the agency's evolving engagement with AI. Drawing from established maturity frameworks and insights from CMS experiences across its AI portfolio, the next section introduces a tailored maturity model that builds on these existing efforts and guides the agency's ongoing path for AI advancement.

### 3.2. AI Maturity at CMS

CMS' AI mission is to drive healthcare transformation through AI innovation, focusing on enhanced service delivery and operational effectiveness. As AI-related efforts increase in number throughout the agency, teams require a shared understanding of how AI maturity can be measured at CMS.

#### 3.2.1. AI Organizational Maturity Model

Building upon frameworks (industry and public) and informed by research across CMS conducted by OIT, the agency has created a CMS AI Organizational Maturity Model (OMM). This section provides an overview of the OMM and the following sections detail how teams can apply the model as well as where CMS currently lies within the model's stages.

##### *Understanding the Five Stages of AI Maturity*

The OMM charts the AI maturity path forward through five progressive stages and represents the vision for how the agency can progress in adopting, implementing, and scaling AI capabilities over time. It starts with *Level 1 - Exploratory*, where initial AI exploration and pilot projects take place, and ends in *Level 5 –Healthcare Transformation*, where AI fundamentally enhances healthcare delivery.

While individual teams and projects may operate at different maturity levels based on their specific contexts, the OMM serves as a roadmap to align the agency's collective journey toward AI transformation. It helps leaders and teams prepare for future opportunities while acknowledging the varying pace at which different parts of the organization may advance.

Figure 6 provides a visual of the maturity model and is followed by descriptions for each maturity level.



Figure 6. CMS AI Organizational Maturity Model

##### Level 1: Exploratory

In this initial level, the organization focuses on AI awareness and early experimentation. Teams begin exploring AI possibilities through proofs of concept and pilot projects, building foundational knowledge and identifying potential use cases. This stage is characterized by learning and discovery, with an emphasis on understanding how AI can benefit CMS operations.

## CMS Artificial Intelligence Playbook

---

### Level 2: Foundation Building

This level centers on developing core AI capabilities and infrastructure. The organization establishes basic frameworks for AI governance, data management and technical implementation. Teams work to create standardized approaches for AI projects while building essential technical and organizational capabilities.

### Level 3: Operational Integration

At this level, AI becomes integrated into operational workflows. The organization implements AI solutions that drive operational effectiveness, with established governance frameworks and policies guiding deployment. Teams begin to see measurable improvements in efficiency and service delivery through AI implementation.

### Level 4: Innovation and Scaling

This level focuses on expanding successful AI initiatives across the organization. The emphasis shifts to new product development and optimization, with AI solutions driving significant operational improvements. Teams work on scaling proven AI applications while continuing to explore new opportunities for innovation.

### Level 5: Healthcare Transformation

The final level represents the vision where CMS' maturity in AI implementation begins to influence broader healthcare transformation. As internal AI innovations enhance the agency's programs, policies, and operations, these improvements accelerate positive changes throughout the healthcare ecosystem. The efficiency and effectiveness of these AI-enabled processes create ripple effects that benefit providers, beneficiaries, and other stakeholders, while positioning CMS as a leader in effective AI adoption across government.

Progression through these stages is not always linear, and different projects or components of the organization may be at different maturity levels simultaneously. The goal is for teams to use this model as a guide for continuous improvement rather than following it as a strict sequential progression that requires uniform adoption across the agency.

### *Applying the Organizational Maturity Model*

While the maturity model provides a path forward for the agency as a whole, AI project teams and CMS internal stakeholders can also use it for various applications listed in the following table.

*Table 5. AI Organizational Maturity Model Applications*

Application	Description
Assessments	Teams can use the OMM to evaluate their current AI maturity level and identify areas for growth. For example, teams pursuing proofs of concept or pilot efforts may be at the exploratory stage, while those developing a product may be further along in maturity (see Section 2.2.1).
Planning	Teams can reference the OMM as a roadmap for AI initiatives or progressive AI development and use it to set realistic goals and milestones.
Communication Framework	The OMM offers common language and reference points for stakeholders to discuss AI initiatives and other AI efforts across the organization.
Resource Alignment	Teams can use the OMM to understand current and target maturity levels, enabling them to allocate resources and plan funding appropriately.

### *Evolution of the Organizational Maturity Model*

As CMS continues to gain experience with AI, the maturity model will evolve. Through ongoing research and engagement with teams across the organization, the AI Explorers team plans to develop more detailed guidance for specific organizational functions. This will help teams better understand their unique paths to AI maturity, including concrete actions they can take and specific goals they can pursue. As the model is refined based on real-world implementation experience, it will provide increasingly practical and targeted guidance for different groups within CMS.

### **3.2.2. CMS Maturity Assessment**

Research conducted by OIT found that teams across CMS are operating at varying levels of AI maturity. This variation is expected in the early stages of AI adoption, as components and teams differ in their capacity, expertise, and strategic priorities for advancing AI capabilities. An assessment of CMS' AI maturity indicates that the agency is currently positioned at *Level 2 (Foundation Building)* of the OMM, with several initiatives beginning to push toward *Level 3 (Operational Integration)* (see Figure 6). This section describes the needs and ongoing AI efforts identified as a result of OIT's research.

#### *Current Maturity Needs*

Despite the variation in maturity across the agency, several trends have been identified as relevant needs to be addressed before CMS can fully move into *Level 3 (Operational Integration)*. These shared needs are summarized in Table 6 below.

*Table 6. CMS AI Maturity Needs*

Need	Description
<b>AI Access and Literacy</b>	Technical and non-technical employees are broadly interested in AI tools and capabilities. They are eager to learn about and incorporate AI into their daily work, yet many lack an in-depth understanding of how AI works, including strengths and weaknesses. This highlights a need for broader AI literacy programs and accessible tools that all staff members can use.
<b>Training and Support for Technical Teams</b>	Many teams lack the specialized knowledge needed to effectively evaluate, develop, and deploy AI solutions. This creates a significant skill gap that limits CMS' ability to capitalize on AI opportunities.
<b>Infrastructure and Data Access</b>	Data scientists and engineers encounter barriers when attempting to access the computing resources, development environments, and data needed to build and test AI solutions. These infrastructure limitations create bottlenecks that slow innovation and implementation.
<b>Governance Frameworks</b>	Staff express a need for structured guidance on AI implementation. Effective governance would not only clarify boundaries and requirements but also actively enable innovation by providing clear pathways for AI development.

#### *Current Maturity Efforts*

These needs are actively being addressed through ongoing efforts across CMS aimed at advancing the agency's AI maturity. Table 7 highlights many of these key initiatives. Additional information about these efforts and guidance on how to get involved can be found in Appendix B.

Table 7. CMS AI Maturity Efforts

Effort	Description	Highlights
<b>AI Explorers Program</b>	A team within OIT which supports organizational maturity at CMS by conducting technical research and development, conducting policy research, and engaging the CMS AI community.	In addition to maintaining the CMS AI Playbook, the AI Explorers team released <b>eight white papers</b> featuring CMS-tailored operational resources in fiscal year 2024.
<b>AI Workspace</b>	A workspace available for CMS employees to instantiate a quick industry-standard AI in a pre-loaded cloud environment.	AI Workspace continues to provide CMS teams with a <b>secure cloud environment</b> for AI research, development, rapid prototyping, and experimentation, leveraging tools like open-source LLMs and Amazon Web Services (AWS) capabilities.
<b>AI Use Cases and Portfolio</b>	The number and variety of AI use cases at CMS.	CMS reported <b>66 use cases</b> across 13 CMS components in the HHS FY2024 <a href="#">AI Use Case Inventory</a> (see Section 3.1).
<b>AI Ignite</b>	A CMS internal micro-training program that provides hands-on, practical training to develop AI literacy across the organization and equips employees with skills to use generative AI in their daily workflows.	The AI Ignite program has <b>trained 4,700+ employees</b> in CMS Chat and generative AI. It strives to support every CMS employee by the end of 2025.
<b>CMS Chat</b>	A secure generative AI tool, custom-developed for internal CMS use.	<b>100% of CMS employees</b> have been granted access to use CMS Chat, with participation from all CMS components and the Office of the Administrator. CMS Chat adoption is supported by the AI Ignite program and a dedicated support channel in Slack which has grown to <b>1,000+ members since launch in January 2025</b> .
<b>Workforce Resilience (WR) Program</b>	Training offered to all CMS employees to learn new skills and technologies for the future, including but not limited to AI/ML.	In addition to AI content that is offered in the existing multi-course series, the WR Program has introduced <b>three new AI workshops</b> : <ul style="list-style-type: none"><li>• What are Large Language Models (LLMs)?</li><li>• How to Write a Better Prompt</li><li>• Doing Prompt Engineering</li></ul> Over <b>440 employees</b> have completed AI-related courses offered in this program.
<b>AI Cross-Cutting Initiative</b>	A team of strategic leaders and AI practitioners who develop and oversee various agency-wide initiatives and AI governance at CMS.	The AI CCI has become <b>CMS' cross agency AI strategy team</b> . Their activities have launched discussions to propose AI policies and governance frameworks for the agency. Additionally, the AI CCI facilitates infrastructure development, workforce training, and information sharing.

Effort	Description	Highlights
AI Community	A CMS-enterprise Slack channel that provides a space for CMS staff to share relevant AI news, policy changes, and technology advancements.	The AI Community Slack channel has grown from fewer than 80 members in 2021 to <b>900+ members in 2025</b> .

CMS' progression toward *Level 3 (Operational Integration)* is reflected by the AI efforts listed in Table 7. The AI Explorers team produces and shares tailored AI operational resources while the AI Workspace enables secure development and prototyping, both of which help teams move from exploration to implementation. Further, the growing number of AI use cases indicates that AI is moving beyond just experimentation and into regular use across CMS. The introduction of enterprise tools like CMS Chat, supported by AI Ignite training and corresponding operational resources (e.g., the CMS Chat Prompt Template (CMS AI Explorers, 2025) in Appendix B, demonstrates CMS's shift from basic AI awareness to actively embedding AI capabilities into everyday operations. Workforce training through AI Ignite and the WR Program equips staff with practical skills to adopt and apply AI tools. Meanwhile, the AI CCI and the expansion of the AI Community reflect progress toward structured governance and a more connected AI culture.

This research-driven assessment into CMS' AI maturity reveals both the challenges and opportunities ahead as CMS works to enhance its AI capabilities. While CMS has established foundational efforts and launched successful initiatives, the agency must now focus on systematically addressing these gaps to advance its organizational AI maturity.

To help guide individual contributions and practices that will accelerate the agency's AI journey, the AI Explorers team has developed four interconnected guiding principles that represent strategic focus areas for evolving from organizational readiness to meaningful impact.

### 3.3. Guiding Principles for AI at CMS

To guide CMS' continued advancements in AI maturity, this Playbook proposes a set of guiding principles to serve as the foundations for effective and collaborative AI adoption across the agency. These principles reflect CMS' commitment to scaling AI in ways that enhance operations, uphold public trust, and support federal priorities for strategic and innovative AI use in government.

The guiding principles are not abstract ideals—they are grounded in the real, ongoing work already taking place across CMS. From workforce training and governance development to pilot implementations and impact analysis, these principles are informed by the current needs of the agency as well as federal guidance to connect agency-wide strategy to day-to-day execution. They offer a shared framework for aligning leadership vision, team-level action, and cross-functional collaboration.

As depicted in Figure 7, CMS' guiding principles create a comprehensive framework where each builds upon and reinforces the others.

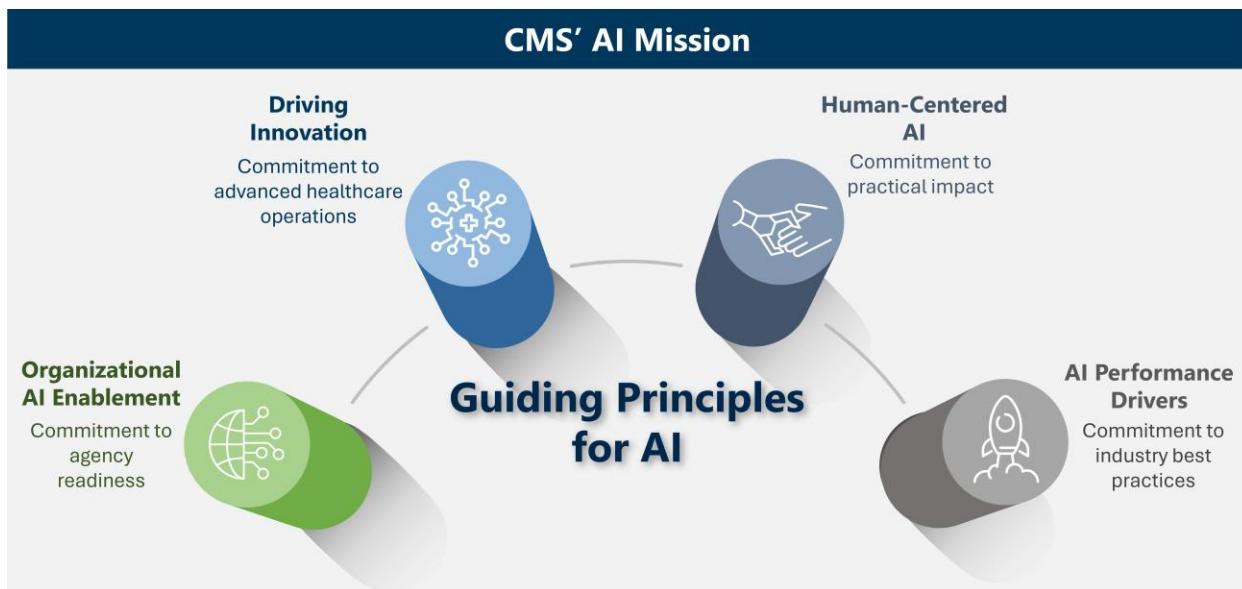


Figure 7. Guiding Principles for AI

- **Organizational AI Enablement** establishes the foundation by developing the technical infrastructure, human capital, governance structures, and organizational culture necessary for AI to thrive. By addressing these factors comprehensively, CMS will create an environment where both technical specialists and non-technical staff can effectively engage with and benefit from AI.
- **Driving Innovation** leverages this organizational readiness to generate value through AI experimentation and implementation. With the right enabling framework in place, CMS can more effectively explore, test, and scale AI-powered solutions that advance its mission and operational goals.
- **Human-Centered AI (HCAI)** ensures these innovations enhance human capabilities. By keeping people at the center of the AI design process, CMS will create solutions that augment its workforce's abilities, improve stakeholder experiences, and ultimately deliver better outcomes for the Americans it serves.
- **AI Performance Drivers** complete the framework by establishing standards for AI excellence and managing AI systems to be not just ethical but technically sound. This principle ensures AI systems at CMS are built and maintained with unwavering commitment to security, privacy, reliability, accountability, and transparency, creating a foundation of trust that enables sustainable and effective AI adoption.

Together, these guiding principles support CMS' AI maturity by establishing a strong foundation for AI efforts, promoting innovation, centering human needs, and providing guidance for an AI project's performance. In the sections that follow, each principle is defined, connected to key roles, and illustrated with example practices and current efforts. Challenges and opportunities for continued growth are also provided. While not every example below applies to every team, the principles guide how AI is approached across the agency.

### 3.3.1. Organizational AI Enablement

**Principle in Brief:** By aligning people, policies, systems, and skills, CMS creates the conditions for AI to thrive. Organizational AI Enablement takes shape through interdisciplinary collaboration, governance, infrastructure development, and workforce transformation at an agency level.

The *Organizational AI Enablement* guiding principle reflects CMS' commitment to equipping its workforce with the processes, tools, and support needed to implement and scale AI initiatives effectively in alignment with CMS' mission.

This first principle establishes the foundation for all other AI maturity efforts. With the right collaborative culture, governance structures, infrastructure, and workforce readiness in place, CMS can be well-positioned to scale AI in a coordinated and impactful way (Fountaine, McCarthy, & Saleh, 2019). Organizational AI enablement ensures that these internal structures are in place to support and sustain long-term success.

Leadership and managers are especially critical to advancing this principle. Their decisions around strategic planning, funding, staffing, and prioritization directly shape CMS' capacity to adopt and integrate AI. By setting a clear vision and investing in the long-term, leaders create the conditions for AI to take root and deliver value (see 6.2.4).

At the same time, progress depends on close collaboration with AI project teams and IT and security leads. These groups translate leadership's vision into practical systems, tools, and environments that lay the groundwork for AI to be embedded into core operations.

Table 8 introduces interdisciplinary collaboration, governance, infrastructure development, and workforce transformation as the domains driving organizational AI enablement at CMS.

Table 8. Organizational AI Enablement Domains

Domain	Description	Example Practices
Interdisciplinary Collaboration	Engagement across roles and departments to support cohesive AI initiatives and facilitate collaboration	<ul style="list-style-type: none"> <li>• Cross-functional AI working groups</li> <li>• Regular engagement between stakeholders and AI project teams</li> </ul>
Governance	Coordinated oversight that enables innovation, ensures compliance with emerging standards, and manages AI-related risks	<ul style="list-style-type: none"> <li>• AI intake forms and self-assessment questionnaires</li> <li>• Governance registries and tracking dashboards</li> <li>• Periodic project review schedules</li> <li>• Third-party procurement guidelines</li> </ul>
Infrastructure Development	Strategic investments in tools, systems, and infrastructure to support AI adoption and development across the organization	<ul style="list-style-type: none"> <li>• Compute investments (e.g., cloud platforms, on-premise processing hardware)</li> <li>• Secure data storage solutions</li> <li>• Model deployment platforms</li> </ul>
Workforce Transformation	Proactive change management for organizational shifts in workflows and employee readiness	<ul style="list-style-type: none"> <li>• Clear communication and feedback loops on AI's role and adoption progress</li> <li>• Workforce training / upskilling programs and bringing in AI talent</li> <li>• Providing access and support for using AI-enabled tools</li> </ul>

Organizational AI Enablement is being led through several fronts at CMS. The establishment of the AI CCI group and AI Community Slack channel demonstrates the agency's commitment to fostering collaboration and knowledge sharing. A governance framework is being developed and refined by the AI CCI (see Chapter 4). In Chapter 6, more details can be found for techniques and frameworks for aligning the Organization towards innovative AI success in the future. OIT has several programs that provide AI guidance and resources to CMS' AI community; AI Explorers white papers (see Appendix B) and CMS Chat that provide employees with implementation guidance and hands-on AI experience. This is complemented by other initiatives like the AI Ignite micro-training program, which trains employees on how to use CMS Chat for streamlining daily tasks.

CMS continues to approach AI enablement with careful consideration of its unique operating environment. The agency works to balance centralized oversight with federated needs, making sure AI efforts are aligned across the organization while allowing each division to address its specific goals. It also recognizes the appeal and challenges of hiring top talent (e.g., AI SMEs) to accelerate workforce transformation. CMS aims to continuously adapt its processes and guidance, such as this Playbook, to keep pace with evolving AI technologies and practices while maintaining operational requirements.

These foundational elements of collaboration, governance, infrastructure, and workforce readiness pave the way for broader organizational preparation. Section 6.2 expands on this by detailing how effective change management strategies can strengthen these enablement efforts.

### 3.3.2. AI Innovation

**Principle in Brief:** CMS supports AI innovation to improve service delivery, reduce costs, enhance and make internal processes more efficient, and ultimately evolve with the needs of those it serves. To drive innovation, CMS takes advantage of data resources, encourages experimentation, and strategically scales AI capabilities for greatest impact.

The *AI Innovation* guiding principle reflects CMS' commitment to leveraging the transformative potential of AI to enhance operational efficiency and develop new solutions that advance healthcare for all Americans. At its core, CMS aims to harness AI-driven innovation to optimize operations and improve healthcare outcomes across its programs.

Innovation through AI occurs along a spectrum, from incremental advancements that enhance existing capabilities to disruptive breakthroughs (Kennedy, 2020) that could upend traditional healthcare delivery models. For CMS today, the pursuit of incremental AI innovations will be more likely as the agency continues to grow its organizational AI maturity. These innovations will provide an iterative path toward realizing AI's transformative potential in progressive steps.

AI project teams play a central role in leading AI innovation through their hands-on work with data, AI development, research, and evaluation. While engaging in these projects, teams will collaborate with IT and security teams who provide the technical environments needed to test and deploy AI systems securely, as well as leadership and managers who play a key role in identifying strategic priorities, approving resources, and championing innovation across scaled initiatives.

Table 9 introduces data utilization, experimentation, and scalable impact as the domains driving AI innovation at CMS.

*Table 9. AI Innovation Domains*

Domain	Description	Example Practices
<b>Data Utilization</b>	Extracting value from CMS' vast data resources through AI	<ul style="list-style-type: none"> <li>• Updating data governance and data sharing policies for use with AI</li> <li>• Data mining, analysis, and predictive modeling</li> </ul>
<b>Experimentation</b>	AI exploration and controlled testing	<ul style="list-style-type: none"> <li>• Developing proofs of concept and pilot programs</li> <li>• Rapid prototyping and iterative development</li> <li>• Testing custom or commercial off-the-shelf (COTS) products</li> <li>• Shareable development environments and workspaces</li> </ul>
<b>Scalable Impact</b>	Pursuit of high-value opportunities that align with strategic goals and deliver measurable outcomes	<ul style="list-style-type: none"> <li>• Cost-benefit and return-on-investment (ROI) analysis</li> <li>• Road mapping and strategic planning</li> <li>• Portfolio management</li> <li>• Modular and flexible AI system design</li> </ul>

CMS has long recognized the value and opportunities that AI innovation can bring to its operations and healthcare services. The AI Explorers program, launched in 2021, and AI Workspace are clear examples of how the agency has invested in innovative AI research and development. Several use cases within CMS' AI portfolio have begun reporting positive ROI, such as OAGM's CMS Labor Analysis Wizard (CLAW) 2.0. The CLAW streamlines the evaluation process of new business proposals by analyzing proposed labor rates against historical pricing data and automates this process so that evaluators no longer have to conduct individual labor rate lookups. The tool informs smarter contracting decisions and saves evaluators more time to focus on negotiations, which has already demonstrated value by saving \$26M through FY29 in cost avoidance decisions.

AI innovation, however, is not without its challenges. The agency's extensive data resources often require significant preparation before use in AI applications. Measuring and reporting ROI across all AI projects remains complex, as teams work to develop standardized evaluation metrics. Through improved data management practices and evaluation frameworks, CMS continues to advance its AI capabilities while addressing these and other implementation challenges. Realizing sustained AI innovation requires more than technology. A proactive plan for aligning evolving policies, engaging people, and nurturing cross-functional collaboration can be found in Section 6.2.

### 3.3.3. Human-Centered AI

**Principle in Brief:** HCAI ensures that AI initiatives and projects are designed to maximize their effectiveness, usability, and real-world impact for the humans they serve. At CMS, human involvement and impact analysis are essential in the design of AI systems that are both operationally efficient and aligned with the practical needs of users.

The *Human-Centered AI* guiding principle reflects CMS' commitment to designing AI systems that are effective, efficient, and meet the real-world needs of the people they serve. At its core, this principle ensures that the people who build, use, or are affected by AI systems remain at the center of design, evaluation, and decision-making.

This approach emphasizes the importance of applying Human-Centered Design practices when designing, developing, and evaluating AI initiatives and projects. By engaging stakeholders and exploring root causes, values, needs, and potential impacts, CMS ensures that AI systems are designed to solve real-world challenges and deliver meaningful value in healthcare operations. The aim is to create truly human-centered solutions. This leads to more effective, efficient AI solutions that not only function as intended but also drive measurable improvements in service delivery.

AI project teams play a leading role in implementing Human-Centered AI by applying design research methods, integrating human feedback, and evaluating system impacts throughout the lifecycle of AI projects. These efforts are strengthened through support from leadership and managers who champion these design priorities and supply the appropriate resources, and IT and security teams who contribute critical expertise that informs system design, risk assessment, and ongoing iteration.

Table 10 introduces human involvement and impact analysis as the domains driving Human-Centered AI at CMS.

*Table 10. Human-Centered AI Domains*

Domain	Description	Example Practices
<b>Human Involvement</b>	Proactive stakeholder engagement, collaboration and oversight throughout the AI system lifecycle	<ul style="list-style-type: none"><li>• Discovery interviews (see Section 5.2.1)</li><li>• User experience design</li><li>• Usability testing and feedback mechanisms</li><li>• Iterative improvements</li><li>• Human-in-loop and human validation</li></ul>
<b>Impact Analysis</b>	Analyze potential risks and impacts of AI systems to ensure they operate as intended and avoid unintended consequences	<ul style="list-style-type: none"><li>• Research to understand socio-technical context</li><li>• Risk identification, analysis, and prioritization</li><li>• Mitigation tactics</li><li>• Metric development</li></ul>

CMS continues to develop valuable HCAI resources for AI project teams and contributors through programs like Workforce Resilience, which offers an HCAI workshop to CMS employees, and the AI Explorers team, which develops focused tools and guidance for HCAI, such as the HCAI Matrix (see Section 5.2.4). These resources are part of a broader effort to integrate human-centered thinking into the way AI is developed and deployed at CMS.

While some CMS teams are already using stakeholder input and impact analysis to improve their AI solutions, fully integrating HCAI into standard development processes is still in progress. CMS is working to close these gaps by building governance, sharing resources, and promoting human-centered design across the organization. The goal is to raise awareness and make HCAI a consistent part of how AI is developed. These efforts are part of CMS' larger journey toward AI maturity, where strong collaboration and culture matter as much as tools and policies. Achieving true human-centered outcomes depends on an organization's readiness to adapt roles, processes, and governance structures. Section 6.2 explores strategies to guide individuals and teams through these critical transformations.

### 3.3.4. AI Performance Drivers

**Principle in Brief:** Adopting AI Performance Drivers ensures AI systems at CMS maintain operational excellence while safeguarding sensitive healthcare data and stakeholder interests. These systems must follow standards and best practices for security, privacy, reliability, transparency, and accountability.

The *AI Performance Drivers* guiding principle reflects CMS' commitment to industry best practices that ensure AI systems are built for effectiveness and long-term viability. As AI becomes a stronger capability and more prevalent across CMS operations, this guiding principle becomes increasingly important. AI must perform reliably to deliver optimized value, and integrate smoothly into operations without causing inefficiencies or disruptions. Performance drivers establish essential criteria for all AI projects at CMS to be technically sound and well-governed.

For AI to be a productive and scalable tool, it must be designed to perform in real-world healthcare environments. This means maintaining security, protecting privacy, and generating trustworthy outputs that inform decision-making and program operations. AI systems should be interpretable and well-documented, with governance to guide their responsible use.

Ensuring AI systems perform effectively within the federal space, and often sensitive context of healthcare, requires technical depth and organizational coordination. IT and security teams play a key role in shaping how systems protect sensitive data, withstand threats, and evolve alongside emerging risks. Their collaboration with AI project teams ensures that performance considerations—such as accuracy, fine tuning, and explainability—are designed and tested with intention. Meanwhile, leadership and managers guide oversight efforts, make informed trade-offs across competing priorities, and promote accountability throughout the AI lifecycle.

Table 11 introduces security, privacy, reliability, transparency, and accountability as the domains driving AI performance standards at CMS.

*Table 11. AI Performance Drivers Domains*

Domain	Description	Example Practices
Security	Protecting AI systems and data integrity with strong cybersecurity measures and risk management approaches	<ul style="list-style-type: none"> <li>• Focused threat modeling</li> <li>• Access controls and encryption</li> <li>• Adversarial testing and system updates</li> <li>• Supply chain risk assessments for third-party and COTS products</li> </ul>
Privacy	Safeguarding personal or sensitive data to protect human autonomy, identity, and dignity.	<ul style="list-style-type: none"> <li>• User data controls and access restrictions</li> <li>• Data anonymization and encryption</li> </ul>
Reliability	Delivering consistent AI performance with accuracy and adaptability to changing conditions and data	<ul style="list-style-type: none"> <li>• Data validation</li> <li>• Model fine-tuning and performance benchmarking</li> <li>• Stress testing</li> <li>• Ongoing monitoring and maintenance</li> </ul>
Transparency	Ensuring stakeholders understand AI functions and limitations to support informed choices and human control	<ul style="list-style-type: none"> <li>• Open-source sharing and code transparency</li> <li>• Explainability tools and documentation</li> <li>• Feedback mechanisms and user support channels</li> </ul>

Domain	Description	Example Practices
Accountability	Establishing clear governance structures, roles and responsibilities for proper AI oversight and compliance	<ul style="list-style-type: none"> <li>• Governance frameworks</li> <li>• Periodic audits and impact assessments</li> <li>• Monitoring of agency and federal policies and regulations</li> </ul>

Addressing these performance domains individually within any governance framework or AI system does not eliminate risk. Trade-offs often arise—for example, improving transparency may reduce privacy. Since no AI system is entirely risk-free, CMS teams must weigh these trade-offs based on system context and agency goals. Formal, well-documented decisions should guide how performance considerations are balanced, based on each AI system's sensitivity and intended use.

As CMS advances in AI performance and compliance, it faces unique challenges. Protecting personally identifiable information (PII) and protected health information (PHI) while leveraging AI for healthcare improvement remains a top priority. CMS must also stay responsive to evolving regulations and ensure that third-party vendors meet the same high standards.

The agency is already making strong progress. CMS' existing security and data policies offer a solid foundation for protecting sensitive information in AI systems. Transparency and accountability remain central to its AI strategy, helping build trust across the agency and with external stakeholders. The AI CCI is refining a governance framework to oversee AI initiatives, incorporating key performance drivers and other critical factors. The suggested governance framework, introduced in the next chapter, is essential for managing AI growth, ensuring alignment with CMS' mission, and supporting effective oversight. Integrating security, privacy, reliability, transparency, and accountability into AI systems hinges on effective organizational change practices. Section 6.2 outlines change management frameworks that help ensure these performance drivers are maintained.

### Key Takeaways - AI at CMS

Chapter 3 provided a snapshot of CMS' latest AI portfolio, maturity level, and principles guiding ongoing AI efforts and strategy. The information is relevant for all audiences seeking awareness of CMS' overall AI status, and how various roles can contribute to AI at CMS.



- Leadership and managers can use the AI Organizational Maturity Model to evaluate current capabilities, set realistic goals, and align resources while providing a common framework for discussing AI initiatives.
- CMS' AI portfolio includes 66 use cases across 13 components, with the agency positioned at *Level 2 (Foundation Building)* maturity and moving toward *Level 3 (Operational Integration)*.
- Success in AI initiatives requires balancing innovation with responsible governance through strategic investment in infrastructure, workforce development, and oversight frameworks.



### KEY TAKEAWAYS

From Chapter 3

#### For AI Project Teams

- AI project teams should innovate by testing ideas carefully, using data effectively, and focusing on scalable, high-impact opportunities that support CMS' goals.
- Human-Centered AI should guide AI development, ensuring systems are designed with user needs in mind and include proper feedback mechanisms throughout the lifecycle.
- Current AI efforts at CMS include multiple development resources and collaboration tools, including the AI Workspace for secure development, CMS Chat, and the AI Community Slack channel (see Appendix B).



### KEY TAKEAWAYS

From Chapter 3

#### For IT and Security Teams

- IT and security teams play a key role in ensuring systems protect sensitive data, resist threats, and adapt to emerging risks. Their subject matter expertise and collaboration with AI project teams and leadership are essential to successful AI implementation.
- Infrastructure development and security considerations are critical enablers for AI maturity, requiring careful attention to data protection, system integrity, and compliance requirements.
- Security, privacy, reliability, transparency, and accountability are AI performance drivers that must be integrated into all AI systems at CMS.

Once CMS teams understand the AI landscape and maturity model, the next step is to explore how the agency is developing governance and managing its AI capabilities. The following chapter outlines a recommended governance framework, which ensures AI initiatives remain both innovative and responsible.

## 4. Governance

This chapter outlines an aspirational governance framework to serve as implementation guidance. As AI policy develops, the final shape of governance may evolve, but the concepts here reflect the current best practices and alignment with federal technical guidelines (e.g., National Institute of Standards and Technology [NIST], Federal Information Security Management Act [FISMA]) and broader federal direction (including the April 2025 OMB Memoranda M-25-21 and M-25-22) that reinforces the importance of balanced oversight and robust risk management in AI initiatives (U.S. Office of Management and Budget (OMB), 2025). This chapter integrates these federal guidance concepts with insights from the AI CCI, the AI Community Slack, the AI Explorers team, industry experts, and a diverse range of cross-agency stakeholders for a CMS specific governance approach.

This AI governance framework aspires to achieve four main objectives:

- **Drive Innovation and Opportunity:** Encourage high-impact, high-value AI projects that can yield meaningful improvements (e.g., cost savings, improved patient outcomes).
- **Ensure Responsible Oversight of High Impact AI:** Provide proportionate risk oversight mechanisms to address security, privacy, fairness, and compliance concerns.
- **Integrate with Existing CMS Structures:** Propose roles and processes that fit within CMS' organizational environment, in partnership with the AI CCI and other existing boards (such as the Technical Review Board (TRB)).
- **Support Continuous and Ongoing Monitoring:** Integrate continuous oversight with Quality Assurance Surveillance Plans (QASP) using regular performance metrics for tracking AI solutions. Utilize Independent Testing and Validation throughout the AI products' lifecycles for performance and risk mitigation.

All recommendations provided throughout the chapter at this time are strategic suggestions. This optional framework may be further tailored for adoption by any CMS center or office for robust but flexible AI oversight.

As such, CMS is currently working towards a cross-component approach further tailored to the organization's needs. Where formal CMS/HHS directives exist (e.g., HIPAA guidelines from HHS), those formal directives take precedence. For the latest updates and information on the AI Governance approach, CMS employees may reach out to the AI Explorers team at [ai@cms.hhs.gov](mailto:ai@cms.hhs.gov).

### 4.1. Governance Roles and Structure

Clearly defined governance roles are essential to ensuring AI initiatives at CMS are responsibly developed, aligned with agency priorities, and effectively managed across varying levels of risk and complexity. Leadership may define how governance roles are staffed or whether existing staff (e.g., Information System Security Officers (ISSOs)) can fulfill them.

## CMS Artificial Intelligence Playbook

---

Below, Table 12 introduces proposed roles and their details:

*Table 12. Governance Roles and Structure*

Role	Description
<b>CAIO:</b> Chief AI Officer	<p>The Chief AI Officer (CAIO) is a high-level executive responsible for driving the overall AI vision, strategy, and compliance for an organization or sub-organization. This role is required to be filled by all agencies within 60 days of the OMB M-25-21 memo issued on April 3, 2025 (U.S. Office of Management and Budget (OMB), 2025). This role ensures that AI initiatives align with the agency's mission, meet regulatory requirements, and deliver impactful outcomes. The CAIO works closely with governance bodies (e.g., AI Review Board [AIRB], AI Review Committee [AIRC]) and other organizational leaders to harmonize AI efforts across various programs and offices. Key responsibilities may include (page 10 of OMB M-25-21):</p> <ul style="list-style-type: none"><li>• Promote agency-wide responsible AI innovation and adoption in accordance with M-25-21 through a governance and oversight process</li><li>• Coordinate with other responsible agency officials to ensure that the agency's use of AI complies with applicable law and governmentwide guidance</li><li>• Serve as the senior advisor on AI to the head of the agency and within their agency's executive decision-making forums</li><li>• Represent their agency in and collaborate with coordination bodies related to their agency's AI activities, including external forums such as AI-related councils, standard-setting bodies, relevant governance boards, or international bodies</li><li>• Maintain the agency's AI Use Case Inventory</li><li>• Ensure processes are in place for the agency's high-impact AI use, including:<ul style="list-style-type: none"><li>– Establishing a process for determining and documenting AI use cases as high-impact</li><li>– Establishing processes to measure, monitor, and evaluate the ongoing performance and effectiveness of the agency's high-impact AI applications</li><li>– Overseeing agency compliance with requirements to manage risks from the use of AI</li><li>– Establishing a process for an independent review of high-impact use cases before risk acceptance</li><li>– Centrally tracking high-impact use cases and use case determinations</li></ul></li><li>• Advise on the transformation of the agency's workforce into an AI-ready workforce</li><li>• Ensure that custom-developed AI code and data used to develop and test AI are appropriately inventoried, shared, and released in agency code and data repositories</li><li>• Provide guidance on AI investments to the agency head and agency CFO related to resourcing requirements necessary to implement this memorandum</li><li>• Support agency efforts to track AI spending</li></ul>
<b>AIRB:</b> AI Review Board (Combination of AI Governance Board in M-25-21 expectations tied and CMS specific alignment)	<p>Panel that oversees AI governance and risk oversight, reporting, and complex evaluation/considerations for an organization or sub-organization. This board is required to be filled by all agencies within 90 days of the OMB M-25-21 memo issued on April 3, 2025 (U.S. Office of Management and Budget (OMB), 2025). To avoid confusion with other CMS boards (e.g., Architectural Review Board), this framework suggests the AIRB naming convention. The AIRB is recommended to include the CAIO and other senior technical leaders, AI legal and policy experts, cybersecurity, and program experts. Its high-level functions may include:</p> <ul style="list-style-type: none"><li>• Coordinate and govern issues related to AI use across the agency through regular meetings led by the Deputy Secretary or equivalent, supporting the CAIO in implementing governance requirements.</li></ul>

## CMS Artificial Intelligence Playbook

Role	Description
	<ul style="list-style-type: none"> <li>Review and approve high-risk or high-impact AI projects, evaluating whether potential benefits justify identified risks before authorizing deployment.</li> <li>Include cross-functional representation from IT, cybersecurity, data, budget, legal, privacy, civil rights, human capital, procurement, and program offices to ensure comprehensive oversight.</li> <li>Partner with the Chief Information Security Officer to ensure threat analysis and risk assessments addressing privacy, security, and other AI-relevant risks are conducted for major projects.</li> <li>Coordinate with existing governance structures like the Technical Review Board and leverage existing bodies where appropriate to avoid duplicative oversight mechanisms.</li> <li>Consult external experts from industry, academia, and sector-specific domains to identify innovative AI use cases and broaden governance perspectives.</li> <li>Establish governance structures within individual Components or Offices as needed to ensure appropriate oversight at all organizational levels.</li> </ul>
<b>AIRC:</b> AI Review Committee	<p>Small groups formed to coordinate with stakeholders and evaluate moderate to high-level projects and reports up to the AIRB. Duties may include:</p> <ul style="list-style-type: none"> <li><b>In-Depth Evaluations:</b> Examine moderate-risk or moderate-opportunity projects, referencing standardized forms such as Algorithmic Risk and Impact Assessment (ARIA) (see Section 5.2.4) or threat modeling (see Section 5.3.3) checklists.</li> <li><b>Periodic Rechecks:</b> Work with project teams to confirm ongoing compliance or to spot shifts in risk/opportunity.</li> <li><b>Collaboration with AIRB:</b> Escalate high-risk/high-opportunity proposals or any project that transitions to a risk/opportunity profile exceeding the AIRC's purview.</li> </ul>
<b>Warden:</b> Embedded Compliance Representative	<p>Individuals embedded in business operations who aid in identification of AI projects and oversee low to moderate-level evaluations and oversight and report up to the AIRC and AIRB. AI Wardens are also focused on day-to-day compliance and education for staff on the AI Governance Processes. AI Wardens are expected to:</p> <ul style="list-style-type: none"> <li><b>Ensure the governance steps</b> (e.g., ARIA questionnaires, threat model updates) are followed.</li> <li><b>First level evaluation</b> of low to moderate level risk. Responsible for elevating concerns to the AIRB and AIRC, when appropriate.</li> <li><b>Keep the risk registry</b> updated as usage, scope, or opportunity changes.</li> <li><b>Flag drifts from initial assumptions</b> or triggers that might prompt reclassification.</li> <li><b>Provide training and guidance</b> for governance processes and expectations.</li> </ul>
<b>Guide:</b> Embedded AI Liaison	<p>Individuals embedded in business operations who encourage growth and impact of AI across the agency by fostering collaboration and understanding. AI Guides are expected to:</p> <ul style="list-style-type: none"> <li><b>Help non-SME components</b> adopt AI effectively by serving as liaisons for best practices, feasibility insights or technical/logistical assistance.</li> <li><b>Share knowledge</b> of agency-wide resources, foster collaboration, and streamline the path to launching an AI pilot.</li> </ul>
<b>Organization:</b> AI CCI	<p>A cross-component group formed to collaborate on AI technology governance, training, collaboration, and acceleration. The AI CCI is exploring enterprise-level considerations to further tailor this framework to AI Governance at CMS. This draft aspirational framework has been developed in partnership with the AI CCI along with various stakeholders across the agency. The AI CCI may in the future adopt, formalize, and operationalize pieces of this approach if it determines the framework aligns with broader agency strategy.</p>

Role	Description
<b>Team:</b> AI Project Team	<p>Team of practitioners who complete appropriate governance documentation. This framework utilizes graduated team reporting to assess impact and innovation. The AI project team consists of development and operations members who are responsible for:</p> <ul style="list-style-type: none"> <li>• <b>Self-Nomination:</b> Should an AI project team be formed, the team is expected to either submit the project to the initiation system or to notify the appropriate Warden, AIRC, or AIRB members of the project's status. This is also true if an AI project shifts considerably in its scope. For example, if a project evolves from a proof of concept into an operational system with sensitive data, it will be required to complete additional reviews due to increased levels of risk.</li> <li>• <b>Documentation:</b> The team must complete all AI governance documentation to the best of their ability and at regular cadences (as requested by the AIRB, AIRC, or Warden).</li> <li>• <b>Collaboration:</b> The team is responsible for implementing mitigations for risks and innovation opportunities to the best of their ability so far as these expectations meet legislative and contractual expectations.</li> </ul>
<b>System:</b> Automated System	<p>Combination of forms, automations, registries, reports, and dashboards. To encourage efficiency, this framework recommends a well-architected system to maintain AI governance for the long-term. This system will have rules-based evaluation criteria to support in evaluation while also acting as the information hub for stakeholders. This system should form the basis of the Annual AI Use Case Inventory as required by OMB Memo M-25-21 (U.S. Office of Management and Budget (OMB), 2025).</p>

## 4.2. The Governance Process

Figure 8 is a visual description of this Playbook's example AI governance approach followed by details of each step. The stages of the process include:

1. **Intake:** The use of simple onboarding forms to determine if an identified project is an AI project, then to evaluate the appropriate initial level risk for the two-axis assessment.
2. **Two Axis Impact Assessment:** The use of graduated risk and opportunity assessment mechanisms (e.g., surveys, discussions, etc.) to determine appropriate governance actions for the AI project.
3. **Tailored Oversight Intensity (Frequency, Roles, Activities):** The roles, review frequency, and actions are assigned and completed for an AI Project given its oversight intensity level.
4. **Auto Re-Evaluation:** Continuous reassessment activities either scheduled by governance roles or through an autonomous schedule based on oversight intensity.

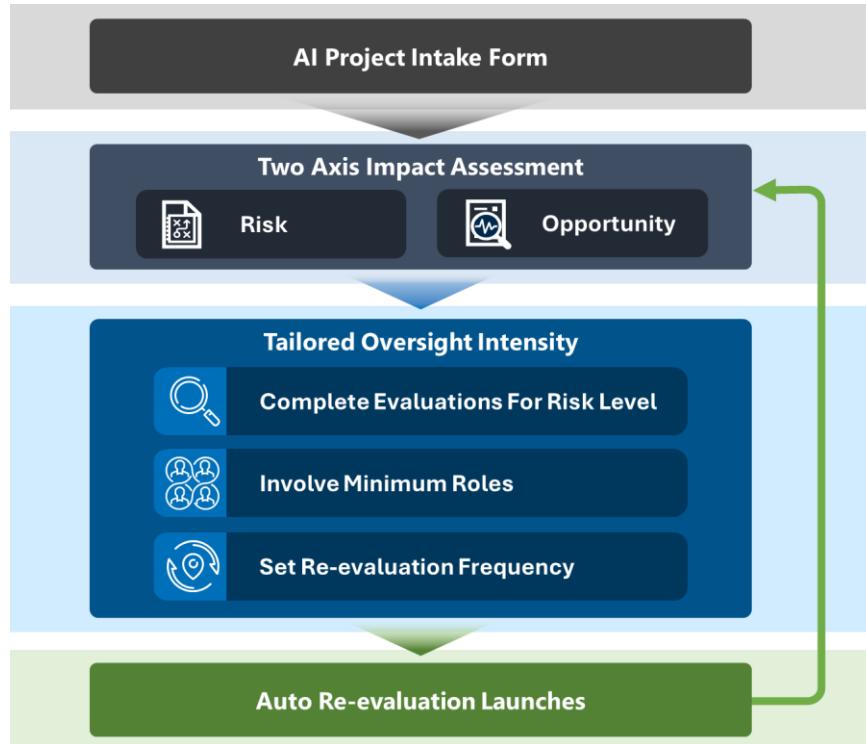


Figure 8. Governance Approach

Each of the stages above result in key information artifacts being created. This is further discussed in Section 4.3 Registries and Dashboards. These may include:

- **Project Registry:** Tracking data for all projects submitted to governance review. This may include key insights into the project such as stakeholders, risk level, opportunity level.
- **Risk Registry:** Tracking data for all risks identified in the governance process. This may include details such as risk status, details, mitigations, and involved stakeholders.
- **Dashboards and Reports:** Dashboards and reports tailored to specific users for information from the governance process and information artifacts.
- **Communications:** Notices that may include discussions, meetings, reviews, and notices to both AI project teams and evaluators to complete actions, or of upcoming re-evaluations.

### 4.2.1. The Two-Axis Approach: Balancing Opportunity and Risk

This framework introduces a two-axis perspective that addresses both risk and opportunity. This has been informed by OMB memo M-25-21, which directs agencies to use rigorous risk management practices and to ensure human oversight for high impact systems (U.S. Office of Management and Budget (OMB), 2025). Below, the two axes are defined as:

- **Risk:** Potential adverse impacts on impact rights, benefits, health, safety, or access to services.
- **Opportunity:** Determining feasibility, desirability, viability and potential benefits (e.g., cost reductions, improved user experience, advanced or novel insights, or better decision-making) using similar approaches to those found in 5.2.



*Figure 9. Two-Axis Impact Assessment Approach*

Table 13 breaks down example factors to consider for each axis, below:

*Table 13. Example Two-Axis (Risk x Opportunity) Evaluation Factors*

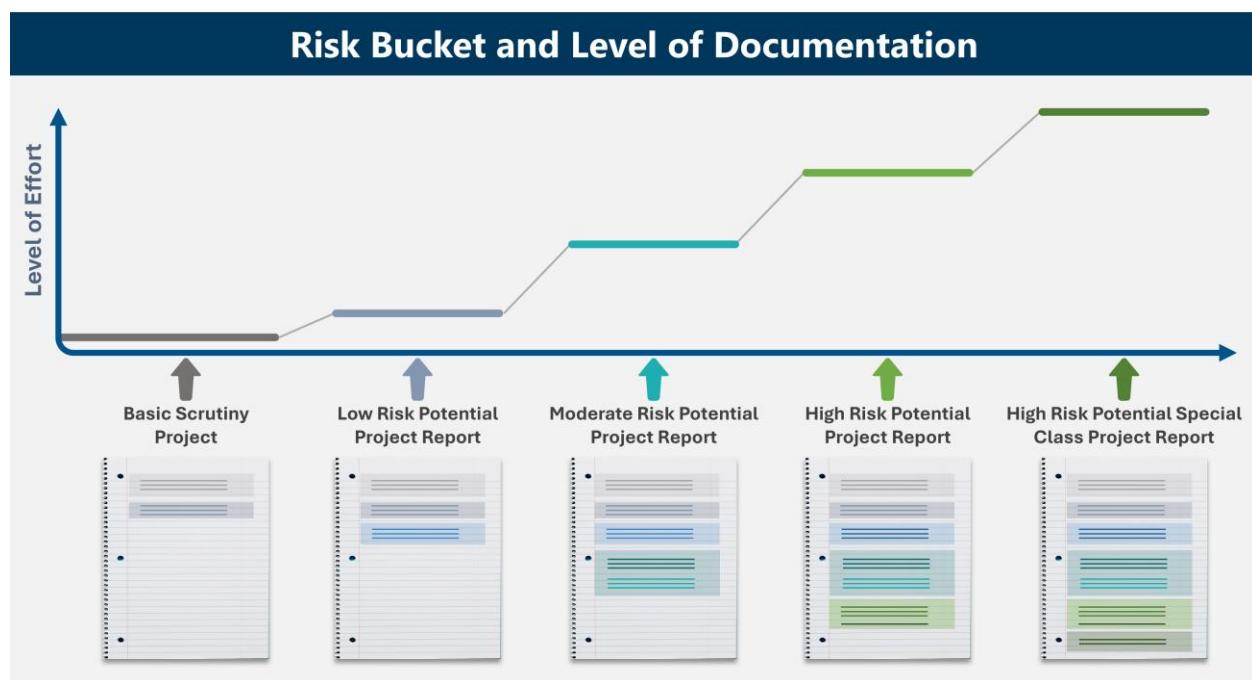
Evaluation Axes	Example Factors
<b>Opportunity Axis</b>	<ul style="list-style-type: none"> <li>Potential cost savings</li> <li>Enhanced beneficiary experience</li> <li>Operational efficiency improvements</li> <li>Novel insights or transformations (in line with “transformative research” concepts from NSF)</li> </ul>
<b>Risk Axis</b>	<ul style="list-style-type: none"> <li>Data sensitivity (PHI/PII)</li> <li>Scale of beneficiary impact</li> <li>Degree of automation (fully autonomous vs. human-in-loop)</li> <li>Legal/regulatory constraints</li> <li>Ethical or reputational concerns</li> <li>Harmful outcomes</li> <li>System inefficiencies</li> </ul>

Some projects may score each dimension on a numeric scale (e.g., 1–5) or a simpler classification (Low / Moderate / High) for each axis. Others may choose to use specific questions and choices to determine levels. Either way, the combined “Risk and Opportunity Profile” guides which path the project follows in the governance process. For more in-depth governance approach details that include questionnaire samples (see Appendix B). For additional documentation, reach out to the AI Explorers team at [ai@cms.hhs.gov](mailto:ai@cms.hhs.gov).

#### 4.2.2. Graduated Evaluation

After the initial evaluation is completed, projects are assigned a level of risk (low, moderate, high or special) or (1, 2, 3 or S) and have graduated requirements for additional documentation aligned to their risk level. This documentation is used to determine the potential impacts and risk of an AI project. By explicitly weighing opportunity alongside risk, CMS can prioritize resources for initiatives with higher transformative value, while still ensuring safe deployment. For instance, a “high-risk/high opportunity” AI project might unlock groundbreaking insights for program integrity but require stricter oversight. Conversely, “low-risk/low-opportunity” AI projects require minimal governance overhead but may offer less value to the agency.

To fully understand the status of an AI project, it may require additional documentation. Figure 10 demonstrates the increased documentation expectations based on an increasingly scrutinized AI project.



*Figure 10. Graduated Documentation Concept*

Table 14 below breaks down the additional documentation categories alongside additional context examples that a team would expect to collect depending on the level of scrutiny required of an AI project. The AIRB and AIRC review these materials to inform actions and decisions.

*Table 14. Example Information Collection for AI Governance*

Documentation	Example Contents
ARIA	<ul style="list-style-type: none"> <li>Documents potential algorithmic pitfalls, identification of data imbalances, or unintended/risky impacts (see Section 5.2.4).</li> <li>Helps teams systematically think through model design, data representativeness, and impartiality.</li> <li>Must be updated whenever the project scope significantly changes.</li> </ul>

Documentation	Example Contents
<b>Threat Model</b>	<ul style="list-style-type: none"> <li>Evaluates security and privacy vulnerabilities of models and algorithms (see Section 5.3.3).</li> <li>Must be updated whenever the project scope significantly changes.</li> </ul>
<b>Risk Level Scorecards</b>	<ul style="list-style-type: none"> <li>Summarize the project's risk, opportunity, potential ROI, compliance checks, and final recommended oversight path.</li> <li>Recommended to be graduated (or layered), where each tier has specific questions required of project teams. Additionally, the responses could indicate a need for more/less scrutiny, resulting in moving up or down a risk tier.</li> </ul>
<b>Data Documentation</b>	<ul style="list-style-type: none"> <li>Data Provenance: Clearly defining where data originates from and how it is processed.</li> <li>Inaccurate Data Mitigation Tactics: Documenting any steps taken to reduce data imbalance and ensure appropriate impacts.</li> <li>Compliance Alignment: Ensuring that data handling adheres to HIPAA, FISMA, and other federal regulations.</li> <li>Lifecycle Tracking: Maintaining records on how data evolves over time and how models are updated.</li> </ul>
<b>Risk Management Plan</b>	<ul style="list-style-type: none"> <li>Risk Categorization: Assigning risk levels based on data sensitivity, system autonomy, and potential for harm.</li> <li>Mitigation Tactics: Outlining steps to reduce high-risk factors (e.g., human-in-the-loop safeguards, data anonymization requirements).</li> <li>Escalation Procedures: Defining how high-risk issues are reported and addressed.</li> <li>Continuous Monitoring: Establishing periodic reviews to ensure risks remain manageable.</li> </ul>

To ensure AI systems continue to align with CMS objectives and compliance requirements, all AI projects in this governance framework undergo scheduled reevaluations and trigger-based reassessments.

#### 4.2.3. Scheduled Reviews by Risk/Opportunity Category

Projects should undergo periodic reviews. Table 15 below is an example of how reevaluation timelines may be structured:

Table 15. Example Reevaluation Cadence

Risk / Opportunity Profile	Example Review Cadence
<b>Low Risk / Low Opportunity</b>	Optional annual check or upon major updates
<b>Moderate Risk / Moderate Opportunity</b>	Every 6–12 months, or if data usage or user base expands significantly
<b>High Risk / High Opportunity</b>	Every 3–6 months, plus ad hoc re-checks if any major performance concerns emerge. Ideally these reviews are built into program management.
<b>High Risk / Low Opportunity</b>	Potentially more frequent reevaluations or <i>discouragement for low opportunity</i> unless strong justification arises.

Note that the exact timeline for review is flexible. A CMS component, AIRC, AIRB or AI Wardens may opt for more frequent audits depending on office, component or project needs. The key is to ensure continuous alignment with security, privacy, fairness, and benefit outcomes.

### *Trigger Events for Reassessment*

While this approach recommends autonomous re-evaluation, not all projects are likely to have triggers that automate initiation. Below are example triggers for evaluation (initialization or reassessment) for AI projects. Some of these examples may trigger with the use of autonomous systems and reviews while others are best initiated by humans acting within the governance loop:

- Expansion in scope (e.g., a pilot moves to full production with thousands of users)
- New data types introduced (e.g., PHI, social data, or other sensitive information)
- Statutory or regulatory changes
- Newly discovered vulnerabilities
- Shift in opportunity (e.g., a previously low-impact AI product may evolve into a high-impact analytics driver)
- Performance anomalies or metrics indicating potential harm to stakeholders
- Other factors as determined by CMS components, AIRB, AIRC, or Warden

## 4.3. Registries and Dashboards

A key requirement for this governance approach is appropriate information tracking and dissemination. This section covers various recommended information stores and reports to create and maintain throughout the governance lifecycle.

### 4.3.1. Registries

- **AI Use Case Inventory:** A single list of all AI projects identified and evaluated as part of this governance program. It can also serve as a resource to the CMS System Census and HHS AI Use Case Inventory when researching key details of AI projects for reporting purposes. This system should form the basis of the Annual AI Use Case Inventory as required by OMB Memo M-25-21 (U.S. Office of Management and Budget (OMB), 2025).
- **Risk Registry:** A registry for tracking risks identified throughout the governance evaluation approach.
- **Policy or Governance Criteria:** A centralized repository (e.g., SharePoint, Confluence) of all relevant legal requirements (e.g., HIPAA, Privacy Act, 42 CFR Part 2) and internal guidance (ARIA templates, threat modeling standards). This repository can help AI Wardens or leaders quickly check compliance and understand expectations.

### 4.3.2. Dashboards and Reports

Users of this governance approach can customize dashboards for different stakeholders. Example dashboard implementations may include:

## CMS Artificial Intelligence Playbook

---

- **Executives** see high-level metrics (e.g., number of AI projects, cumulative cost savings, risk distribution).
- **Wardens/AIRC** see detailed risk scores, upcoming re-check deadlines, risk indicators, etc.
- **AIRB** sees portfolio-level performance to help determine resource allocations or expansions.

This approach ensures that each stakeholder receives information that is appropriate to their role and level of responsibility.

### Key Takeaways - Governance

Chapter 4 described an AI Governance Framework intended as an aspirational, nonbinding guide for CMS components pursuing both opportunities (benefits, high-value innovation) and safeguards (risk management, compliance).

As formal CMS directives evolve, leadership and the AI CCI may adopt or modify this governance model. As a result, teams may encounter new or updated guidelines, official role assignments, or more detailed intake forms. This governance outline provides a clear and flexible structure to guide AI efforts across CMS, with all Playbook audiences playing a role in its implementation.



- Leadership and managers can contribute to CMS' efforts in developing its AI governance framework to meet four proposed objectives: driving innovation, ensuring responsible risk management, integrating with existing CMS structures, and supporting continuous monitoring.
- A governance framework requires establishing key roles including the AIRB, AI Review Committee (AIRC), AI Wardens, and AI Guides to ensure appropriate oversight and support.
- Regular re-evaluation and monitoring schedules should be implemented based on risk/opportunity profiles, with higher-risk projects requiring more frequent reviews.



- AI project teams should prepare appropriate documentation based on their project's risk / opportunity profile, including ARIA, threat models, and risk management plans.
- Project scope changes, new data types, or performance anomalies may trigger reassessment of AI projects through the governance process.
- Collaboration with AI Wardens and Guides will help ensure compliance with any formalized governance requirements while fostering effective AI adoption.



From Chapter 4

### For IT and Security Teams

- IT and security teams play a crucial role in threat modeling coordination, working with the CISO to analyze AI-relevant risks including privacy and security concerns.
- Implementation of appropriate registries, dashboards, and monitoring systems is essential for tracking AI projects and associated risks throughout their lifecycle.
- Regular security assessments and updates to threat models are required whenever project scope significantly changes.

While governance provides the framework for AI management at CMS, success of individual projects ultimately depends on effective execution. The next chapter offers a detailed roadmap for conducting AI projects, from initial conception through deployment, focusing on practical implementation approaches and organizational considerations.

## 5. Conducting an AI Project

Following the AI governance framework outlined in the previous chapter, this chapter provides teams with a structured approach for conducting an AI project at CMS. For reference, an AI project is a time-bound effort to explore, develop, or implement an AI solution and may encompass an AI product, an AI pilot, or proof of concept (see Section 2.1.2). While considering the steps described in this chapter, teams are encouraged to adopt an [agile mindset](#) by prioritizing small-scale implementations first before committing significant time and resources to larger efforts.

The chapter begins with key considerations teams should address before initiating a project. It then outlines the three key stages—research and approach, design and development, and deployment and integration—that guide teams through the practical steps of planning, building, and implementing AI solutions. Following these steps will support teams in implementing AI efforts effectively and contribute to CMS' broader goal of advancing AI maturity.

### 5.1. Starting a Project

Before any project kicks off, team members should have clarity on what to expect before getting started. The following sections offer an orientation for Conducting an AI Project, where contributors can develop an understanding of the level of effort the proposed AI project might require, the decision point to buy or build an AI product, the resources needed, and stakeholders the team may need to collaborate with across the agency.

#### 5.1.1. Stages

When conducting an AI project, AI project teams should prepare to engage in the following three stages shown in Figure 11. More information is provided on each of these stages in the subsequent sections of this chapter.



Figure 11. Stages of an AI Project

1. **Research and Approach:** First, the team will need to conduct discovery research to understand what the business problem is, the needs of the humans involved, and whether the team should consider AI for its solution.

2. **Design and Development:** Next, the team will need to decide whether to design and develop a custom product or procure a commercial solution. To validate the approach, teams should begin with iterative practices such as developing a proof of concept. The solution should ultimately balance user experience, technical requirements, system performance, and impact.
3. **Deployment and Integration:** Finally, the team will need to ensure the AI system is securely deployed and meets the AI Performance Drivers of security, privacy, reliability, transparency, and accountability while encouraging user adoption.

Based on research findings and available resources, some projects may not move forward to the next stage if they lack technical feasibility, user desirability, or business viability. Some projects may stop after initial research, while others may reach only the pilot stage before showing limited potential. This is expected, and AI maturity at CMS relies on teams being willing to embody an agile mindset, pursue innovative projects, and iterate on what they have learned.

### 5.1.2. AI Decision Framework

After reviewing the stages of an AI project, teams can use the AI Decision Framework introduced in Figure 12 to determine whether AI is the right solution and how to proceed with their project. A custom AI product is not always the best solution for a business problem, and the following steps help guide that decision.

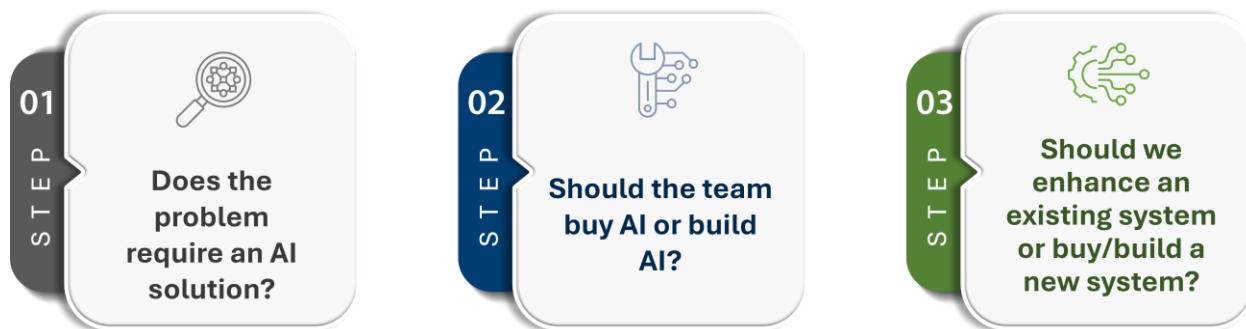


Figure 12. AI Decision Framework Overview

**Step 1. Determine if the problem requires an AI solution:** Before implementing AI, assess whether AI is necessary, feasible, and appropriate for solving the business problem through research and iteration from a proof of concept. Not all challenges require AI, and a traditional solution may be more efficient, cost-effective, and easier to maintain.

**Step 2. Determine if the team will buy a COTS product / service or build a custom AI solution:** Decide whether to purchase a COTS solution or to pursue custom AI development. Viability and desirability of either option can be informed through pilot testing, vendor demonstrations, or limited trials.

**Step 3. Determine if the solution will enhance an existing system with AI or transition to a new system:** Whether the team decides to buy or build AI, they will also need to determine where it will be introduced. In some cases, AI can enhance an existing system within the current infrastructure. In other cases, transitioning to a new system or product may be more effective.

Section 5.2.3 dives deeper into this AI Decision Framework with additional guidance and reasoning for each approach. Guidance that accounts for the variability in AI approaches is provided throughout Chapter 5, however, greater attention is given to custom-built solutions for CMS.

### 5.1.3. Team Skillsets

An interdisciplinary team is required for the effective design, development, deployment, and integration of an AI product. Assign clear roles to team members to ensure accountability; individual members may fulfill multiple roles if needed. Table 16 provides a list of recommended roles for an AI project team. Not all projects will require every role, and teams should determine which roles are necessary based on their decision to buy a COTS product / service or build custom AI.

*Table 16. Team Roles and Responsibilities*

Role	Responsibilities
<b>Business Owner (BO)</b>	Acts as the CMS liaison between stakeholders and the AI project team; defines the vision of the project or product and translates the vision into a prioritized list of features and requirements; supports vendor selection, contract negotiation, and budget management if there is a need to procure AI.
<b>Product Managers</b>	Set the product vision, roadmap, and success metrics; align business objectives with user needs; coordinate team members to manage product development, monitor performance, and continuously improve the product.
<b>Domain Experts</b>	Provide industry-specific knowledge that guides AI projects; support the product manager in helping to define project goals; validate model outputs; ensure alignment with real-world scenarios.
<b>HCD Researcher</b>	Conducts research via discovery interviews and workshops to uncover the needs, behaviors, and expectations of end users regarding the AI project or AI-based tool; continues to conduct concept and usability testing to guide iteration on the solution as the project progresses.
<b>UX Designer</b>	Designs user interfaces (UIs) to ensure the final product is user-friendly, intuitive, and aligned with user needs; collaborates with HCD researchers, end users, and developers to ensure the solution meets user expectations.
<b>Data Engineers</b>	Develop and maintain data pipelines to ensure data is accessible, clean, and usable for AI models; work closely with data scientists to provide necessary data infrastructure.
<b>Data Scientists</b>	Prepare data; develop and implement algorithms to extract insights from data; evaluate algorithm behavior, performance, and outputs against necessary criteria.
<b>Developers and System Architects</b>	Translate AI models into functional software solutions while managing system architecture; develop user interfaces for interaction; design the overall structure and framework of the system, ensuring scalability, efficiency, and reliability; collaborate to ensure seamless integration into existing systems.

### 5.1.4. Stakeholder Collaboration

When starting an AI project, teams must account for the various groups they will need to collaborate with across the agency. Effective coordination with Development Security Operations (DevSecOps), internal and external agency stakeholders, and end users is essential to ensure alignment with security requirements, policy expectations, and user needs. Balancing the priorities of these groups supports a smoother development process and increases the likelihood of a successful, well-integrated AI product. Table 17 lists roles of agency groups that the AI project team is expected to collaborate with.

*Table 17. Roles of Agency Groups*

Role	Responsibilities
<b>Development Security Operations (DevSecOps)</b>	Ensures that AI systems and infrastructure operate securely, with a focus on the technical implementation of continuous monitoring, threat detection, vulnerability management, and integration of secure coding practices throughout the AI development pipeline.
<b>Security Governance Stakeholders</b>	Ensure alignment with security policies, regulatory requirements, and organizational risk management approaches; collaborate with DevSecOps teams to implement necessary security measures and oversee compliance; establish risk-based security frameworks, ensuring that AI systems meet appropriate security levels based on their risk exposure (see Section 4.2).
<b>Internal Agency Stakeholders</b>	Represent various CMS entities with vested interests in AI projects by providing input and feedback which help determine, or evaluate, project requirements and success criteria. This role may include end users, executives, clients, regulatory bodies, and technical system owners, as well as teams that represent other groups, divisions, or components that the solution would impact.
<b>External Agency Stakeholders</b>	Provide input and feedback from parties separate from the agency. Depending on the scope and impact of the AI project, these stakeholders can include community members, beneficiaries, healthcare providers, and end users.
<b>End Users</b>	Provide feedback to the team on the tool's performance, usability, and features.
<b>Policy Experts</b>	Ensure that AI projects comply with relevant laws, regulations, and agency policies. They help interpret policy requirements, assess potential implications of AI use, and support alignment between technical development and broader policy goals.

Team member alignment on the progression of project stages, AI approaches, team member roles, and understanding with whom to collaborate with across the agency will set the stage for project success. Next, the team can begin research to define the problem and select the AI approach that best fits the project needs.

### 5.1.5. Case Study Example - Starting a Project

In 2024, project work started on CMS Chat, a secure, custom generative AI tool designed to enhance employee productivity and efficiency across a broad range of tasks. The initial MVP focused on providing a general-purpose AI tool, while future iterations will integrate with internal CMS knowledge bases and introduce specialized CMS Chat assistants and agents tailored to agency-specific needs. After each section in Chapter 5, a visual like Figure 13 will explain how the CMS Chat team applied the concepts from that section to develop and implement this evolving AI solution.

### Case Study: CMS Chat



Hi CMS Chat! I want to learn about how OIT followed the AI project stages to build you. Can you tell me what it took to **Start a Project**?



Certainly! I'll return at the end of each stage within Chapter 5 to revisit this example of how I was developed.

As early adopters and explorers of AI, OIT's journey to **Starting a Project** involved building a robust community of AI enthusiasts and subject matter experts across CMS. The AI Explorers program was established to research emerging technologies and develop proofs of concept that could address agency needs. For CMS Chat specifically, the groundwork was laid when the AI Explorers team conducted Retrieval-Augmented Generation (RAG) technology research in 2023. This research provided critical technical insights that would later become the foundation for CMS Chat's development. By leveraging this existing expertise and research, the CMS Chat project was able to move from concept to implementation more efficiently, building upon organizational knowledge rather than starting from scratch.



Figure 13. Starting a Project for CMS Chat

## 5.2. Research and Approach

Once a team has considered what is necessary to start a project, it can move to conducting research and selecting the appropriate AI approach. This section provides guidance on the research needed to identify the business problem, outline requirements, and explore AI approaches while implementing HCAI. Taking the appropriate steps to uncover stakeholder requirements, user needs, and technical aspects for AI implementation will ensure the solution addresses the business problem while keeping aligned to CMS' organizational goals.

### 5.2.1. Identify the Business Problem

If the business problem is not well-defined through research, the team risks wasting resources on solutions that fail to meet user needs, fall short of organizational goals, or neglect core requirements for effective AI Performance Drivers. By defining a precise business problem, teams can make sure that the AI investment delivers measurable value and improved outcomes.

#### *Conducting Discovery Research*

To identify the business problem, the team should conduct discovery research and can use the [USDS Discovery Sprint Guide](#) to support their discovery process. Through the discovery sprint, the team [interviews stakeholders and end users](#) to uncover stakeholder goals, user needs, and solution requirements. If from discovery research, the team determines that AI may be an appropriate solution, they can reference the HCAI Matrix (see Section 5.2.4) during the discovery research phase. The HCAI Matrix prompts teams to consider the human impact, human-AI interactions, and human needs throughout the AI project's lifespan.

### *Delivering Findings and Iterating*

Once the initial discovery sprint is complete, the team should synthesize findings (such as key themes from user and stakeholder interviews) into a [sprint report](#) and present the report to stakeholders. This documentation supports alignment on the business problem, user needs, and serves as a reference point throughout the project lifecycle.

Discovery is not a one-time activity. As the project progresses, the team should plan to revisit discovery research by conducting follow-up interviews or feedback sessions. These checkpoints help validate assumptions, refine requirements, and respond to emerging challenges or opportunities.

### *Determining Feasibility, Desirability, and Viability*

Incorporating user and stakeholder insights from the beginning and revisiting them throughout iterative development helps ensure the AI solution remains aligned with user needs and business goals. As depicted in Figure 14, a successful solution must be *desirable* (users want or need it), *feasible* (it can be built using available or emerging technology), and *viable* (it offers value through efficiency, scalability, or cost-effectiveness) (Vinney, 2022).



Figure 14. Feasibility, Desirability, Viability

To validate early assumptions, teams often develop artifacts such as a *proof of concept* (to test technical feasibility), a *pilot* (to evaluate viability and desirability in a limited real-world setting), and an *MVP* (to deliver core functionality with room for iteration) (see Section 2.2.1). These early-stage efforts help teams reduce risk and ensure the solution is grounded in practical needs before scaling.

As teams conduct research, they begin to uncover the specific business, functional, technical, data, and operational requirements needed to support a solution that is both effective and sustainable. The next section outlines these key requirement types and considerations in more detail.

## 5.2.2. Establishing Requirements

Starting to establish requirements during the discovery research phase lays the foundation for a project's success. Rather than attempting to define all requirements at the outset, a practice that can limit flexibility, an iterative approach allows project teams to start small, refine, and adapt project requirements as needed. This involves identifying needs across several categories—business, functional, technical, data, and operational—that collectively guide the project's design, development, and evaluation. Since many requirements are directly influenced by business objectives, requirements are likely to evolve over time and will require continuous reassessment to ensure alignment with project goals and any constraints.

Table 18 provides an overview of the different requirement types for AI project teams to consider and refine iteratively throughout their project.

*Table 18. Requirement Types*

Requirement Type	Description	Considerations	Impacts
<b>Business Requirements</b>	The business goals and objectives that the AI-enabled solution must achieve to be considered successful	Alignment with agency mission and priorities; expected outcomes and success metrics; stakeholder needs; policy and regulatory constraints	Ensures the AI solution delivers value to the organization; guides prioritization and resource allocation; supports stakeholder buy-in
<b>Functional Requirements</b>	The necessary features and functions for the end user to interact with the AI-enabled solution effectively	User roles and permissions; accessibility and usability standards; integration with existing workflows; system interactions and expected outputs	Enhances user experience; ensures the system meets operational and stakeholder needs; reduces usability issues
<b>Technical Requirements</b>	The infrastructure, tooling, and system specifications necessary to support AI development, deployment, and monitoring	Compute, power, storage, and environment specifications; scalability; security and compliance; integration with existing tools; cost efficiency; leveraging CMS resources such as Open Source Program Office (OSPO) and AI Workspace (see Appendix B); testing for performance, security, scalability, and integration; risk assessment	Prevents costly delays; ensures performance reliability, and security; optimizes resource allocation and operational efficiency; mitigates risks related to system failures, security threats, and compliance issues
<b>Data Requirements</b>	Datasets, quality standards, and governance needed for AI model training, validation, and deployment	Data sources, quality, and quantity; data collection, storage, and processing; compliance with privacy and security regulations	Fosters model accuracy; prevents security and compliance risks; optimizes data pipelines (reducing latency and bottlenecks); enhances decision-making and operational efficiency
<b>Operational Requirements</b>	The background operations essential for maintaining continuous functioning, efficiency, reliability, availability, and security of the system over time	Monitoring and logging; incident response and troubleshooting; system updates and maintenance; user training and support; scalability and performance management; threat modeling	Ensures system reliability and availability; reduces downtime; enhances security and operational efficiency; mitigates operational security risks

### 5.2.3. Implementing the AI Decision Framework

After defining the business problem and establishing requirements, the team must determine the most effective approach to address their problem. This begins with assessing whether the problem requires an AI solution or can be solved without AI. These decisions should be guided by the feasibility, desirability, and viability of the proposed solution.

The decision framework and supporting questions depicted in Figure 15 will help a team navigate selecting the appropriate approach for their project.

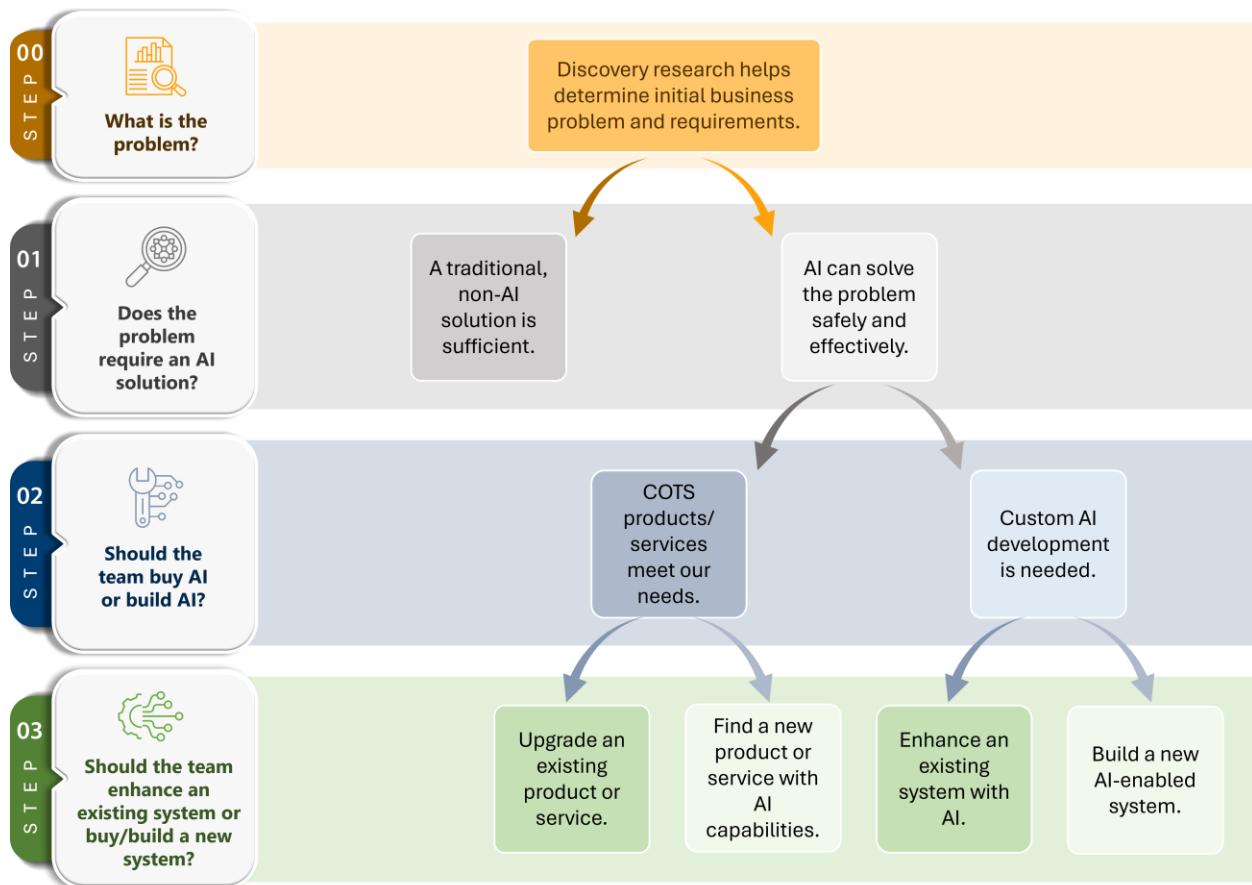


Figure 15. AI Decision Framework Process Flow

*Step 1: Determine if the business problem requires AI in the solution.*

Before adopting AI, consider whether the outcome will truly benefit from AI and if a traditional solution can effectively address the problem instead. Non-AI solutions are often more resource efficient and require less time, cost, and complexity to implement and maintain.

*Table 19. Indicators For and Against AI Suitability*

 <u>Indicators for AI Suitability</u>	 <u>Indicators against AI Suitability</u>
<ul style="list-style-type: none"> <li>• Problem requires adaptive, self-learning capabilities that are inappropriate for a rules-based system.</li> <li>• Involves unstructured data (e.g., text, images, videos).</li> <li>• Demands autonomous actions.</li> <li>• Demands large-scale data analysis, dynamic decision-making, or adaptability.</li> <li>• Technology is reusable once it is built.</li> </ul>	<ul style="list-style-type: none"> <li>• Domain is extremely sensitive to ethics and accountability.</li> <li>• Data is insufficient in volume, quality, or relevance to train or support an effective AI system.</li> <li>• Organization lacks capacity or resources to support ongoing human oversight, model monitoring, error handling, and system updates after deployment.</li> <li>• Costs less to address the problem manually than to invest in building AI for one-time use.</li> </ul>

When determining if the problem requires AI, a team should consider developing a proof of concept to test key technical assumptions and explore whether AI is technically feasible in a low-risk environment. This early iteration helps reduce uncertainty before investing significant resources. If the team determines that the business problem does require AI, consider next whether to purchase a COTS product / service or if the solution will need to be custom built.

*Step 2: Determine if the team will buy a COTS product / service or build a custom AI.*

The decision to buy or build will significantly impact the team roles needed, processes, and requirements.

Purchasing a COTS product or service will require resources put into market research and evaluation. Using pre-built services will likely accelerate implementation, however, may impose limitations in customization, integration, long-term adaptability, and potentially lead to over-reliance on a single vendor. In contrast, building custom AI will require deeper investment into technical roles and agile development processes, while offering flexibility tailored to specific needs. These procurement considerations align with the Administration's April 2025 guidance on eliminating barriers to federal AI acquisition, which emphasizes maximizing competition, favoring interoperable solutions, and prioritizing American-made AI technologies (The White House, 2025).

When evaluating options, consider factors such as the cost of procurement versus development, long-term maintenance and update requirements, dependency on a single vendor, integration with existing systems, and the solution's ability to scale with the agency's needs. Additionally, assess the availability and implications of open-source versus closed-source solutions.

*Table 20. Indicators for Buying vs. Building AI*

 <u>Indicators to Buy</u>	 <u>Indicators to Build</u>
<ul style="list-style-type: none"> <li>• There are existing solutions on the market that meet most or all the requirements.</li> <li>• Implementation needs to be quick and cost-effective.</li> <li>• Data privacy and security requirements can be met by a third-party solution.</li> <li>• The agency lacks technical expertise or resources to build the solution.</li> </ul>	<ul style="list-style-type: none"> <li>• The problem is highly-specific and no COTS solution adequately meets the requirements.</li> <li>• The agency has access to unique datasets or expertise that can be leveraged.</li> <li>• Strict data privacy and security needs require full control over data handling and model training.</li> <li>• Customization, control, or scalability is critical.</li> </ul>



### Indicators to Buy



### Indicators to Build

- The agency wants to avoid dependency on a specific model provider and maintain flexibility in future development.

Before committing to full-scale implementation, teams are advised to validate assumptions through iterative design and development. This can include building a proof of concept or acquiring one from a vendor. For a COTS solution, a proof of concept might take the form of a vendor demo, a limited trial, or a pilot agreement to explore integration potential and usability in a CMS context.

Both buying and building require clear alignment with organizational goals, rigorous requirements gathering, and proactive planning to ensure successful integration and adoption. In accordance with [OMB M-25-22 Driving Efficient Acquisition of Artificial Intelligence in Government](#), teams pursuing a COTS approach should ensure that selected product / service aligns with federal modernization priorities. They can also reference [GSA Purchasing Guidance](#) to support informed acquisition decisions.

*Step 3: Determine if the solution will enhance an existing system with AI or transition to a new system.*

Enhancing an existing system means leveraging AI to improve the functionality and efficiency of an available system without fully replacing it. Either through purchase of new AI features in COTS systems or hands-on customization, this could include adding ML for automation, AI-driven data analysis, or natural language processing to an existing system or service with its core preserved. The alternative, whether purchased or built, would be to transition users to a completely new system with AI embedded from the start.

*Table 21. Indicators for Enhancing an Existing System vs. Transitioning to a New System*

Indicators for Enhancing an <u>Existing</u> System	Indicators for Transitioning to a <u>New</u> System
<ul style="list-style-type: none"><li>• The system can be incrementally improved with AI (e.g., upgrading to a higher COTS licensing tier with AI features, enabling custom predictive analytics or automation).</li><li>• Existing workflows are functional but need optimization.</li><li>• Budget or time constraints favor enhancement over a complete transition.</li><li>• The system has an engaged user base.</li></ul>	<ul style="list-style-type: none"><li>• The existing system is outdated, inflexible, or unable to meet evolving needs.</li><li>• The problem requires new capabilities that would require a substantial rewrite of the system.</li><li>• Strategic goals require a solution that existing systems cannot support.</li></ul>

Enhancing an existing system by adding AI-driven features can streamline workflows, improve efficiency, and extend capabilities while minimizing effort and change management. However, this approach may be constrained by the system's current architecture and integration limitations. Transitioning to a new AI-enabled system offers greater flexibility and customization but requires more time, resources, and development effort. In either case, teams are encouraged to begin with a proof of concept or pilot to validate assumptions, assess integration challenges, and iterate on the solution before committing to full-scale implementation.

The rest of Chapter 5 provides generalized guidance for buying and building AI products, with greater detail on how to build custom AI.

### 5.2.4. Designing AI with the Human in Mind

Before beginning *Design and Development*, it is important for the team to establish frameworks that ensure the AI project is designed with the human in mind. This increases the likelihood of adoption, improves impact, and helps prevent unintended consequences such as user confusion or disruption to existing workflows and workforce roles. Teams can incorporate [design workshops](#) and [impact assessments](#) to guide these processes.

The CMS HCAI Matrix Guide (Figure 16) and the Algorithmic Risk and Impact Assessment Framework (Table 18) described below are examples of a design workshop and impact assessment respectively. These resources were custom-created for CMS to support teams in designing Human-Centered AI products.

#### *The CMS HCAI Matrix*

The CMS HCAI 3x3 Matrix is a practical tool designed to guide teams through the formulation and development of AI projects from a Human-Centered perspective. This framework can be used within a design workshop to prompt teams to intentionally consider human impact, human-AI interactions, human concerns, and human needs throughout the project's life (CMS AI Explorers, 2025). A full version of the tool and accompanying white paper can be found in Appendix B.

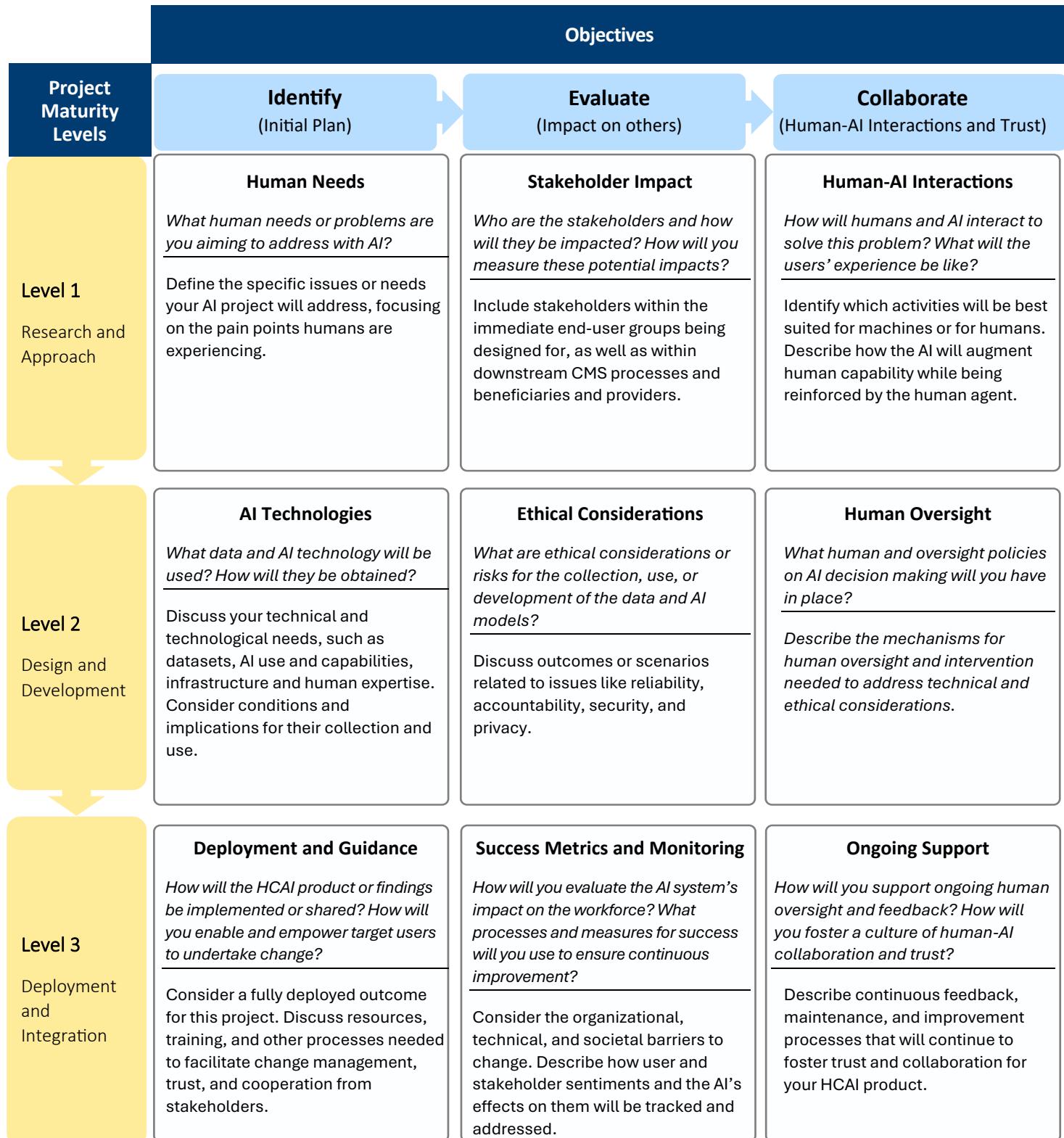


Figure 16. Human-Centered AI Matrix Guide

## CMS Artificial Intelligence Playbook

Additional resources for Human-Centered Design include the [HCD Guide by the U.S. General Services Administration](#) and the [HCD Playbook](#) for digital innovation at the CMS Center for Clinical Standards and Quality.

### *Assessing Risk and Impact*

Without proper oversight, AI systems can pose risks to both CMS and public stakeholders (NIST, 2023). Risks for the agency during an AI project can be any that contradict or counteract the guiding principles for AI at CMS (see Section 3.3). For individual stakeholders of an AI system, using HCAI principles and evaluating against CMS' AI Performance Drivers will be especially critical for assessing system risk and impact. ARIAs (introduced in Chapter 4) can provide a structured way for teams to identify, evaluate, and manage risks and potential negative impacts an AI project may pose to human stakeholders and users. There are two primary goals of ARIAs: (1) to influence design processes by embedding considerations of stakeholder risks early on and (2) to provide documentation of the algorithm for comparison to accountability expectations (Selbst, 2021).

The ARIA process depicted in the table below provides a guide for teams to conduct their own risk and impact assessments for AI projects (CMS AI Explorers, 2025). Risk management is a continuous and timely practice, ideally begun in the earliest design phases of AI use, performed throughout the AI system lifecycle, and re-evaluated with every major change influencing the AI system (NIST, 2023). Appendix B offers more information about the ARIA and a supplementary self-assessment questionnaire template for CMS use.

*Table 22. Algorithmic Risk and Impact Assessment Framework*

Activity	Description
 Scope and Discovery	<ul style="list-style-type: none"><li>Define the scope for the assessment.</li><li>Understand objectives, intended purpose, and use.</li><li>Identify data components.</li><li>Understand decision procedures.</li><li>Identify stakeholders.</li></ul>
 Preliminary Risk Identification and Mapping	<ul style="list-style-type: none"><li>Delineate stakeholder roles and their relevant expertise.</li><li>Begin a preliminary list of potential risks/impacts for each stakeholder and the suspected mechanisms that would lead to them.</li></ul>
 Stakeholder Consultation	<ul style="list-style-type: none"><li>Schedule interviews.</li><li>Prepare open-ended questions for information gathering.</li><li>Conduct interviews.</li><li>Schedule follow-ups as necessary.</li></ul>
 Impact Analysis	<ul style="list-style-type: none"><li>Expand on initial list of risks from research and consultation findings, documenting each potential impact with the type of impact, its mechanisms and impacted stakeholders.</li><li>Determine impact severity and likelihood of each risk.</li><li>Assign a corresponding risk level that captures severity and likelihood).</li></ul>
 Measuring and Evaluation	<ul style="list-style-type: none"><li>Generate evaluation criteria and metrics aligned to contextual needs and values.</li><li>Evaluate the system and update/iterate upon negative impact analysis.</li></ul>
 Risk Management	<ul style="list-style-type: none"><li>Prioritize risk mitigation based on risk level and cost/level of effort.</li><li>Create auditing criteria or guidelines to measure sufficiency and residual risk.</li><li>Implement mitigation tactics.</li><li>Reassess periodically.</li></ul>

Designing with the human in mind supports use-case design and evaluation by helping teams think critically about the implications of their AI systems from the earliest stages. By anticipating human-AI interactions, potential risks, and benefits, these processes encourage teams to make informed decisions that reflect the needs of stakeholders and the broader impact of the proposed AI solution. They offer a means to align technical innovation with responsible AI principles, ensuring the systems designed, developed, and deployed in the following stages are done with care.

### 5.2.5. Case Study Example - Research and Approach

Learn about the research the CMS Chat team conducted and how they validated the need for a custom-built Chatbot in Figure 17 below.

**Case Study: CMS Chat**

What steps were taken during the **Research and Approach** stage that prompted the creation of the CMS Chat tool?



The **Research and Approach** stage involved conducting stakeholder interviews and mapping user needs against organizational goals like efficiency and workforce development. The team embraced human-centered design principles, iteratively testing and refining the approach with users to ensure technical feasibility and organizational value alignment. This process validated that an AI-powered chatbot would genuinely enhance how employees access organizational knowledge.



Figure 17. Research and Approach for CMS Chat

## 5.3. Design and Development

After the team has conducted discovery research and selected an AI approach, the next steps in the process include designing for human-AI interactions, planning for versioning, preparing data, selecting appropriate models, and developing and testing those models.

This section provides guidance for both buying and building AI products. Taking an iterative approach to design, development, and testing enables teams to validate their assumptions early and often. Teams should first start with proofs of concept to test technical feasibility, build a pilot to explore usability and performance, and then develop an MVP that delivers core functionality and informs further refinement. This process helps ensure the final solution effectively addresses the business problem while minimizing risk.

### 5.3.1. Designing Human-AI Interactions

Regardless of whether the team is buying a COTS product or building a custom AI tool, prioritizing human needs and goals is essential when designing an AI solution that will be incorporated into a user's workflow. Teams should reference the HCAI Matrix (see Section 5.2.4), which helps identify broad human-centered considerations.

Once these foundational needs are established, a team designing an AI product can focus on human-AI interactions, a term that refers to how humans communicate and collaborate with AI-enabled systems (whether directly through user interfaces or indirectly as AI operates in the background) to inform decisions and automate processes (Interaction Design Foundation, 2025).

[Microsoft's Human-AI Experience \(HAX\) Toolkit](#) provides structured and easy-to-follow guidance for designing AI products and interfaces that align with human needs and support seamless collaboration between humans and AI.

All AI software interfaces should comply with [Section 508 of the Rehabilitation Act](#) to ensure software meets the needs of all users. The team should reference [WCAG 2.1 Guidelines](#) (Web Content Accessibility Guidelines) to implement best practices for perceivable, operable, and understandable AI-driven interfaces. As the team designs the product's interface and interactions, they should conduct usability testing on each version to gather feedback from users and make iterative changes to ensure that the product meets their needs (Moran, 2019).

### 5.3.2. Planning for Versioning

As development activities begin, it is important to establish tracking and versioning practices to ensure traceability across multiple iterations of datasets and models. Tracking refers to monitoring changes over time, including who made them and when. Versioning is the practice of saving and labeling specific states of data, code, or models so they can be reused, compared, and stored. Regardless of the selected approach (buy or build), versioning protects both COTS solutions and custom development models, which can change and mature over time.

There are five main advantages of versioning for AI projects: traceability, reproducibility, rollback, comparison for debugging, and collaboration (Intro to MLOps: Data and Model Versioning, 2023). These advantages are further explained below in Table 23.

*Table 23. Advantages of Versioning for AI Projects*

Advantage	Description	Implications if Not Considered
Traceability	Tracks who made changes, when, and the impact of each change.	Hard to audit changes or understand the root cause of issues.
Reproducibility	Enables recreation of past results using specific file or project versions.	Inability to verify or validate past project outcomes, slowing progress and reviews.
Rollback	Allows quick reversion to earlier stable versions if issues arise.	Delays in recovery and increased downtime during code failures.
Comparison and Debugging	Supports version comparison to identify and resolve issues.	Difficult to isolate bugs or assess performance differences.
Collaboration	Facilitates teamwork through branching and merging in version control.	Risk of overwriting or duplicative work, code conflicts, and reduced team efficiency.

Together, the first three advantages—traceability, reproducibility, and rollback—enable teams to track experimental iterations, align specific datasets with models, and restore previous iterations as needed. These practices help ensure results can be repeated and verified over time.

## CMS Artificial Intelligence Playbook

---

There are three common approaches to implementing version control: local, centralized, and distributed. Local version control stores all files and change history on a single computer, making it simple but limited to individual use. Centralized version control uses a shared, network-connected repository that team members access directly, supporting collaboration but relying on constant connectivity. Distributed version control builds on the centralized model, by allowing team members to work from local copies, then sync changes back to the central repository. This approach supports more flexible collaboration and offline work. Table 24 provides AI project teams with examples of tools for each of the version control types described above.

*Table 24. Common Tools for Version Control Approaches*

Version Control Type	Examples/Tools
Local	File backups, RCS (Revision Control System), SCCS (Source Code Control System)
Centralized	Apache Subversion (SVN), Microsoft Team Foundation Version Control (TFVC)
Distributed	Git, GitHub, GitLab, Bitbucket

### 5.3.3. Preparing Data

Once a team has completed initial research and determined whether to buy or build an AI solution, the next step is to prepare the data. Data preparation refers to the process of identifying, gathering, exploring, cleaning, and transforming data to ensure it is suitable for use in AI models. This step lays the foundation for all downstream AI development.

If a team chooses to buy an AI solution, data preparation may focus on ensuring internal data is compatible with the vendor's input requirements. If a team chooses to build an AI solution, data preparation often requires gathering raw data from multiple sources, performing exploratory data analysis (EDA), addressing data quality issues, and transforming data into a usable format.

#### Data Preparation Tasks

Data preparation includes a series of structured tasks that help ensure datasets are reliable, representative, and ready for modeling. Table 25 outlines common data preparation steps, how they are performed, key tools, and considerations for each.

*Table 25. Data Preparation Tasks for AI Projects*

Task	Definition/How to Perform	Relevant Tools and Resources
1. Data Identification and Collection	Identify relevant data sources (structured and unstructured) and collect data into a centralized location.	<ul style="list-style-type: none"><li>• Data catalogs</li><li>• Data integration platforms</li><li>• Stakeholder Interviews</li></ul>
2. Exploratory Data Analysis (EDA)	Perform initial exploration to understand data distributions, identify patterns, correlations, anomalies, and potential data issues. Common methods include descriptive statistics and visualization techniques.	<ul style="list-style-type: none"><li>• Python libraries (<a href="#">Pandas</a>, <a href="#">Matplotlib</a>, <a href="#">Seaborn</a>)</li><li>• R packages (<a href="#">ggplot2</a>, <a href="#">dplyr</a>)</li></ul>

## CMS Artificial Intelligence Playbook

Task	Definition/How to Perform	Relevant Tools and Resources
<b>3. Data Cleansing</b>	Identify and address quality issues such as missing, incorrect, inconsistent, or malformed data. Techniques include removing duplicates, handling missing values, and correcting errors.	<ul style="list-style-type: none"><li>OpenRefine</li><li>Python libraries (<a href="#">Pandas</a>, <a href="#">NumPy</a>)</li><li>AWS Glue DataBrew</li></ul>
<b>4. Data Transformation and Feature Engineering</b>	Transform data into formats optimized for AI models. Tasks include normalization, encoding categorical variables, and creating new features from existing data.	<ul style="list-style-type: none"><li>AWS SageMaker Data Wrangler</li><li>Python libraries (<a href="#">Pandas</a>, <a href="#">Scikit-learn</a>)</li></ul>
<b>5. Data Splitting</b>	Partition data into separate sets for training, validating, and testing the model. Common data splitting practices are 60-20-20 or 70-15-15.	<ul style="list-style-type: none"><li>Scikit-learn <code>train_test_split</code> function</li><li>R's <a href="#">caret</a> package</li></ul>

These tasks lay the groundwork for developing reliable AI models that are aligned with project goals. While presented in a logical sequence, many data preparation steps require iteration as teams gain new insights or updated data.

### Privacy, Governance, and Data Protection

In addition to complying with CMS' existing [information security and privacy](#) policies, AI project teams must be aware of sensitive data used in or collected by AI systems. Given that CMS project data may include sensitive details (e.g., processed claims, provider information, PII, PHI, and third-party data), teams need to practice careful handling and safeguarding during data preparation and throughout the AI system's life. Proper data management protects individual privacy, autonomy, identity, and dignity. Governance practices need to align with broader CMS and federal guidance to address accountability and transparency, and to ensure data-related decisions are clear and documented (see Chapter 4).

Since data is essential for both AI model training and operations, it must be protected to prevent security breaches associated with model outcomes. The AI model itself must also be protected from malicious inputs that can lead to manipulated predictions or outputs. Finally, operational security ensures measures are in place throughout the AI lifecycle to protect against vulnerabilities.

### Threat Modeling

AI projects have unique data, model, and operational security concerns, making threat modeling a key step for protection. Mapping the data flow and architecture of the AI system is the first step in creating a threat model, which is a mental model of the potential threats against a system and what needs to be mitigated to minimize the impact of any threats. While related to the ARIA, the process of threat modeling specifically targets security and privacy characteristics.

There are generally four (4) questions used to create a threat model (Braiterman, et al., 2020) which gradually become more complex:

- What are we working on?
- What could go wrong?
- What are we going to do about it?
- Did we do a good enough job?

Figure 18 below depicts these practical threat modeling questions in a cycle along with a step to “do the work”, which encompasses the *doing* of mitigation tactics (CMS AI Explorers, 2024). See Appendix B for an AI threat model template and its associated AI Explorers white paper.

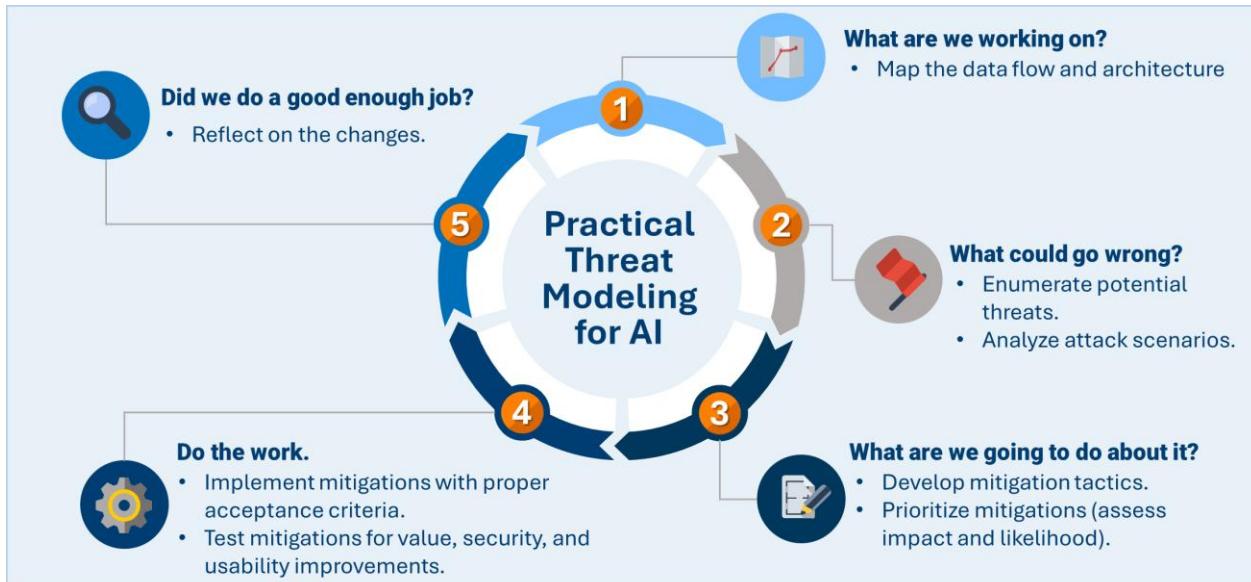


Figure 18. Practical Threat Modeling for AI

As such, the preparation of data for an AI project expands beyond typical data manipulation to require accountability in safe and responsible data handling before it is introduced to any AI system model.

### 5.3.4. Selecting the Right Models

Once the AI project team has completed data preparation, the next step is to evaluate what kind of model best fits the needs of the project. This includes identifying the type of model (e.g., classification, regression, clustering, LLM) and the approach for obtaining it (see Section 5.2.3).

#### *Model Selection Decision Guide*

The table below outlines key considerations to help AI project teams determine what kind of model might be appropriate and whether to acquire an existing model or develop one.

Table 26. Model Selection Decision Guide

Consideration	Recommendation
Does your team need a plug-and-play solution with minimal development effort?	Consider a COTS product with built-in AI/ML capabilities. (Some COTS products allow model selection or fine-tuning; others do not. Review vendor documentation carefully.)
Do you have access to domain-specific data that can improve outcomes?	Consider building or fine-tuning an existing open-source model to tailor performance.
Is the problem structured (e.g., predicting values, classifying inputs)?	Traditional ML models (e.g., regression, random forest, SVM) might suffice.

## CMS Artificial Intelligence Playbook

Consideration	Recommendation
Does the task involve understanding or generating natural language?	Consider a LLM. Use Retrieval-Augmented Generation (RAG) when outputs need to reference source documents.
Are there constraints around cost, compute, or privacy?	Evaluate model size, efficiency, and on-premise deployment options (especially important for LLMs).

### Evaluating and Selecting Large Language Models

If your team determines that an LLM is the best fit, further evaluation is needed to identify the most appropriate one. Selecting the right LLM involves balancing tradeoffs across multiple criteria:

- **Aligning the LLM capabilities** with user needs, stakeholder priorities and the business problem.
- **Experimentation** to assess model versatility.
- **Flexibility** to adapt to evolving models, allowing seamless updates to maintain effectiveness.
- **Cost** to acquire and run the model.
- **AI Performance Drivers**, i.e., security, privacy, transparency, reliability, and accountability.

A helpful tool for comparing LLMs is the LLM Evaluation Label, developed for LLMs currently available for use within CMS' AI Workspace, a cloud-based lab environment that enables teams to experiment with AI models instantly (CMS AI Explorers, 2025). These labels summarize each model across key metrics, as shown in Figure 19. A scoring template and methodology for creating these LLM evaluation labels can be found in Appendix B.

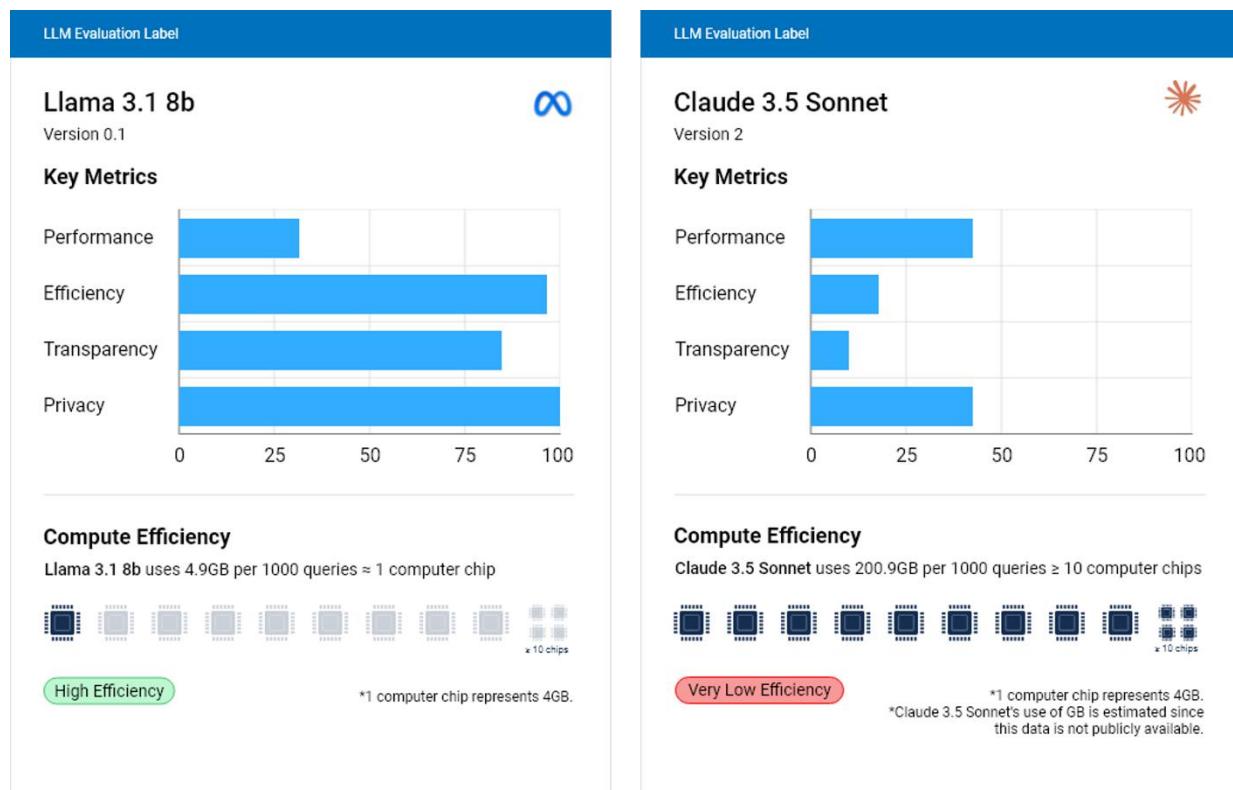


Figure 19. Evaluation Labels for Llama 3.1 8b and Claude 3.5 Sonnet

## CMS Artificial Intelligence Playbook

---

Key considerations included in Table 27, provide guidance on how to tailor LLM implementation to specific AI project needs.

Table 27. Key Considerations for LLM Implementation

Description	Implications
<b>Identify the hardware available to you.</b>	The amount of graphics memory (Video Random Access Memory [VRAM]) on your device affects which AI models you can run. Less VRAM limits the project to smaller, simpler language models.
<b>Consider whether the task is time sensitive, accuracy dependent, or cost-conscious.</b>	Understanding this helps balance the trade-offs between performance, cost and speed, tailoring the LLM deployment to your priorities.
<b>Consider Retrieval-Augmented Generation (RAG) if the LLM's responses need to come from document sources.</b>	RAG enables LLMs to pull accurate information from specific documents, reducing hallucination risk.
<b>Make benchmarks relevant to the specific domain to assess real-world performance.</b>	Domain-specific benchmarks can help predict real-world effectiveness and guide further tuning.
<b>Prioritize metrics based on the application's goals.</b>	For example, emphasize groundedness over creativity in sensitive contexts (e.g., healthcare, policy interpretation).

This checklist is adapted from the *Proposed LLM Implementation Checklist* and its associated *LLM Cost and Quality Comparisons* white papers (CMS AI Explorers, 2025) found in Appendix B.

### 5.3.5. Developing and Testing Models

After selecting the appropriate model, the next step is developing and testing the model to ensure robustness, reliability, and suitability for production. This phase is critical for refining model performance, ensuring reproducibility, and validating the model's ability in real-world scenarios. To guide this process, teams can leverage structured frameworks such as [CRISP ML\(Q\)](#) and Machine Learning Technology Readiness Level ([MLTRL](#)), which provide best practices for model development and testing.

The following table outlines the key activities within model development and model testing.

Table 28. Model Development and Testing

Phase	Purpose	Key Activities
<b>Model Development</b>	Refines and optimizes the selected model through an iterative, experiment-driven approach in which AI project teams train, validate, and test to ensure alignment with business objectives and performance requirements.	<ul style="list-style-type: none"><li>• Train, refine, and optimize models iteratively.</li><li>• Align model performance with business objectives.</li><li>• Use hyperparameter tuning (Grid Search, Randomized Search, Bayesian Optimization).</li><li>• Apply cross-validation techniques (K-fold, Stratified K-fold).</li><li>• Implement techniques like regularization and early stopping to prevent overfitting.</li></ul>

Phase	Purpose	Key Activities
<b>Model Testing</b>	Confirms that the model performs reliably in real-world applications and maintains robustness across diverse scenarios. In some cases, independent evaluation should be employed as outlined in OMB memo M-25-22 (U.S. Office of Management and Budget (OMB), 2025).	<ul style="list-style-type: none"> <li>Validate performance on unseen, diverse data.</li> <li>Assess generalization across different conditions.</li> <li>Perform stress testing and adversarial testing in controlled sandbox environments.</li> <li>Evaluate using standard metrics (accuracy, precision, recall, F1-score, specificity, and Receiver Operating Characteristic/Area Under Curve [ROC/AUC]) and advanced metrics (coherence, groundedness, context relevance, answer relevance, and bias)</li> </ul>
<b>Reproducibility (Cross-Cutting Process)</b>	Enables repeatability, enhances transparency, and ensures models can be reliably compared and improved over time.	<ul style="list-style-type: none"> <li>Document experiments and results (e.g., model cards, metadata records).</li> <li>Version control all models and datasets (see Section 5.3.2).</li> <li>Standardize testing and deployment environments using containerization (e.g., Docker) and orchestration tools (e.g., Kubernetes).</li> <li>Ensure consistency in training and inference environments.</li> </ul>

As models are being developed and tested, part of the iterative process includes ensuring the AI tool meets users' needs. The team can do this by conducting usability testing, which identifies pain points, improves user experience, and refines system behavior before deployment. This ensures the AI solution is human-centered and aligned with user expectations, ultimately increasing adoption and trust (Introduction to Remote Moderated Usability Testing, 2018).

### 5.3.6. Case Study Example - Design and Development

Learn more about what the CMS Chat team did during the Design and Development stage in Figure 20 below.

**Case Study: CMS Chat**

What steps took place during the **Design and Development** stage of the CMS Chat tool?



Within the **Design and Development** stage, the team conducted model evaluation, testing multiple LLM options including Llama 2 variants and Claude via Amazon Bedrock. Testing covered both performance metrics like response time and latency, as well as quality measures such as response accuracy and relevance. The team also evaluated various embedding models and vector database configurations to optimize the RAG implementation. After rigorous testing, the team selected the architecture that demonstrated the best balance of performance and accuracy.

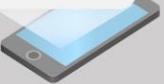


Figure 20. Design and Development for CMS Chat

## 5.4. Deployment and Integration

Deployment and integration, the final stage of the AI project lifecycle, mark the transition from a pilot to an MVP, and potentially to a fully scaled solution used in real-world environments. An MVP is a production-ready but narrowly scoped version of the product, designed to validate core functionality, gather user feedback, and guide future development. Not all pilots progress to the deployment and integration stage, and those that do may not require every aspect of the process that follows.

Deployment and integration apply to both COTS products and custom-built AI solutions, each requiring careful planning to ensure feasibility, desirability, and viability at scale. At this stage, the AI project team configures and launches the AI product and establishes continuous monitoring and performance management. They also foster trust among users and stakeholders by providing transparent communication around model limitations and potential biases (Arsanjani 2023). This chapter will outline deployment, acceptance, adoption, and scaling for teams ready to move from delivering a pilot to developing a large-scale AI product with greater impact.

### 5.4.1. Deploying an AI Product

After an AI product has been developed and tested, it can begin its transition to production. Deploying an AI product means ensuring the solution is production ready. This includes infrastructure provisioning, security, governance, and risk management to enable smooth rollout and reliable operations. Table 29 provides actions that a team should complete before deployment. In addition to these, teams should consult the [CMS Production Readiness Checklist](#), which outlines mandatory steps and artifacts required for applications before moving to production in CMS Hybrid Cloud. This checklist reinforces key readiness areas such as logging configuration, disaster recovery, incident response planning, and security verification.

*Table 29. Deploying Your AI System*

Deployment Step	Description	Key Actions
<b>Team and Stakeholder Alignment</b>	Cross-functional collaboration between technical, compliance, and operational teams supports smooth integration and long-term adoption.	<ul style="list-style-type: none"> <li>Define clear roles and responsibilities across engineering, legal, and operational teams.</li> <li>Set shared goals and performance metrics to align efforts across teams.</li> <li>Plan for ongoing communication to address concerns</li> </ul>
<b>Launch a Pilot</b>	Allows the team to gather feedback, refine the solution, and gain buy-in before scaling.	<ul style="list-style-type: none"> <li>Gather feedback from a variety of users</li> <li>Ensure the pilot represents real-world use cases.</li> <li>Refine the system.</li> <li>Build stakeholder engagement early.</li> </ul>
<b>Finalize MVP</b>	Ensure that the MVP is production-ready, meets success criteria, and incorporates feedback from earlier stage pilots.	<ul style="list-style-type: none"> <li>Confirm core functionality is stable and aligned with user needs.</li> <li>Validate that feedback from pilots and testing has been integrated.</li> <li>Complete security, compliance, and performance readiness</li> </ul>

## CMS Artificial Intelligence Playbook

Deployment Step	Description	Key Actions
<b>Technical Considerations</b>	Ensure the AI system is built on infrastructure with appropriate hardware, software, and monitoring tools to support scalability, performance, and maintainability.	<ul style="list-style-type: none"> <li>Acquire computing resources (e.g., Graphics Processing Units [GPUs], cloud infrastructure) suited to performance needs.</li> <li>Implement version control (e.g., Git) for rollback capabilities and workflow stability.</li> <li>Establish performance monitoring to detect bottlenecks and optimize efficiency.</li> <li>Use observability frameworks to track system behavior and anticipate issues.</li> </ul>
<b>Compliance, Governance, and Risk Management</b>	As outlined in OMB memo M-25-21, ensure the AI system adheres to regulatory requirements through security measures, legal oversight, and transparent documentation (U.S. Office of Management and Budget (OMB), 2025).	<ul style="list-style-type: none"> <li>Implement data security measures such as encryption and role-based access controls.</li> <li>Collaborate with legal and privacy experts to ensure compliance with CMS regulations.</li> <li>Maintain clear documentation for audits, regulatory reviews, and accountability.</li> <li>Embed compliance checks early in the deployment process to reduce risks.</li> </ul>
<b>Use ARIA for Risk Management</b>	An ARIA helps identify, evaluate, and mitigate risks such as bias, system failures, and emerging vulnerabilities (see Section 5.2.4). In some cases, independent evaluations should be employed as outlined in OMB memo M-25-22 for strict governance purposes (U.S. Office of Management and Budget (OMB), 2025).	<ul style="list-style-type: none"> <li>Conduct periodic reassessments triggered by model updates, data shifts, or regulation changes.</li> <li>Document and communicate residual risks, ensuring transparency for decision-makers.</li> <li>Identify biases or system errors early in deployment to prevent harmful outcomes.</li> <li>Integrate ARIA findings into governance frameworks for ongoing risk management.</li> </ul>
<b>Performance Monitoring, Observability, and Governance</b>	Continuous monitoring ensures AI system reliability and compliance throughout its lifecycle.	<ul style="list-style-type: none"> <li>Develop specific metrics tailored to use cases.</li> <li>Establish automated monitoring for detecting performance issues, drift, or bias.</li> <li>Use observability tools to provide insights into system health and AI decision-making.</li> <li>Align governance policies (see Chapter 4) with ARIA findings to refine auditing and compliance procedures.</li> <li>Maintain iterative improvements based on system observations and regulatory updates.</li> </ul>

Successfully deploying an AI product ensures the system is technically ready and compliant for real-world use, but technical readiness alone is not enough. To achieve long-term success, teams must also focus on building user trust and integrating the product into daily workflows through methods that promote acceptance and adoption.

### 5.4.2. Acceptance and Adoption

While often used interchangeably, *acceptance* and *adoption* of a technology tool are distinct but related concepts. The Technology Acceptance Model (TAM) defines *acceptance* as an individual's perception that a technology tool is both useful and easy to use (Davis, 1989). It lays the psychological foundation for initial engagement but does not guarantee ongoing use. *Adoption* refers to the sustained use of a technology tool within workflows. Adoption can occur without acceptance (e.g., when use is mandated), and acceptance can occur without adoption (e.g., when the tool is liked but not usable due to barriers).

These dynamics apply to both COTS products and custom-built AI, as both require users to accept and integrate the tool into their daily work.

Iterative practices such as proofs of concept and pilots support both acceptance and adoption by allowing users to engage with the technology in a low-risk setting and provide feedback before deployment.

To support both outcomes, the team should apply tactics that build trust and understanding as well as those that promote integration and sustained use. Table 30 provides the key actions a team can take to promote the acceptance of an AI tool with users.

*Table 30. AI Technology Acceptance*

User Acceptance Step	Description	AI Project Team Key Actions
Awareness	Employees become aware that a new AI tool exists.	<ul style="list-style-type: none"> <li>• Communicate clearly and early about the new AI tool.</li> <li>• Use email, internal forums, and leadership messaging.</li> </ul>
Understanding	Users learn what the AI tool does, why it is being introduced, and how it fits into their work.	<ul style="list-style-type: none"> <li>• Offer brief overviews, FAQs, and quick-start guides.</li> <li>• Explain how the AI helps solve relevant problems</li> </ul>
Perceived Usefulness	Users believe the AI tool will improve or support their work outcomes.	<ul style="list-style-type: none"> <li>• Highlight use cases, testimonials, or pilot success stories.</li> <li>• Show how the tool enhances productivity or reduces workload.</li> </ul>
Perceived Ease of Use	Users believe the AI tool is easy to use and can be integrated into their tasks with minimal effort.	<ul style="list-style-type: none"> <li>• Provide simple user interfaces and user-friendly training materials.</li> <li>• Include guided walk-throughs or demos.</li> </ul>
Reliability and Transparency	Users feel confident that the AI behaves predictably and is aligned with governance standards.	<ul style="list-style-type: none"> <li>• Share information on how the model works, its limitations, and how results are monitored</li> <li>• Provide access to ARIA summaries or governance documentation.</li> </ul>
Willingness to Try	Users are open to using the AI tool and taking the first step toward engagement.	<ul style="list-style-type: none"> <li>• Create opt-in pilot programs or low-risk opportunities to test the product.</li> <li>• Offer support for first time use.</li> </ul>

Table 31 provides the key actions a team can take to promote the adoption of an AI product with users.

*Table 31. AI Technology Adoption*

User Adoption Step	Description	AI Project Team Key Actions
<b>Onboarding and Training</b>	Users receive hands-on guidance to learn how to use the AI system effectively.	<ul style="list-style-type: none"> <li>Deliver role-specific training sessions and materials.</li> <li>Establish ongoing support systems (e.g., help desks, AI advisors, dedicated liaisons).</li> </ul>
<b>Workflow Integration</b>	The AI tool is embedded into daily tasks and existing systems.	<ul style="list-style-type: none"> <li>Automate or simplify integration steps.</li> <li>Coordinate with IT to integrate the AI product with existing systems.</li> </ul>
<b>Organizational Support</b>	Leadership, IT, and operations provide consistent support and resources for use	<ul style="list-style-type: none"> <li>Allocate time and technical resources for support</li> <li>Set clear expectations for use and reinforce the value of the tool through leadership</li> </ul>
<b>Feedback and Iteration</b>	User feedback is collected to refine the tool and improve alignment with work needs.	<ul style="list-style-type: none"> <li>Use surveys, interviews, and usage metrics to gather feedback.</li> <li>Continuously improve the tool based on insights.</li> </ul>

Establishing user acceptance and promoting adoption are essential steps in ensuring that an AI product becomes a trusted and integrated part of daily work. With these foundations in place, the next phase focuses on preparing the product and the organization for broader scaling and sustained impact.

### 5.4.3. Scaling

Scaling is the final step when deploying an AI product. Scaling is the process of expanding an AI system's capacity to handle increased usage, integrate seamlessly into existing workflows, and adapt to evolving organizational needs, while maintaining performance. Whether the AI solution is a COTS product or a custom-built AI product, scaling requires careful planning to ensure it continues to deliver value at a larger scope.

Table 32 details multiple aspects of the scaling process with key actions to take at each step. A team should use this information to assess if the AI product is prepared for scaling.

*Table 32. Scaling an AI Product*

Scaling Step	Description	Key Actions
<b>Assessing Readiness for Scaling</b>	Ensures the AI system is stable, well-adopted and capable of handling higher demand without performance degradation.	<ul style="list-style-type: none"> <li>Gauge stability by assessing uptime, performance consistency, and frequency of incidents. If these metrics are within accepted ranges, then scaling is possible.</li> <li>Consider whether the current system can handle increased load, both technically and operationally.</li> <li>Gather user feedback to track adoption rates, user satisfaction, and common pain points.</li> <li>Evaluate ROI and confirm financial or operational benefits.</li> </ul>

Scaling Step	Description	Key Actions
Technical Considerations	Addresses potential infrastructure bottlenecks and ensures that models and systems can scale effectively.	<ul style="list-style-type: none"> <li>Review infrastructure by assessing cloud/hardware resources, like GPU availability.</li> <li>Conduct load testing to evaluate how models perform under peak usage.</li> <li>Ensure AI accelerator scalability by enabling dynamic resource allocation and efficient routing to new nodes.</li> <li>Redesign system architecture if original design was not built for scalability.</li> </ul>
Data Management Practices	Ensures capacity for growing data volume and variety while automating ingestion processes.	<ul style="list-style-type: none"> <li>Scale storage solutions using techniques like database sharding, redundancy, and distributed instances.</li> <li>Automate data ingestion to handle larger volumes while maintaining quality and integrity.</li> <li>Implement data governance policies to manage access levels for different user groups securely.</li> <li>Monitor system capacity to prevent performance bottlenecks from increased read/write operations.</li> </ul>
Monitoring the System	Shifts from basic performance metrics to proactive anomaly detection, preserving system trust and uptime.	<ul style="list-style-type: none"> <li>Expand monitoring scope to track broader system performance, usage trends, and potential failures.</li> <li>Enhance dashboards and alerts to detect anomalies like data drift, latency issues, or load imbalances.</li> <li>Proactively address risks by setting up automated responses to prevent disruptions.</li> <li>Ensure visibility across technical and user metrics to quickly identify and resolve issues at scale.</li> </ul>

After scaling an AI product, maintaining its reliability involves regular performance monitoring, updating models and data pipelines, addressing technical debt, and ensuring compliance with evolving governance standards. These efforts help sustain the product's value and functionality over time. At the same time, staying informed about new AI developments is essential. The next chapter explores emerging AI technologies and offers guidance on how organizations can prepare to anticipate and adapt to future innovations.

#### 5.4.4. Case Study Example - Deployment and Integration

CMS Chat has been successfully deployed within the agency; learn more about what the team did and how it can be accessed in Figure 21 below.

### Case Study: CMS Chat

Finally, what's the status of the CMS Chat tool within the **Deployment and Integration** stage?

As part of the **Deployment and Integration** stage, CMS Chat launched in December 2024 with a measured rollout strategy which includes robust observability tools to monitor usage patterns and costs. All CMS employees currently have access to CMS chat and currently there is a waitlist for contractors to gain access.

That's great! How can I use CMS chat?

CMS federal employees can access CMS Chat by going to <https://chat.cms.gov>. Guidance for effective prompt crafting is available in Appendix C.6. of this Playbook. Our adoption approach includes demonstrations, AI Ignite training sessions, and a dedicated Slack channel for user feedback and support.



Figure 21. Deployment and Integration for CMS Chat

### Key Takeaways - Conducting an AI Project

Chapter 5 stepped through various stages and approaches that teams at CMS may follow while designing and implementing AI solutions. Despite the emphasis on AI project team activities, all primary audiences of the Playbook contribute throughout these stages of conducting an AI project.



- Leadership and managers will need to form interdisciplinary AI project teams that fulfill necessary skills and expertise, with clearly assigned roles and access to relevant stakeholders.
- Plan accordingly for AI projects to follow three key stages: Research and Approach, Design and Development, and Deployment and Integration. Note that each stage requires appropriate resources and oversight.
- Effective adoption and scaling of deployed AI systems depends on building trust with stakeholders, providing adequate training and support, and measuring success through both quantitative and qualitative metrics.



### KEY TAKEAWAYS

From Chapter 5

For AI Project Teams

- AI project teams should conduct thorough discovery research to define the business problem, establish requirements, and determine if AI is the appropriate solution before committing resources to development.
- Research and iterative testing, such as through development of concept proofs, will inform whether the project team should buy a COTS solution or build custom AI based on factors such as cost, technical expertise, data privacy requirements, and long-term maintenance needs.
- Human-centered AI and performance driver principles must guide the entire project lifecycle, from initial design through deployment and continuous maintenance, ensuring systems are both secure and designed with stakeholder needs in mind.



### KEY TAKEAWAYS

From Chapter 5

For IT and Security Teams

- IT and security teams must support the infrastructure development and security considerations to be integrated throughout a project lifecycle, from data preparation through deployment and scaling.
- Teams will need to implement robust version control, threat modeling, and security measures to protect sensitive data and ensure system integrity.
- Continuous monitoring and observability frameworks are essential for maintaining system reliability, detecting issues, and ensuring compliance with security requirements.

The structured approach to AI project implementation outlined in this chapter provides teams with the foundation needed for successful AI adoption. The final chapter of the Playbook builds on this foundation by examining future technologies and considerations that will shape CMS' ongoing journey toward AI-enabled healthcare transformation.

## 6. Looking Ahead

As AI continues to reshape healthcare, CMS will need to anticipate new technologies to remain an innovative leader. This chapter explains how CMS is moving from basic predictive tools to smarter AI systems that help staff, handle tasks on their own, and adjust in real time. It also covers the social, regulatory, and security reasons why using AI responsibly is so important. By examining near-term as well as more distant horizons, CMS can maximize benefits while safeguarding public trust.

### 6.1. A Glimpse into Future Technologies

Within CMS, there is a growing evolution from purely predictive tools to assistants that augment employees' everyday work to autonomous agents. The agency is moving from using narrow one-off solutions to general-purpose tools that can resolve tasks proactively, integrate multiple data types, and adapt to evolving user demands. These tools come with their own challenges. New social, regulatory, and security considerations are expanding as LLM and agentic AI use become more ubiquitous that simultaneously reign in their impact while driving better technology evolution. As AI models improve, their efficacy in implementation requires compounding evolution in coupled factors such as adoption, policy, system interoperability and more (Rutherford C. , 2024). This is especially true in sensitive domains. Despite these challenges, the strong momentum currently in play is an opportunity for CMS to evaluate the feasibility of cutting-edge technology while learning from experience and using those lessons for strategic planning.

Not all technologies are equally ready for CMS use. Technologies like assistants are being planned for near-term implementations, but agentic AI enabled systems (which blend prediction with the capacity to orchestrate complex, multipart processes and tools) remain outside feasibility for most use cases now. These emerging capabilities (alongside others such as advanced wearables offering preventative care capabilities), currently offer a glimpse of what the healthcare ecosystem could look like when technology is integrated into every level of patient engagement, payment models, and policy administration. The timeline in Figure 22 shows estimates of when each technology will be implemented (or directly impactful) within CMS based on research within industry, federal guidelines, and government (U.S. Office of Management and Budget (OMB), 2025) (U.S. Office of Management and Budget (OMB), 2025) (GSA Blog, 2025) (GAO, 2024) (GAO, 2023) (GAO, 2023).

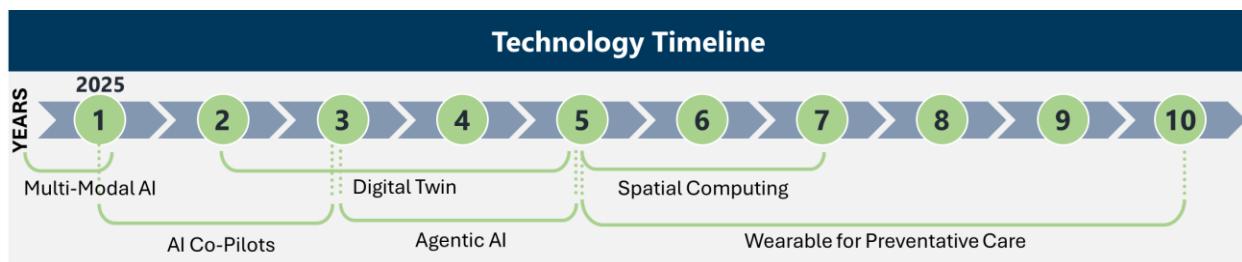


Figure 22. Technology Timeline

The next two sections provide additional insights into each of the technologies mentioned.

### 6.1.1. Near Future Technology

Advancements such as assistants, agentic AI, multimodal AI, and digital twins are becoming more commonplace and offer near-term practicality for CMS.

#### *Integrated Assistants*

AI assistants are embedded in software or workflows to support users by suggesting content, automating tasks, and enhancing productivity without replacing human judgment. Table 33 below further discusses assistants:

*Table 33. Integrated Assistants*

Technology	Details and Examples
<b>How it works</b>	Assistants typically LLMs tailored to the domain, responding to natural language commands or context, and pulling from relevant data to help with decisions or document creation.
<b>General example</b>	Microsoft's Dragon Copilot for clinicians can automatically draft patient visit summaries and referral letters, reducing administrative burden on doctors (Microsoft).
<b>CMS context</b>	In program oversight, AI assistants could help staff navigate complex policies and data. For instance, assistants driven by LLMs can parse evolving healthcare payment rules and highlight billing edits.

#### *Agentic AI*

Autonomous AI agents adapt to new information, make decisions, and proactively perform tasks with minimal human input to achieve defined goals. Table 34 below further discusses agentic AI:

*Table 34. Agentic AI*

Technology	Details and Examples
<b>How it works</b>	Agentic AI systems often link LLMs with process automation, allowing them to not only generate insights but also trigger actions.
<b>General example</b>	In customer service, agentic AI "virtual agents" can schedule appointments, check insurance benefits, or refill prescriptions without involving a live operator.
<b>CMS context</b>	Agentic AI could automate claims handling. Early implementations include AI-driven fraud detection systems that independently scan billing patterns and notify investigators if they detect anomalies in real time.

\*An *Understanding Agentic AI and Its Potential at CMS* white paper by the AI Explorers is linked in Appendix C.

#### *Integrated Multimodal AI*

Integrated multimodal AI combines multiple types of data—such as text, images, and audio—with a single model or workflow to enable richer, more comprehensive analysis. By correlating information across formats, these systems generate more complete insights than single-modality approaches. Table 35 below further discusses integrated multimodal AI:

*Table 35. Integrated Multimodal AI*

Technology	Details and Examples
<b>How it works</b>	Multimodal AI leverages “fusion” models or combined pipelines to handle various inputs (clinical notes, labs, medical images, voice transcripts).
<b>General example</b>	Clinical AI tools can answer complex queries by analyzing both text and imaging data. By merging multiple data sources, it improves diagnostic accuracy and situational awareness.
<b>CMS context</b>	Multimodal AI can synthesize unstructured and structured data for better Medicare or Medicaid oversight. For instance, it might analyze claims data, beneficiary complaints, and call center audio transcripts to detect issues.

### Digital Twins

A digital twin is a virtual replica of a real-world entity or process that uses real data to mirror physical behavior and respond to simulations. It can represent facilities, populations, or individuals to test and analyze scenarios in a virtual environment. Table 36 below further discusses digital twins:

*Table 36. Digital Twins*

Technology	Details and Examples
<b>How it works</b>	Digital twins use real-time data from sources like Internet of Things (IoT) sensors and Electronic Medical Records (EMRs) to keep their virtual models up to date. This creates a dynamic simulation that supports performance testing and strategic forecasting.
<b>General example</b>	Hospitals may use digital twins to predict bed shortages or optimize workflows, proactively addressing operational bottlenecks.
<b>CMS context</b>	In oversight for Medicare or Medicaid, a digital twin of the claims process could test changes to reimbursement policies or prior authorization rules in simulation, revealing impacts on costs and service quality before real-world implementation.

Collectively, these innovations present an immediate opportunity for CMS to leverage cutting-edge technologies to streamline services and inform future policy.

### 6.1.2. Distant Future Technology

The above technologies are all near-term and have practical applications being developed. This section focuses on technologies that are in their infancy. Table 37 below introduces each technology expected to impact CMS in the long term:

*Table 37. Distant Future Technologies*

Technology	Description
<b>Fully Integrated Wearables for Care</b>	Wearables are rapidly evolving from fitness trackers into continuous health monitors capable of detecting early signs of disease. In the future, they may integrate advanced biosensing and AI to offer real-time insights that anticipate medical issues before symptoms appear while offering support for care management. This has potential to transform long-term care and reduce costs (Ferguson, et al., 2022).

Technology	Description
Spatial Computing	Spatial computing merges the digital and physical worlds, allowing users to interact with data through gestures, voice, and movement. As devices like AR glasses become more powerful, spatial computing may reshape healthcare training, diagnostics, and remote care delivery. This will fundamentally shift how Medicare and Medicaid coverage of care work. However, widespread adoption will depend on overcoming significant hardware and access barriers (Dickson, 2024).
Quantum Computing	Quantum computing uses the principles of quantum mechanics to solve complex problems far faster than today's computers. Though still in early development, it may eventually enable breakthroughs in drug discovery, predictive health analytics, and even CMS program integrity by processing massive datasets in parallel (Reymond, 2025).
Artificial General Intelligence (AGI) & Artificial Superintelligence (ASI)	AGI would possess human-like reasoning across any domain, while ASI would far surpass it, potentially reshaping science, governance, and healthcare. While still speculative, these technologies could revolutionize CMS oversight if developed—but also pose major risks around safety, fairness, and control (Altman, 2023).

## 6.2. Organizational Preparation

Building on the AI principles in Section 3.3 and the emerging technologies described in Section 6.1, organizations must align their administrative structures, policies, resources, and workforce to effectively harness AI's evolving capabilities. A well-prepared organization understands that technology alone does not guarantee transformation; rather, success hinges on the interplay of governance frameworks, people, change management, and collaboration. This section outlines how organizations can proactively adapt in anticipation of assistants, agentic AI, multimodal AI, and other rapidly advancing technologies.

### 6.2.1. Evolving Policies and Governance

#### *Updating Policies for New AI Capabilities*

The rise of agentic and multimodal AI calls for continuous policy review and updates. While this playbook offers governance pathways (see Chapter 4) that focus on balancing risk and opportunity, emerging technologies may introduce novel risk profiles, such as autonomous decision-making in claims processing or merging text and image data for more comprehensive analyses.

#### 1. Flexible Governance

- **Supportive Review Over Restrictive Regulation:** Rather than imposing rigid rules on every AI application, aim for governance models that encourage experimentation while mitigating risks through structured checkpoints (e.g., an AI review board or AI wardens) (Rutherford C. D., 2024).
- **Scenario-Based Policies:** Develop policies that address specific use cases, such as agentic systems capable of initiating tasks independently. This approach helps tailor oversight to the unique demands of each technology.

#### 2. Ensuring Fairness and Transparency

- **ARIAs:** Integrate ARIAs early (see Section 5.2.4) to identify bias and ethical concerns, especially for AI systems that shape beneficiary outcomes.

- **Clear Communication:** Publicly communicate model purposes, limitations, and oversight mechanisms to build trust among stakeholders, especially when AI impacts high-stakes decisions (e.g., coverage determinations).

### 6.2.2. Continuing Workforce Transformation

#### *Identifying New Roles*

Advancements in AI often demand specialized positions—such as AI ethics leads, data curation specialists, ontologists, and “futurists” who anticipate long-term AI trends. These roles complement existing teams of product managers, data scientists, developers, and domain experts, ensuring a holistic approach to AI development and deployment.

#### 1. Upskilling Current Teams

- **Targeted Training Programs:** Build on CMS’ existing Workforce Resilience initiatives by offering additional practical, scenario-based modules covering advanced data engineering, AI threat modeling, interpretability tools, human-in-the-loop integration, assistant and agentic development, and more.
- **Cross-Functional Literacy:** Encourage domain experts, policy analysts, and IT/security staff to gain foundational AI knowledge, enabling better project collaboration and informed decision-making.

#### 2. Cultivating a Learning Culture

- **Peer-Led Groups:** Support forums or communities of practice (like the AI Community and CMS Chat Slack channels) to exchange best practices and innovations.
- **Mentorship and Knowledge Sharing:** Pair junior data analysts with experienced AI professionals or AI Explorers to accelerate learning and foster cross-pollination of ideas.

#### 3. Hiring Specialized Talent

- **Strategic Recruitment Initiatives:** Develop tailored recruitment campaigns by partnering with top academic institutions, industry conferences, and specialized job boards to attract candidates with expertise in AI ethics, data curation, and long-term technology forecasting. Emphasize the organization’s commitment to pioneering innovative AI practices.
- **Competitive Onboarding and Growth Pathways:** Create robust onboarding programs and clear career trajectories for new hires. Offer mentorship opportunities, continuous professional development, and the chance to contribute to groundbreaking AI projects. This will ensure that top talent is not only attracted but retained within the organization.

### 6.2.3. Continuing Interdisciplinary Collaboration

#### *Internal Collaboration*

Siloed teams can slow innovation or create inconsistent standards. As AI grows more complex—combining data, algorithms, and human-centered design—effective collaboration becomes essential:

- **Cross-Team Partnerships:** Form dedicated working groups involving product managers, data scientists, HCD researchers, security analysts, and other subject matter experts. These groups can ensure alignment with AI principles (see Section 3.3) and maintain open communication channels.

- **Shared Governance Tools:** Maintain a central repository (e.g., a registry or dashboard; see Section 4.3) for all AI projects, allowing each office or component to stay informed about ongoing initiatives. A working group within the AI CCI is actively collecting use cases for this purpose.

### External Partnerships

Engagement with vendors and research institutions expands CMS' access to emerging AI technologies. However, external collaborations must align with CMS' mission and guiding principles. Establish clear guidelines around data sharing, intellectual property, and compliance requirements to mitigate risks while capitalizing on external expertise.

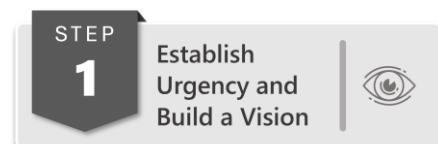
### 6.2.4. Managing Organizational Change for AI Adoption

Adopting AI at scale is not just a technical endeavor—it requires guiding people through change. The urgency is clear: *in the 2023 AI Index Report* from Stanford's Institute for Human-Centered Artificial Intelligence, 72% of surveyed global executives reported significant concerns about their organization's ability to implement AI responsibly at scale (Stanford Institute for Human-Centered Artificial Intelligence, 2023). To bridge this gap, leaders need a structured approach to change management.

This Playbook recommends a three-stage approach:



To accomplish these stages, this section employs evidence-based approaches found in the following frameworks: [Kotter's 8-Step Process](#), the [Prosci ADKAR Model](#), the [Technology Acceptance Model \(TAM\)](#), and the [Diffusion of Innovation](#) theory. By using these frameworks as foundations to build upon principles of explainability, trust, and human-in-the-loop AI collaboration, organizations can progress in their maturity from exploring AI to sustainable integration and innovation. The following subsections dive deeper into each stage.



(Kotter Steps 1–3)

John Kotter's model for leading change begins with *creating a sense of urgency* about the need for change. This means clearly articulating *why* AI is critical to the organization's future. Framing AI as essential for staying ahead (whether to meet customer expectations or improve efficiency) helps generate buy-in from staff and the public.

Kotter's next steps emphasize *forming a guiding coalition* and *developing a compelling vision*. Leaders should build a cross-functional coalition of AI advocates (e.g. CMS initiatives such as AI Explorers and the AI CCI) who can lead by example. This coalition is responsible for defining a shared, concrete vision for AI adoption that aligns with CMS' goals and values. A strong vision answers essential questions like, "What will AI improve?" and "How will it augment our services or workflows?" Vision and strategy should position AI as a mission-critical transformation, not just an IT project (Goswami, 2025).

Equally important is laying the groundwork at the individual level. The ADKAR model (Awareness, Desire, Knowledge, Ability, Reinforcement) emphasizes that “*organizations don’t change, people do*,” (Brusati, 2025). Early stages should focus on building Awareness and Desire by making the need for AI specifically relevant (e.g. showing how it addresses pain points or secures the organization’s future.) When employees understand how AI benefits them and the organization, they are less likely to resist the change. Research confirms that human factors like trust, communication, and training are often more important than the tech itself in successful AI efforts (Creasey, 2025). Leadership plays a crucial role in this process by advocating a clear vision while listening empathetically to concerns. A strong foundation is laid when urgency is matched with a unifying vision that motivates both the organization and its people toward shared success.

These early steps in creating urgency and forming a guiding coalition exemplify the Organizational AI Enablement principle. By aligning leadership, workforce readiness, and governance early in the change process, organizations lay the foundation for AI to thrive as part of their daily operations and strategic initiatives.



(Kotter Steps 4–5, ADKAR & TAM)

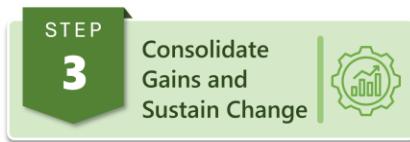
Once a shared vision for AI is in place, the next step is broad communication and empowerment. Kotter emphasizes the need to communicate the vision continuously, not just through one-time announcements but via ongoing dialogue (e.g. town halls, pilot project stories, and open conversations) to help clarify AI’s role. Staff need clarity on where AI will be used, what it will do, and how decisions are made; lack of information often leads to fear and misinformation (Creasey, 2025). Concerns such as job loss or biased applications can be addressed by communication plans tailored to different groups (Valiance Solutions, 2024). Managers play a key part in facilitating two-way discussions, making staff feel involved and heard. Staff are more likely to support AI if they have had a voice in the process.

Enablement must go hand-in-hand with communication. Kotter’s Step 5 stresses removing obstacles, while ADKAR highlights the need to build *Knowledge* and *Ability*. This includes training staff on how to use AI, explaining workflow changes, and supporting skill development. Different roles will interact with AI differently, so each group needs to learn in context of their role. Training should be practical and role-specific, using formats like hands-on workshops, online modules, and peer learning to spread AI capability throughout the organization (Valiance Solutions, 2024); (Creasey, 2025).

The TAM reinforces that adoption depends on perceived usefulness and ease of use (Davis, 1989). To boost usefulness, show how AI benefits daily work (e.g. saving engineers hours of analysis or helping sales teams target leads) (see Section 5.4.2). Sharing results from pilots can help staff visualize value. To improve ease of use, invest in intuitive design and iterative testing (see Section 5.3.1). Simplifying interfaces and integrating tools into existing systems helps increase adoption (Singh, 2025). Strong support systems like help desks and FAQs further ease the transition (Valiance Solutions, 2024).

Finally, fostering trust is critical. Many change efforts fail because people feel technology is being imposed without their input (Creasey, 2025). Leaders should directly address concerns, frame AI as a partner that reduces tedious work, and highlight how roles will evolve—not disappear. Demonstrating human–AI collaboration, like using AI to sift data while humans focus on complex analysis, reinforces this message.

Trust also grows when employees understand how AI makes decisions. Explainability (showing why an AI made a specific recommendation) helps users feel confident in the tools (Doshi-Velez, 2017). When employees see that AI is being implemented with care, transparency, and principled oversight, their confidence in the technology grows (Mittelstadt, 2016). Effective engagement means equipping staff with knowledge and tools while also “winning hearts and minds” through transparency, support, and shared purpose.



(Kotter Steps 6–8)

Launching AI initiatives is just the beginning; organizations must sustain momentum to achieve full adoption. In Kotter’s framework, Step 6 encourages generating short-term wins - quick, visible successes that validate the effort. Identify early pilot projects or use-cases where AI can deliver a tangible win in a short time frame. For example, an early win might be an internal general purpose chatbot, such as CMS Chat, that supports employees in a range of tasks and earns high satisfaction scores, or a predictive maintenance system that cuts downtime by 20%. Celebrate and publicize these wins. Early successes provide proof points to skeptics and create positive buzz, which is essential to influence the early majority in the organization’s own adoption curve (Rahn, 2024). In Diffusion of Innovation terms, the innovators and early adopters within the company help “sell” the innovation to the more cautious groups by showing that it works and has advantages. Observability of AI’s benefits is key - when others can plainly see the improvements AI brings, peer pressure and curiosity build for others to jump onboard. Additionally, encourage those early adopters to become “change ambassadors,” mentors, or igniters as they are positioned in the AI Ignite program. This social influence can accelerate uptake as colleagues trust recommendations from peers. By deliberately leveraging these dynamics, change leaders can broaden the adoption from a few teams to many (Rahn, 2024).

Kotter’s Step 7 is about consolidating gains and driving more change. After initial wins, don’t declare victory too early - use the momentum to tackle additional areas where AI can add value. Continuously expand and iterate on the AI strategy: perhaps after success with internal workplace efficiencies, move to applying AI in external facing use cases that will impact external stakeholders. Each rollout should incorporate lessons learned from earlier ones, refining the approach (for instance, improving training programs or adjusting communication tactics based on feedback). It’s helpful to maintain a feedback loop: gather input from users about what’s working or where they face challenges and use that to make adjustments (Valiance Solutions, 2024 ). This iterative improvement keeps the adoption moving forward and demonstrates a commitment to getting it right. Over time, these practices contribute to building an AI-adaptive culture—one that values learning, experimentation, and continuous improvement. When employees see that their feedback is acted on and that AI tools keep getting better (and easier to use) with each iteration, they remain engaged and open to further change.

Finally, Kotter’s Step 8 calls for anchoring the new approaches in the culture. Sustainable AI adoption means that using AI becomes the “new normal” in how the organization operates. To anchor AI in the culture, integrate it into everyday processes and standard operating procedures. Update job descriptions, performance metrics, and incentives to reflect AI-augmented roles (for example, rewarding teams not just for results, but for smart use of data and AI insights in decision-making). Recognize and reward employees who embrace the AI tools, so that others see that the behavior is valued (this ties in with ADKAR Reinforcement - ensuring people don’t slip back into old ways) (Brusati, 2025). It’s also crucial to continue leadership involvement at this stage (Valiance Solutions, 2024 ).

## CMS Artificial Intelligence Playbook

---

Leaders and managers should model the desired mindset by using AI in their own work and endorsing its use in meetings and decisions. When top leadership routinely ask questions such as: “*What do the AI-driven insights suggest?*,” it signals that data-driven, AI-enabled thinking is part of the company’s DNA.

Another aspect of anchoring is maintaining the trust and governance structures that were built during the change. As AI is scaled enterprise-wide, continue to uphold transparency and ethical practices. This might involve establishing an AI governance committee or ongoing audits of AI decisions for fairness and accuracy. Such measures ensure that as dependency on AI grows, trust in AI remains high. Studies have noted that trust (or distrust) can significantly influence the *rate of diffusion* of AI in an organization (Afroogh, 2024). Thus, anchoring AI in culture isn’t only about technology integration, but also about solidifying a trust-based, human-centered approach to AI use.

### In Summary

To summarize, this playbook recommends weaving together multiple change frameworks into a cohesive plan for change. Table 38 provides a quick reference to these frameworks and how they inform an AI change management:

Table 38. Change Management Frameworks and Concepts

Framework / Concept	Focus	Application to AI Adoption
Kotter's 8 Steps	Organizational change process (top-down)	Roadmap from urgency to anchoring. Begin by establishing urgency (e.g. highlighting AI's competitive edge) and building a guiding coalition; end by anchoring AI practices in culture.
Prosci ADKAR Model	Individual change journey (bottom-up)	Ensures each person transitions: build Awareness of AI's need and Desire to participate; provide Knowledge & Ability via training and support; and Reinforce adoption through recognition and ongoing support.
TAM	User acceptance of technology	Emphasizes <i>perceived usefulness</i> and <i>ease of use</i> . Choose AI tools that demonstrably help employees work better and make them user-friendly with sufficient training. These factors strongly influence the intention to use AI.
Diffusion of Innovation	How innovation spreads in a social system	Leverage early adopters to pilot AI solutions and showcase results. Their success stories (and the observable benefits) encourage the early majority to follow. Plan for a phased rollout (trials, then broader adoption) to accommodate different adopter readiness levels.
Explainability & Trust	Building confidence in AI systems	Make AI decisions explainable and communicate the safeguards in place. When users understand and trust AI outputs, adoption increases. Embed ethical guidelines and highlight AI as a tool to augment human work (not replace it) to foster a collaborative human–AI culture.

Managing AI-driven change is a journey that blends technology deployment with careful change leadership. By creating urgency, crafting a vision, and enabling people at all levels (while continually building trust) organizations can navigate from initial AI experiments to a point where AI is an integral aspect of how work gets done. The frameworks above serve as guideposts and keeps efforts people-centered. The result is a successful implementation of new AI tools with *sustainable adoption* where the organization and its people are continuously learning, improving, and flourishing with AI.

## Key Takeaways - Looking Ahead

Chapter 6 explored emerging AI technologies and organizational preparations that will shape CMS' future. While particularly relevant for leadership and managers who must guide strategic planning and resource allocation, this forward-looking perspective helps all CMS audiences understand and prepare for continued innovation in healthcare transformation.



- Near-term technologies like integrated assistants, agentic AI, multimodal AI, and digital twins are becoming more commonplace and offer immediate practical applications for CMS.
- Organizations must align their administrative structures, policies, resources, and workforce to effectively harness AI's evolving capabilities. This will require flexible governance, hiring and upskilling, open collaboration, and ongoing change management.
- Embrace a structured, human-centered change management strategy—grounded in frameworks like Kotter's 8-Step Process, the Prosci ADKAR Model, and the Technology Acceptance Model—to build trust, ensure transparent communication, and empower staff. This approach not only addresses technical challenges but also reassures employees that AI will augment their roles rather than replace them.

Through its examination of AI fundamentals, current maturity efforts, AI governance, implementation approaches, and future directions, this Playbook establishes a foundation for CMS to advance its AI maturity while safeguarding public trust. References and additional resources and templates can be found in the Appendices that follow.

## References

- Afroogh, S. A. (2024, November 18). *Trust in AI: Progress, challenges, and future directions*. *Humanities and Social Sciences Communications*, 11, Article 1568. Retrieved from Nature: <https://doi.org/10.1057/s41599-024-04044-8>
- Altman, S. G. (2023, February 24). Retrieved from OpenAI Blog: <https://openai.com/blog/planning-for-agi-and-beyond>.
- ASTP/ONC. (2025, February 13). *Artificial Intelligence (AI) at HHS*. Retrieved from Official Website of the Assistant Secretary for Technology Policy/Office of the National Coordinator for Health IT: <https://www.healthit.gov/topic/artificial-intelligence>
- Braiterman, Z., Shostack, A., Marcil, J., de Vries, S., Michlin, I., Wuyts, K., . . . French, M. (2020). *Threat Modeling Manifesto*. Retrieved from <https://www.threatmodelingmanifesto.org/>
- Brusati, I. (2025, March 21). *ADKAR vs Kotter: Which Change Model Should You Choose?* Retrieved from Prosci: <https://www.prosci.com/blog/adkar-vs-kotter>
- CMS AI Explorers. (2024, December 19). *A Practical Understanding of Threat Modeling for AI Systems*. Retrieved from Confluence: <https://confluenceent.cms.gov/x/o5HDN>
- CMS AI Explorers. (2025, February 10). *Assessing Algorithmic Risk and Impact in Artificial Intelligence Systems*. Retrieved from CMS Enterprise Confluence: <https://confluenceent.cms.gov/x/i4QKNw>
- CMS AI Explorers. (2025, January 8). *Crafting Prompts at CMS*. Retrieved from CMS Enterprise Confluence: <https://confluenceent.cms.gov/x/FcpiNg>
- CMS AI Explorers. (2025, February 10). *Human-Centered Artificial Intelligence (AI) at CMS*. Retrieved from CMS Enterprise Confluence: <https://confluenceent.cms.gov/x/v7fHLQ>
- CMS AI Explorers. (2025, February 10). *LLM Cost and Quality Comparisons*. Retrieved from CMS Enterprise Confluence: <https://confluenceent.cms.gov/x/QoRPMQ>
- CMS AI Explorers. (2025, April). *Methodology for Creating Large Language Model (LLM) Evaluation Labels*. Retrieved from CMS Enterprise Confluence: <https://confluenceent.cms.gov/x/BAF0NQ>
- CMS AI Explorers. (2025, February 18). *Technical Application Resources*. Retrieved from CMS Enterprise Confluence: <https://confluenceent.cms.gov/x/vbfHLQ>
- Creasey, T. (2025, February 13). "Mapping AI Adoption Research Insights to ADKAR." . Retrieved from LinkedIn Pulse: <https://www.linkedin.com/pulse/mapping-ai-adoption-research-insights-adkar-tim-creasey-ptjrc>
- Davis, F. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, 319-340.
- DHS Artificial Intelligence Roadmap*. (2024, March 17). Retrieved February 11, 2025, from <https://www.dhs.gov/publication/ai-roadmap>
- Dickson, B. (2024, January 19). *What Is Spatial Computing? A Basic Explainer*. Retrieved from PC Magazine: <https://www.pc当地.com/how-to/what-is-spatial-computing-a-basic-explainer>

## CMS Artificial Intelligence Playbook

---

- Doshi-Velez, F. &. (2017). *Towards a rigorous science of interpretable machine learning.* arXiv:1702.08608. Retrieved from arXiv : <https://arxiv.org/abs/1702.08608>
- Ferguson, T., Olds, T., Curtis, R., Blake, H., Crozier, A., & Dankiw, K. (2022, August). *Effectiveness of wearable activity trackers to increase physical activity and improve health: a systematic review of systematic reviews and meta-analyses.* Retrieved from The Lancet Digital Health 4 (8): e676–e689: [https://doi.org/10.1016/S2589-7500\(22\)00111-X](https://doi.org/10.1016/S2589-7500(22)00111-X)
- Fountaine, T., McCarthy, B., & Saleh, T. (2019, July). Building the AI-Powered Organization. *Harvard Business Review.* Retrieved February 6, 2025, from <https://hbr.org/2019/07/building-the-ai-powered-organization>
- GAO. (2023, February 14). *Science & Tech Spotlight: Digital Twins—Virtual Models of People and Objects.* Retrieved from <https://www.gao.gov/products/gao-23-106453#:~:text=For%20example%2C%20one%20company%20builds,and%20visualize%20the%20effects%20of>
- GAO. (2024, August 22). *Immersive Technologies: Most Civilian Agencies Are Using or Plan to Use Augmented Reality, Virtual Reality, and More.* Retrieved from <https://www.gao.gov/products/gao-24-106665#:~:text=Federal%20civilian%20agencies%20use%20immersive,transportation%20security%20training%2C%20and%20fire>
- Goswami, R. (2025, January 7). "Change Management Tech and Kotter's Eight-Step Change Model.". Retrieved from CTO Magazine: <https://ctomagazine.com/change-management-tech>
- GSA Blog. (2025, January 10). *Artificial Intelligence delivers real results through GSA.* Retrieved from <https://www.gsa.gov/blog/2025/01/10/artificial-intelligence-delivers-real-results-through-gsa#:~:text=GSA%20was%20one%20of%20the,pilot%2C%20which%20we%E2%80%99re%20now%20expanding>
- Interaction Design Foundation. (2025). *What is Human-AI Interaction (HAX)?* Retrieved 3 12, 2025, from Interaction Design Foundation: [https://www.interaction-design.org/literature/topics/human-ai-interaction?srsltid=AfmBOoq84BhW3IJRVW7qqJ1-RfizyIP9gzYT4lzJBiyuYk2Ai2JN1wDv#what\\_is\\_human-ai\\_interaction\\_\(hax\)?-0](https://www.interaction-design.org/literature/topics/human-ai-interaction?srsltid=AfmBOoq84BhW3IJRVW7qqJ1-RfizyIP9gzYT4lzJBiyuYk2Ai2JN1wDv#what_is_human-ai_interaction_(hax)?-0)
- Intro to MLOps: Data and Model Versioning.* (2023, 2 3). Retrieved from Weights & Biases: <https://wandb.ai/site/articles/intro-to-mlops-data-and-model-versioning/>
- Introduction to Remote Moderated Usability Testing.* (2018, November 14). Retrieved February 12, 2025, from <https://18f.gsa.gov/2018/11/14/introduction-to-remote-moderated-usability-testing-part-1/>
- Kennedy, R. (2020). *Strategic Management.* Blacksburg, VA: Pamplin College of Business in association with Virginia Tech Publishing. Retrieved from <https://pressbooks.lib.vt.edu/strategicmanagement/chapter/7-4-types-of-innovation/>
- Microsoft. (n.d.). *Microsoft Dragon Copilot.* Retrieved March 15, 2025, from microsoft: <https://www.microsoft.com/en-us/health-solutions/clinical-workflow/dragon-copilot>

## CMS Artificial Intelligence Playbook

---

- Mittelstadt, B. D. (2016). *The ethics of algorithms: Mapping the debate*. 3(2) 2053951716679679. Retrieved from Big Data & Society, : <https://doi.org/10.1177/2053951716679679>
- Moran, K. (2019, December 1). *Usability (User) Testing 101*. Retrieved April 6, 2025, from <https://www.nngroup.com/articles/usability-testing-101/>
- NIST. (2023, January). *AI Risk Management Framework*. Retrieved from National Institute of Standards and Technology: [https://airc.nist.gov/AI\\_RMF\\_Knowledge\\_Base/Playbook](https://airc.nist.gov/AI_RMF_Knowledge_Base/Playbook)
- Office of Management and Budget (OMB). (2025, April 3). *Accelerating Federal Use of AI through Innovation, Governance, and Public Trust*. Retrieved from <https://www.whitehouse.gov/wp-content/uploads/2025/02/M-25-21-Accelerating-Federal-Use-of-AI-through-Innovation-Governance-and-Public-Trust.pdf>
- Rahn, C. (2024, November 13). *Technology adoption models & AI*. . Retrieved from NathanRahn.com: <https://www.nathanrahn.com/operations-strategy-leadership/2024/11/13/technology-adoption-models-amp-ai>
- Reymond, G.-O. (2025, January 03). *How Quantum Computing Is Changing Drug Development at the Molecular Level*. Retrieved from World Economic Forum: <https://www.weforum.org/stories/2025/01/quantum-computing-drug-development/>
- Rutherford, C. (2024, July 18). Debunking LLMs for Non-Techies and Leaders.
- Rutherford, C. D. (2024, November 27). Responsible AI Governance - Right-Sizing Risk Evaluation and Mitigation for AI Use.
- Saad, F., & Elson, A. (2025, February). *Next-Generation AI Architectures: Comparative Analysis of Neural, Symbolic, and Hybrid Learning Approaches*. Retrieved March 26, 2025, from <https://www.researchgate.net/profile/Aric-Elson/publication/389389247>
- Selbst, A. D. (2021). *An Institutionalized View of Algorithmic Impact Assessments*. Retrieved from Harvard Journal of Law & Technology: <https://jolt.law.harvard.edu/assets/articlePDFs/v35/Selbst-An-Institutionalized-View-of-Algorithmic-Impact-Assessments.pdf>
- Singh, P. D. (2025, April 6). "Generative AI through the Lens of Technology Acceptance Model." *SSRN Working Paper No. 4953174*. Retrieved from SSRN : [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=4953174#:~:text=how%20Perceived%20Usefulness%20,factors%20include%20effective%20organizational%20training](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4953174#:~:text=how%20Perceived%20Usefulness%20,factors%20include%20effective%20organizational%20training)
- Stanford Institute for Human-Centered Artificial Intelligence. (2023). *AI Index Report 2023*. Stanford, CA: Stanford University.
- Stanford University. (2021, September). *Gathering Strength, Gathering Storms: The One Hundred Year Study on Artificial Intelligence (AI100) 2021 Study Panel Report*. Retrieved March 27, 2025, from [https://ai100.stanford.edu/sites/g/files/sbiybj18871/files/media/file/AI100Report\\_MT\\_10.pdf](https://ai100.stanford.edu/sites/g/files/sbiybj18871/files/media/file/AI100Report_MT_10.pdf)
- The White House. (2025, April 7). *Fact Sheet: Eliminating Barriers for Federal Artificial Intelligence Use and Procurement*. Retrieved from <https://www.whitehouse.gov/fact-sheets/2025/04/fact-sheet-eliminating-barriers-for-federal-artificial-intelligence-use-and-procurement/>

## CMS Artificial Intelligence Playbook

---

- Types of Harm.* (2023, January 24). Retrieved from Microsoft Learn: <https://learn.microsoft.com/en-us/azure/architecture/guide/responsible-innovation/harms-modeling/type-of-harm>
- U.S. Office of Management and Budget (OMB). (2025, April 3). *Driving Efficient Acquisition of Artificial Intelligence in Government*. Retrieved from Office of Management and Budget: <https://www.whitehouse.gov/wp-content/uploads/2025/04/M-25-22-AI-Acquisition.pdf>.
- U.S. Office of Management and Budget (OMB). (2025, April 3). *Memorandum M-25-21: Accelerating Federal Use of Artificial Intelligence Through Innovation, Governance, and Public Trust*. Retrieved from Office of Management and Budget: <https://www.whitehouse.gov/wp-content/uploads/2025/04/M-25-21-AI-Governance.pdf>.
- US Code.* (2025, February 10). Retrieved February 11, 2025, from [https://uscode.house.gov/view.xhtml?req=\(title:15%20section:9401%20edition:prelim](https://uscode.house.gov/view.xhtml?req=(title:15%20section:9401%20edition:prelim)
- Valiance Solutions. (2024 , January 30). "Streamlining Work with Generative AI: A Guide to Change Management." . Retrieved from <https://valiancesolutions.com/2024/01/30/streamlining-work-with-generative-ai-a-guide-to-change-management/>
- Vinney, C. (2022, October 6). *Desirability, feasibility and viability diagram: What does it mean?* Retrieved April 1, 2025, from <https://uxdesigninstitute.com/blog/desirability-viability-and-feasibility/>
- Vipra, J., & Myers West, S. (2023, September 27). *Computational Power and AI*. Retrieved March 26, 2025, from <https://ainowinstitute.org/publication/policy/compute-and-ai>

## Appendices

### Appendix A. External Resources

The following table lists external resource artifacts that were referenced throughout this Playbook.

*Table 39. External Resources*

Tool	Description	Page
<a href="#">MLTRL</a>	A framework that adapts the concept of technology readiness levels (TRLs) to machine learning, guiding teams through stages from research to deployment and ongoing monitoring.	57
<a href="#">USDS Discovery Sprint Guide</a>	A structured, five-day process used to rapidly prototype and test digital services for the federal government.	42, 43
<a href="#">Microsoft AI Maturity Model for Government Agencies</a>	Provides a roadmap for government agencies to assess their current AI capabilities and plan for strategic growth.	13
<a href="#">Principles behind the Agile Manifesto</a>	Twelve core principles that emphasize individuals and interactions, working software, customer collaboration, and responsiveness to change.	38
<a href="#">GSA Purchasing Guidance</a>	Supports federal agencies in acquiring technology and professional services effectively and in compliance with regulations.	47
<a href="#">CRISP ML(Q)</a>	An extension of the CRISP-DM methodology tailored to machine learning projects, adding quality assurance and operational readiness checks.	57
<a href="#">Pandas</a>	A Python library providing high-performance, easy-to-use data structures like DataFrames for data manipulation and analysis.	53, 54
<a href="#">Matplotlib</a>	A foundational Python plotting library for creating static, animated, and interactive visualizations.	53
<a href="#">Seaborn</a>	A Python data visualization library built on top of Matplotlib, offering a high-level interface for drawing attractive and informative statistical graphics.	53
<a href="#">ggplot2</a>	An R-based data visualization package implementing the grammar of graphics for building complex plots layer-by-layer.	53
<a href="#">dplyr</a>	An R package focused on efficient data manipulation, providing a grammar for data transformation through verbs like <code>filter</code> , <code>mutate</code> , and <code>summarize</code> .	53
<a href="#">NumPy</a>	A fundamental Python package for scientific computing, offering support for large, multi-dimensional arrays and a wide array of mathematical operations.	54

## CMS Artificial Intelligence Playbook

---

Tool	Description	Page
<a href="#">Scikit-learn</a>	A widely-used Python library for machine learning, offering simple and efficient tools for data mining and predictive modeling.	54
<a href="#">caret</a>	An R package that streamlines the process of building and evaluating machine learning models by wrapping around numerous algorithms and preprocessing tools.	54
<a href="#">Microsoft's Human-AI Experience (HAX) Toolkit</a>	Provides designers and developers with practical tools and principles to build AI systems that are useful, trustworthy, and human-centered.	52
<a href="#">WCAG 2.1 Guidelines</a>	Provide internationally recognized standards for making web content perceivable, operable, understandable, and robust for all users.	52
<a href="#">Kotter's 8-Step Process</a>	Outlines eight steps for leading organizational change, starting with creating urgency and ending with anchoring new practices into the culture.	71-75
<a href="#">Prosci ADKAR Model</a>	Focuses on individual change through five key stages: Awareness, Desire, Knowledge, Ability, and Reinforcement.	71-75
<a href="#">Technology Acceptance Model (TAM)</a>	Explains how users come to accept and use technology based on perceived usefulness and perceived ease of use.	61, 71-74
<a href="#">Diffusion of Innovation</a>	Describes how new ideas and technologies spread through populations over time, influenced by factors like adopter categories and social systems.	71, 73, 74

## Appendix B. Internal Resources

The following tables list AI resources internal to CMS that were referenced throughout this Playbook. Access to most links in this section are restricted to users with appropriate CMS credentials, such as those with CMS Enterprise Confluence access. CMS federal employees and contractors should reach out to [CMS\\_IT\\_Service\\_Desk@cms.hhs.gov](mailto:CMS_IT_Service_Desk@cms.hhs.gov) or [ai@cms.hhs.gov](mailto:ai@cms.hhs.gov) regarding any access issues. These lists are not comprehensive of all resources available across the agency.

- **Table 40:** General AI tools and training that are available for CMS staff to use.
- **Table 41:** AI technical tools and workspaces for teams working on AI projects.
- **Table 42:** Links to CMS tools and white papers developed by the AI Explorers team.
- **B.1.- B.8.**: Full versions of CMS tools intended to support teams with AI project development.

*Table 40. General AI Tools and Trainings*

Resource	Description
<a href="#">AI Community Slack Channel</a>	A channel in CMS-enterprise Slack that provides a space for CMS staff to learn about AI, share relevant AI news, policy changes, and technology advancements.
<a href="#">CMS Chat</a>	A secure, generative AI tool custom-created for internal CMS use. Troubleshooting support and community feedback can be shared in its dedicated <a href="#">CMS Chat Slack channel</a>
<a href="#">AI Ignite</a>	A micro-training program that teaches CMS employees how to use CMS Chat to enhance their daily work (e.g. streamlining tasks, drafting reports, editing documents, and improving workflows)
<a href="#">Workforce Resilience Program</a>	Training offered to all CMS employees to learn new skills and technologies for the future, including but not limited to AI/ML; its 3 new AI workshops include: <ul style="list-style-type: none"> <li>• What are Large Language Models (LLMs)?</li> <li>• How to Write a Better Prompt</li> <li>• Doing Prompt Engineering</li> </ul>
<a href="#">Microsoft 365 Copilot</a>	A secured AI assistant embedded into Microsoft 365 apps (Word, Excel, Outlook, Teams, etc.) that helps users be more productive with tasks like drafting emails, summarizing meetings, generating reports, and analyzing data.

*Table 41. AI Development Resources*

Resource	Description
<a href="#">AI Workspace</a>	A CMS-built secure cloud environment that enables AI development, rapid prototyping, and experimentation while reducing technical barriers to adoption. Login can be accessed at <a href="http://aiworkspace.cms.gov">aiworkspace.cms.gov</a> and requires prior approval.
<b>Integrated Data Repository (IDR) Customer Analytic Environments (CAEs)</b>	The new IDR capability offers fully managed Customer Analytic Environments (CAEs) that integrate with the IDR Snowflake platform, allowing customers to perform AI, ML, and data science workloads securely within the IDR's FISMA boundary—eliminating the need to manage their own infrastructure.

## CMS Artificial Intelligence Playbook

---

Resource	Description
<a href="#">Python Workspace</a>	The Python workspace is a secure, pre-configured development environment that enables users with pseudo privileges to perform data analysis, scripting, and AI/ML development using Python and R. The workspace is built on Anaconda. It is connected to the IDR via Snowflake, ACO-OS (DB2), and the Informatica server, allowing users to access and manipulate CMS datasets efficiently within a controlled, compliant environment.
<a href="#">Amazon Web Services</a>	A comprehensive cloud service platform that offers scalable infrastructure and tools for building, deploying, and managing applications, including AI and machine learning solutions
<a href="#">Azure</a>	Microsoft's cloud computing platform providing integrated services for computing, storage, analytics, and AI to support enterprise-grade application and model development
<a href="#">GitHub Copilot</a>	An AI-powered code completion tool, developed by GitHub and powered by OpenAI, that helps developers write code faster by suggesting code snippets, functions, and entire modules as they type. <i>Procurement is in progress.</i>
<a href="#">Open Source Program Office (OSPO)</a>	OSPO serves as the center of competency for CMS' open source operations and structure. It is responsible for defining and implementing strategies and policies to guide these efforts.
<a href="#">Infrastructure Users and Services Group (IUSG) Production Prerequisites Checklist</a>	A standardized readiness checklist that outlines the required tasks, documentation, and compliance steps for deploying applications into the CMS Hybrid Cloud environment, ensuring production readiness across security, monitoring, continuity, and governance requirements.

Table 42. CMS Tools and White Papers from the AI Explorers Program

ID	Tool	Associated White Paper	Refer to Section
B.1.	CMS Chat Prompt Template	<a href="#">Crafting Prompts at CMS</a>	3.2.2
B.2.	Governance Approach Questionnaire Samples	<a href="#">Governance Sample Question Sets</a>	4.2.1
B.3.	Human-Centered AI Matrix	<a href="#">Human-Centered AI at CMS</a>	5.2.4
B.4.	AI System Algorithmic Risk and Impact Self-Assessment	<a href="#">Assessing Algorithmic Risk and Impact in Artificial Intelligence Systems</a>	5.2.4
B.5.	AI Threat Model Template	<a href="#">A Practical Understanding of Threat Modeling for AI Systems</a>	5.3.3
B.6.	Large Language Model (LLM) Evaluation Labels	<a href="#">Choosing the Right Large Language Model (LLM) for Your Project: A Task Specific Approach</a>	5.3.4
B.7.	LLM Implementation Checklist	<a href="#">LLM Cost and Quality Comparisons</a>	5.3.4
B.8.	Building Blocks of Agentic AI	<a href="#">Understanding Agentic AI and Its Potential at CMS</a>	6.1.1

Note: Up-to-date list of AI Explorers white papers is available on [Confluence](#) (CMS AI Explorers, 2025).

### B.1. CMS Chat Prompt Template

#### Basic Structure

[Task] + [Role] + [Guidelines] + [Reference Context (Reference Materials)]

Each are further defined below:

- **[Task]** - The action(s) you wish the AI to take based on the [Role], [Guidelines], [Format Requirements], and [Reference Context (Reference Materials)]
- **[Role]** - The role that the AI is to take on to complete the [Task]
- **[Guidelines]** - Rules or guidance for the AI to ensure it accomplishes the task as expected. This may include the expected output structure/format or instructions on how to write (perspective, style, etc.), and other guidance for the AI.
- **[Reference Context (Reference Materials)]** - relevant attachments, details, documentation, resource materials, etc. that can act as the source of truth for the AI to accomplish its [Task]

#### Template

As a **[Role]**, complete the task **[Task]**, following the Guidelines: **[Guidelines]** using the relevant details from **[Reference Context]**.

#### Examples

##### Document Analysis

As a Medicare policy analyst with expertise in coverage determinations, complete the task reviewing the attached Local Coverage Determination document and analyzing the key changes from the previous version and their potential impact on providers. Consider the current national coverage guidelines for [specific treatment/service], following the Guidelines: present findings in a structured format with main changes in bullet points and supporting policy citations using the relevant details from [the attached Local Coverage Determination document and national guidelines for [specific treatment/service]].

##### Policy Interpretation

As a senior CMS compliance expert, complete the task analyzing whether [specific scenario] meets current requirements in light of the recent updates to telehealth reimbursement policies for 2024, following the Guidelines: present the analysis in a clear, structured format with specific recommendations and include relevant regulatory citations using the relevant details from [the updated telehealth reimbursement policies for 2024].

##### Content Generation

As a healthcare communications specialist familiar with CMS beneficiary materials, complete the task drafting a clear explanation of [specific policy/change], following the Guidelines: adhere to plain language requirements, structure the content with headers and bullet points for key takeaways, and follow Medicare Communications and Marketing Guidelines using the relevant details from [specific policy/change materials and Medicare Communications and Marketing Guidelines].

### Information Synthesis

As a Medicare program analyst, complete the task synthesizing the key requirements from the provided documents about [specific topic] into a consolidated summary. Focus particularly on areas where requirements overlap or potentially conflict, following the Guidelines: organize the analysis into sections and provide clear citations to source documents using the relevant details from [the provided documents about [specific topic]].

### B.2. Governance Approach Questionnaire Samples

Question Set	Question
Intake	What is the project name?
Intake	Who is your client (name or organization)?
Intake	What mission area(s) does your project align with?
Intake	Who is on your team?
Intake	What is the purpose of your project?
Intake	Does the project involve AI or advanced data analytics activities?
Intake	6a. People on your team use AI specifically for this project's success?
Intake	6b. People on your team are building AI/ML models for this project?
Intake	6c. People on your team are building software with AI-enabled functionality?
Intake	6d. People on your team are building decision-support or advanced analytics (non-AI but complex data science)?
Intake	6e. Our project has no data science, AI, or ML involvement.
Intake	Any additional comments or clarifications?
Low Scorecard	What type of data will the project use?
Low Scorecard	Will this project influence mission-critical functions or government policy?
Low Scorecard	Who are the primary end users of this project?
Low Scorecard	Does the project require access to or use of secure or classified data?
Low Scorecard	Does this project require considerations related to fairness, ethics, or regulatory compliance (Could it cause harm)?
Low Scorecard	Does this project require considerations related to fairness, ethics, or regulatory compliance?
Low Scorecard	Is this project governed by a client's own AI governance or compliance process?
Low Scorecard	What stage is the project currently in, and do you plan to move to production soon?
Moderate Scorecard	Will the project share data with external partners or clients who do not have a formal data protection agreement?
Moderate Scorecard	Does this project fall under specific regulatory frameworks (e.g., HIPAA, GDPR, CJIS, ITAR) requiring special compliance?
Moderate Scorecard	Have you identified potential biases in the dataset(s) that require specific mitigation or fairness measures?
Moderate Scorecard	Approximately how many end-users or stakeholders will directly rely on the project?
Moderate Scorecard	Have you performed a basic ethical or fairness review (internal or external) on the project's approach?

## CMS Artificial Intelligence Playbook

---

Question Set	Question
Moderate Scorecard	Will sensitive or proprietary data be stored or transmitted on external cloud or partner environments?
Moderate Scorecard	Does your project do any of the following?
Moderate Scorecard	Is there a risk of adversarial attacks (e.g., data poisoning, model manipulation) that require special controls?
High Scorecard	Does this project handle or influence areas tied to national security, defense, or classified missions?
High Scorecard	How widespread is the deployment or usage of this solution (e.g., user base, agencies, public impact)?
High Scorecard	Are advanced or specialized encryption methods required to safeguard data at rest or in transit due to legal or gov. mandates?
High Scorecard	Have we established a formal monitoring program for adversarial attacks (poisoning, model evasion)?
High Scorecard	Do you anticipate legal disputes or high liability from the AI's decisions (e.g., health/life, civil rights)?
High Scorecard	Could unintended consequences or model bias result in large-scale harm or discrimination?
High Scorecard	Is the AI integrated into mission-critical business processes or systems at an enterprise/government scale?
High Scorecard	Does the project have a formal incident response plan covering AI malfunctions or data breaches?
Special Scorecard	Are you utilizing or generating classified data, such as Secret, Top Secret, or compartmentalized information (SCI)?
Special Scorecard	Does the project directly support national security or defense-related objectives (e.g., intelligence, weapon systems, critical infrastructure)?
Special Scorecard	Are multiple government agencies (federal or international) involved, each requiring classified data sharing or specialized compliance?
Special Scorecard	Is there a documented threat of foreign intelligence services or advanced adversaries specifically targeting this system or data?
Special Scorecard	Does this project invoke special policy boards or committees (e.g., DoD ethics boards, intelligence oversight bodies) for formal approval?
Special Scorecard	Does the project involve ITAR/EAR-controlled technology or sensitive export controls with foreign partners?
Special Scorecard	Is there a classified incident response process that integrates with federal or DoD-level protocols (e.g., real-time intelligence sharing)?
Special Scorecard	Are OPSEC protocols (operational security) in place to minimize the risk of unauthorized data disclosure or infiltration?

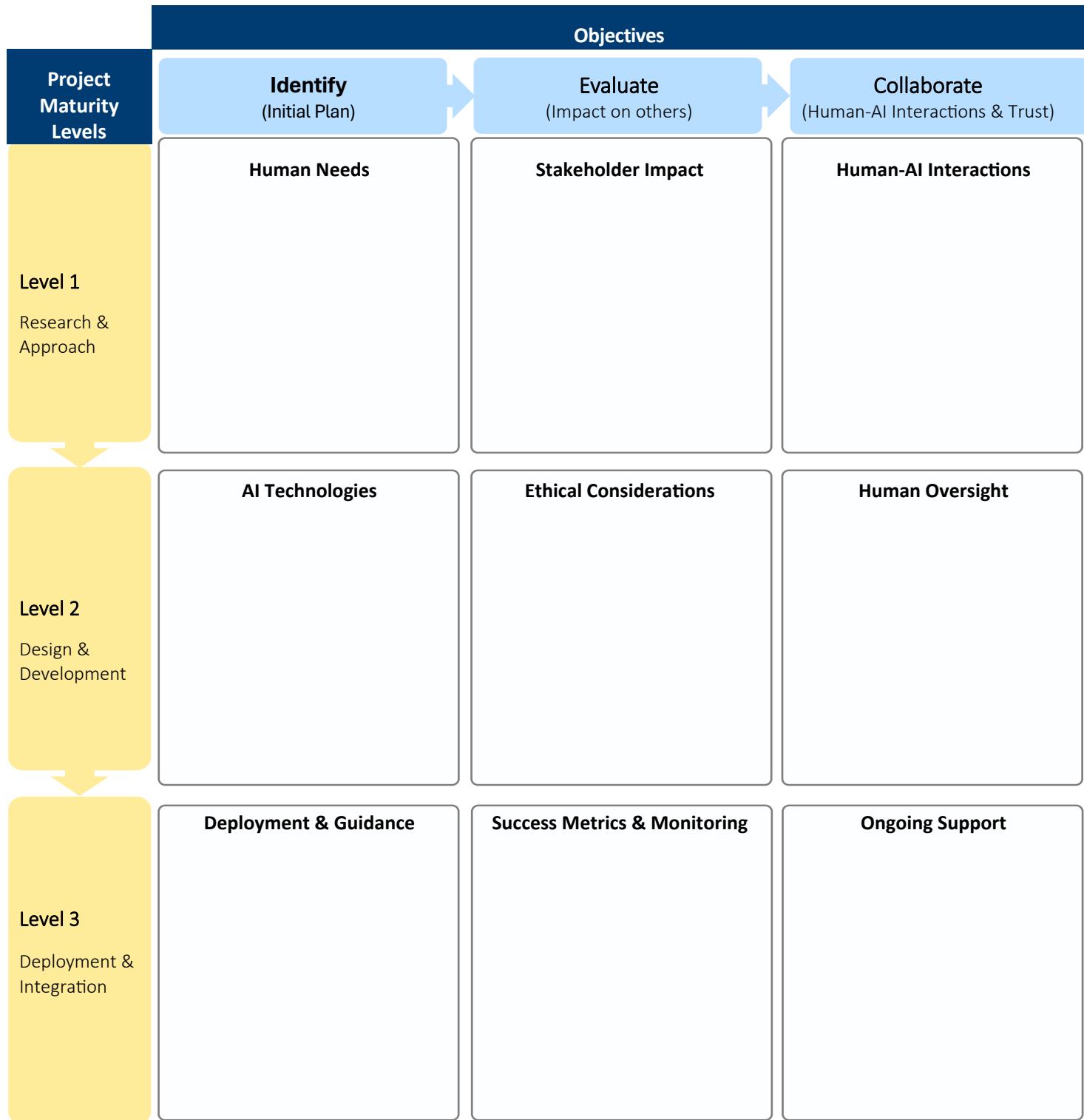
**B.3. Human-Centered AI Matrix**

Project Name:

Date:

Designed for:

Designed by:



### B.4. AI System Algorithmic Risk and Impact Self-Assessment Template

#### **Part 1: General Information**

- 1a) AI system / project name:
- 1b) Representative(s) performing the ARIA (Name and OpDiv):
- 1c) ARIA date completed:
- 1d) Circle or highlight the furthest stage your AI system has reached. Complete up to parts 2, 3, and 4 respectively for your project's stage of maturity. (*e.g. an AI tool that has been developed but not released to beta users would complete parts 2 and 3, but not part 4.*)  
[ DESIGNED / DEVELOPED / DEPLOYED ]

#### **Part 2: AI System Design**

Reference: [CMS Human-Centered AI Matrix](#), Levels 1 and 2

- 2a) **Human Needs:** Describe the purpose and objectives of your AI system:
- 2b) **Human-AI Interactions:** Describe how each intended user of the AI system will use and interact with the system (i.e., user journeys and how decisions or actions are influenced by AI outputs):
- 2c) **AI Technologies:**
  - i. Describe the AI capabilities being used in your system:
  - ii. Describe the data and information your AI system will have access to (model training/testing data, expected user inputs, outputs):
  - iii. Describe what levels of automation, if any, will be present in your AI system:
- 2d) **Human Oversight:** Describe which areas will need human oversight in your AI system:

**Stakeholder Impact and Ethical Considerations:** List all risks identified within your system in the table below by providing a brief description of the impact, classification for the impact type, the subjected stakeholder(s), and the primary mechanism(s) inducing the impact. Negative impacts and mechanisms glossaries are available in Appendices A1 and A2. Consider severity and likelihood to estimate risk level (low, moderate, high). Then, document any intended mitigation tactics in your design.

Impact Description	Impact Type	Stakeholder(s)	Mechanism(s)	Risk Level	Mitigation Tactics

### ***Part 3: Developed AI System***

3a) Provide evaluation information / results for the following:

- **Accuracy** (e.g., correct classifications / total classifications):
- **Precision and Recall** (especially important for imbalanced datasets; Precision: predicted positives correct; Recall: actual positives correct):
- **F1 Score** (harmonic mean of Precision and Recall, useful when both false positives and false negatives matter):
- **ROC-AUC or PR-AUC** (captures the model's ability to distinguish between classes at various thresholds):
- **Latency** (time from input to output, important for user experience):
- **Throughput** (number of processed inputs per unit of time, important for scalability):

3b) Provide results from additional testing or tailored metrics you identified as essential for your system's context (e.g., explanation quality measures, domain-specific Key Performance Indicators (KPIs), resource utilization, or error consistency):

3c) Update the Impact Analysis table from Part 2. List the mitigation tactics, human guardrails, and other risk management in place throughout your AI system to handle higher priority risks:

### ***Part 4: Deployed AI System***

4a) Describe how the AI system's real-world performance is being monitored (processes, KPIs, frequency, responsible parties):

4b) Describe who and approximately how many people are current users of the AI component or outputs of your system:

4c) Summarize any stakeholder feedback received and observed impacts:

4d) Update the Impact Analysis table from Part 2.

### B.4.1. Negative Impacts Glossary

The following negative impact categories (Types of Harm, 2023) are adapted and expanded to align with CMS contexts. These serve as examples, not an exhaustive list.

Negative Impact Type	Definition	Example
<b>OPPORTUNITY LOSS</b>	Reduced or denied access to essential services, benefits, or opportunities due to the AI system's outputs or decisions.	An AI scheduling tool excludes certain patients from specialist referrals, limiting timely access to care.
<b>ECONOMIC LOSS</b>	Financial negative impact to beneficiaries, providers, or CMS due to misinformed decisions, resource allocation, or reimbursements.	An AI-claims adjudication model unreasonably denies legitimate claims, causing financial strain on a provider.
<b>MANIPULATION</b>	Influencing stakeholder behavior or decisions through hallucinatory, misleading, or covertly persuasive outputs.	An AI-driven recommendation nudges providers to order unnecessary tests that increase costs without improving patient outcomes.
<b>SOCIAL HARM</b>	Any negative impact on social structures, relationships, or community trust that is influenced by the use of or outputs from AI	A model consistently making predictions based on statistical errors, eroding trust in services.
<b>DIGNITY LOSS</b>	Undermining an individual's sense of respect, autonomy, or worth, often by misrepresenting identities or using disrespectful language.	A person-facing chatbot uses dismissive language toward certain cultural groups, causing distress and disrespect.
<b>LIBERTY LOSS</b>	Restricting freedom of choice or autonomy, often through influencing decisions without transparency or limiting user options.	An AI-driven ranking system pressures providers to follow only certain treatment pathways, reducing clinical judgment.
<b>PRIVACY LOSS</b>	Exposure, unauthorized inference, or unintended disclosure of personal or sensitive information, reducing individual privacy.	An AI model inadvertently re-identifies patients in what was intended to be de-identified data sets.
<b>EMOTIONAL OR PSYCHOLOGICAL INJURY</b>	Causing distress, anxiety, or reduced well-being through negative impactful AI interactions, content, or decisions.	An AI-generated health prediction message creates undue alarm or confusion for beneficiaries without human follow-up.
<b>PHYSICAL INJURY</b>	Direct or indirect negative impact to an individual's physical health and safety influenced by AI outputs or decisions.	A diagnostic model's inaccurate recommendation leads to a negative impactful treatment choice if followed blindly.
<b>ENVIRONMENTAL IMPACT</b>	Negative ecological consequences related to AI's computational operations, data centers, or required resources.	High computational loads for model training cause increased energy consumption and carbon emissions.
<b>OTHER</b>	Any negative impact not captured above, as relevant to the system's unique context.	Regulatory non-compliance resulting from AI outputs, leading to legal consequences or public trust erosion.

### B.4.2. Mechanisms Glossary

These mechanisms describe how identified negative impacts may occur. Use the mechanism(s) that best fit your system's context. Add or adapt as needed.

- **Statistical Error:** Statistical and analytical errors in predictions or decisions resulting from unbalanced training data sets and/or poor data analysis capabilities of the AI algorithms.
- **Explainability/Transparency:** The clarity and interpretability of the AI's decision-making process. Poor explainability can hide negative impactful logic or errors.
- **Security:** Vulnerabilities that allow adversaries to manipulate the system or access sensitive data, potentially leading to negative impact.
- **Access:** Difficulties regarding a user's ability to gain authorization to, or digital or physical interface with, AI capabilities.
- **Privacy:** Risks related to data handling, including unauthorized disclosure of personal information or unethical data inference.
- **Reliability:** Lack of consistent performance or stability, leading to unpredictable negative impacts over time.
- **Accuracy:** Inaccurate predictions or classifications that may cause flawed decisions and related negative impacts.
- **Accountability:** Absence of clear responsibility or oversight structures can worsen or fail to prevent negative impacts.
- **Data Quality:** Poor or unrepresentative data leading to skewed outputs and unintended consequences.
- **Interoperability:** Inability to integrate with other systems safely and effectively, potentially causing cascading negative impacts.
- **OTHER:** Provide details if a unique or project-specific mechanism induces negative impact.

An additional set of simplified mechanisms can be found in this journal by Brown, S., Davidovic, J., & Hasan, A. (2021): [\*The algorithm audit: Scoring the algorithms that score us.\*](#)

### B.5. AI Threat Model Template

This template is usually paired with a Mural Diagram ([Threat Modeling Mural Template](#)) to help create *Data Flow Diagrams* (DFDs) to facilitate threat models. You are welcome to use other drawing tools for DFDs. For more information on DFDs and *STRIDE Threat Modeling Methodology*, see [Getting Started With Threat Modeling](#).

#### System

**Name:**

**Description:**

**Date:**

Data Type	Data Sensitivity	Workflow/Use case
PII	High	Customer Identification

---

#### Attendees:

- <List who attended any threat model sessions>

#### Session Information:

**Mural Team Room Link:**

**Recordings:**

**Diagram(s):**

**Notes**

- <Notes, questions, etc. about the system - identify potential threats>

#### Threats and Mitigations Table

ID	Threat Description	STRIDE Property(ies) - if applicable	Mitigation(s)	Action Item(s)	Notes
1	Sample Spoofing Threat	Spoofing, Repudiation	1. Add authentication controls and logging of successes and failures	1. Review current controls 2. Test / verify controls work as expected	1. Sample Notes

#### Parking Lot

- <A useful area for other notes and observations>

### B.5.1. Add-ons

#### *Data Flow Diagrams*

DFDs are images that represent how data moves through a system. These are not rote architecture diagrams but can occasionally reference aspects have implemented in the architecture.

Name	External Interactor	Process	Data Store	Dataflow	Trust Boundary
Icon	External Interactor(s)	Process (FE/APP)	Data Store (Name & Type)	Dataflow	Trust Boundary
Purpose	People, Other System(s), Workflows	Applications, Component Services, Team-owned	Database(s), Queues, Artifact Storage	Network Traffic, Data Manipulation	Access or Validation Boundaries

#### *Four-Question Frame for Threat Modeling*

The Four Question Frame for Threat Modeling is a set of questions to help teams build better systems.

1. What are we working on?
2. What could go wrong?
3. What are we going to do about it?
4. Did we do a good enough job?

#### *STRIDE Threat Modeling Methodology*

STRIDE (see below) is a mnemonic to help remember the type of common threats to protect against. Other mnemonics, like CIA (Confidentiality, Integrity, Availability) and AAA (Authorization, Authentication, Auditing) will also be used. These help identify functional threats early and with ubiquitous language.

Threat	Property Violated	Threat Definition
Spoofing	Authentication	Pretending to be something or someone other than yourself
Tampering	Integrity	Modifying something on disk, network, memory, or elsewhere
Repudiation	Non-Repudiation	Claiming that you didn't do something or were not responsible; can be honest or false
Information Disclosure	Confidentiality	Providing information to someone not authorized to access it
Denial of Service	Availability	Exhausting resources needed to provide service
Elevation of Privilege	Authorization	Allowing someone to do something they are not authorized to do

## CMS Artificial Intelligence Playbook

### STRIDE Application Matrix

The diagram below shows how certain aspects of STRIDE do or do not apply to certain elements.

External Entity	Spoofing	Tampering	Repudiation	Information Disclosure	Denial of Service	Escalation of Privileges
Process	Spoofing	Tampering	Repudiation	Information Disclosure	Denial of Service	Escalation of Privileges
Data Store Name and Type	Spoofing	Tampering	Repudiation	Information Disclosure	Denial of Service	Escalation of Privileges
→ Dataflow	Spoofing	Tampering	Repudiation	Information Disclosure	Denial of Service	Escalation of Privileges

I've got an Artifact, Now What? Create Sprint Tasks

Three-Part User Story Format

As an audience member, I would like [a mitigation description] so that I may avoid [a threat description]. I need [action items #1 and #2] and spike on [question #1] to do this.

Gherkin Format

**Feature:** Mitigation Description

**Scenario:** threat actor performs threat description

**When:** audience performs threat description

**And:** ...

**Then:** the mitigation effect

**And:** additional validation of mitigation effect

## B.6. Evaluation Labels Methodology

Key Metric	Description	Calculation	Score	Supporting Links
Performance	The model's ability in tasks like following instructions, answering science questions, and reasoning	Performance = MEAN(IFEval Raw + BBH Raw + GPQA Raw + MMLU-PRO Raw)		Find raw Instruction-following Evaluation (IFEval), BIG-Bench Hard (BBH), Graduate-Level Google-Proof Q&A Benchmark (GPQA), and Multitask Language Understanding (MMLU) - PRO scores at <a href="#">Hugging Face Open LLM Benchmark</a>
Efficiency	The cost (in \$) of running the model	Efficiency = $100 \cdot e^{(-100 \cdot (\text{dollar price per 1000 input tokens} + \text{dollar price per 1000 output tokens}))}$		Use price per 1,000 input and output tokens from <a href="#">Amazon Bedrock</a> .
Transparency	The level of public information available about the model's training process, fine-tuning capabilities, and output generation	+50 points for being open source. +20 points for publishing architecture details (partial points allowed). +10 points for ability to control parameters such as temperature, max_k, system prompts, and max output tokens (2.5 points for each control feature). +10 points for token/training data transparency (partial points allowed). +10 points for ability to fine-tune weights of the model.		Check <a href="#">model cards on Hugging Face</a> as a resource for architecture details, training data, and other information for many open-source models.
Privacy	Considerations for safety of data use and data access in sensitive contexts	100 points total if the model is self-hosted and open-source, or 75 points total if the model is self-hosted but closed-source, or 50 points total if the model is available via a FedRAMP-authorized cloud provider, or 0 points given otherwise; e.g., the model is only available via API from a non-FedRAMP provider		Use the official <a href="#">FedRAMP Marketplace</a> database to check for FedRAMP authorization.

## CMS Artificial Intelligence Playbook

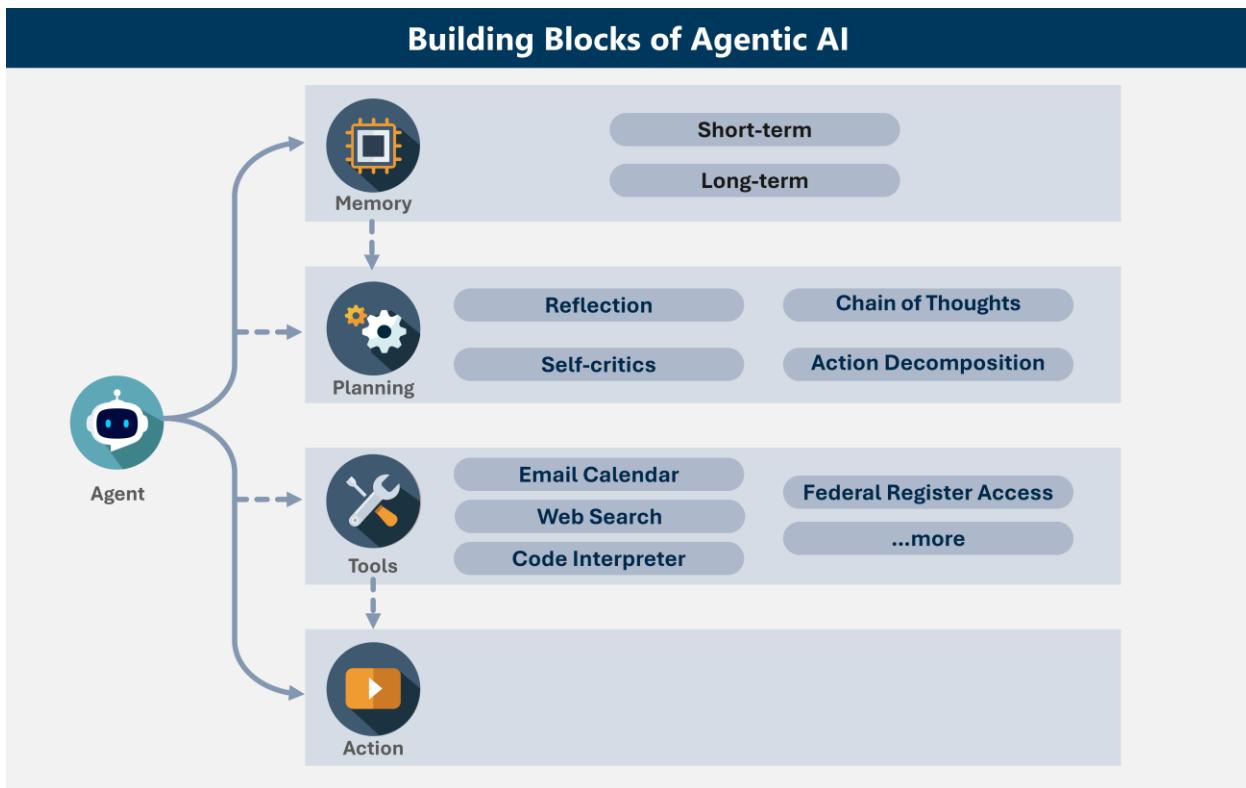
Key Metric	Description	Calculation	Score	Supporting Links
Compute Efficiency	How effectively a model utilizes available hardware, particularly focusing on GPU memory consumption	Total memory footprint required to process 1000 queries at a given precision (in GB) divided by 4 to represent a 4GB chip <b>0 - 2 chips</b> = High Efficiency ( <i>0 - 8GB per 1000 queries</i> ) <b>3 - 5 chips</b> = Moderate Efficiency ( <i>12 - 20GB per 1000 queries</i> ) <b>6 - 9 chips</b> = Low Efficiency ( <i>24 - 36GB per 1000 queries</i> ) <b>10+ chips</b> = Very Low Efficiency ( <i>40GB+ per 1000 queries</i> )		Identify memory usage using <a href="#">Ollama</a> , or estimate using price ratios between two models on a service such as <a href="#">Amazon Bedrock</a> .

### B.7. LLM Implementation Checklist

Use Case Name		
Team		
Use Case Description	Stakeholders	
Action	Implication	Team Notes / Rationale
<input type="checkbox"/> Determine the specific task or domain where the LLM will be applied.	Identifying the task or domain will guide the customization of the LLM to meet specific needs and improve its effectiveness (e.g. If an LLM serves scripted customer service, it can be less complex than one tackling Massive MMLU benchmark problems).	
<input type="checkbox"/> Apply the CMS HCAI Matrix (Artz, 2024) to guide the development of the project.	Ensures that all stakeholders are considered at every phase of the project, enhancing the relevance and impact of the AI solution on actual user needs.	
<input type="checkbox"/> Identify the hardware available to you.	The hardware limits the model size you can use; low VRAM supports only smaller LLMs.	
<input type="checkbox"/> Consider whether the task is time-sensitive, accuracy-dependent, or cost-conscious.	Understanding this helps balance the trade-offs between performance, cost, and speed, tailoring the LLM deployment to your priorities.	
<input type="checkbox"/> Consider RAG if the LLM's responses need to come from document sources.	RAG ensures the LLM pulls accurate information from relevant documents (Artz, 2024). This lowers the risk of hallucinations.	

Action	Implication	Team Notes / Rationale
<input type="checkbox"/> Make benchmarks relevant to the specific domain to assess how the LLM will perform in real-world applications	Creating or using domain-specific benchmarks can help predict the LLM's real-world effectiveness and guide further tuning.	
<input type="checkbox"/> Prioritize metrics based on the application.	Prioritizing metrics helps the team focus efforts, such as emphasizing groundedness over relevance when using ungrounded information could be harmful.	
<input type="checkbox"/> Plan for a pilot phase where the LLM is tested in a controlled environment.	This allows for the testing of the LLM under controlled conditions, providing a chance to address issues before full-scale deployment.	

## B.8. Building Blocks of Agentic AI



## Appendix C. Acronyms

Term	Full Form
<b>AAA</b>	Authorization, Authentication, Auditing
<b>ADKAR</b>	Awareness, Desire, Knowledge, Ability, Reinforcement
<b>AGI</b>	Artificial General Intelligence
<b>AI</b>	Artificial Intelligence
<b>AI CCI</b>	AI Cross-Cutting Initiative
<b>AIRB</b>	AI Review Board
<b>AIRC</b>	AI Review Committee
<b>ARIA</b>	Algorithmic Risk and Impact Assessment
<b>ASI</b>	Artificial Superintelligence
<b>AUC</b>	Area Under Curve
<b>BBH</b>	BIG-Bench Hard
<b>BO</b>	Business Owner
<b>CAIO</b>	Chief AI Officer
<b>CCI</b>	Cross-Cutting Initiative
<b>CCIO</b>	Consumer Information & Insurance Oversight
<b>CCSQ</b>	Center for Clinical Standards & Quality
<b>CIA</b>	Confidentiality, Integrity, Availability
<b>CISO</b>	Chief Information Security Officer
<b>CLAW</b>	CMS Labor Analysis Wizard
<b>CM</b>	Center for Medicare
<b>CMCS</b>	Center for Medicaid & CHIP Services
<b>CMMI</b>	Center for Medicare & Medicaid Innovation
<b>CMS</b>	Centers for Medicare & Medicaid Services
<b>COTS</b>	Commercial off-the-shelf
<b>CPI</b>	Center for Program Integrity
<b>DevSecOps</b>	Development Security Operations
<b>DFD</b>	Data Flow Diagram
<b>EDA</b>	Exploratory Data Analysis
<b>EMR</b>	Electronic Medical Records
<b>EPRO</b>	Emergency Preparedness & Response Operations
<b>FISMA</b>	Federal Information Security Management Act
<b>GPQA</b>	Graduate-Level Google-Proof Q&A Benchmark
<b>GSA</b>	General Services Administration
<b>GPU</b>	Graphics Processing Unit
<b>HAX</b>	Human-AI Experience

## CMS Artificial Intelligence Playbook

---

Term	Full Form
<b>HCAI</b>	Human-Centered AI
<b>HCD</b>	Human-Centered Design
<b>HHS</b>	Department of Health and Human Services
<b>HIPAA</b>	Health Insurance Portability and Accountability Act
<b>IDR</b>	Integrated Data Repository
<b>IFEval</b>	Instruction-following Evaluation
<b>IoT</b>	Internet of Things
<b>ISSO</b>	Information System Security Officer
<b>IT</b>	Information Technology
<b>KPIs</b>	Key Performance Indicators
<b>LLM</b>	Large Language Models
<b>ML</b>	Machine Learning
<b>MLTRL</b>	Machine Learning Technology Readiness Level
<b>MMLU</b>	Massive Multitask Language Understanding
<b>MVP</b>	Minimal Viable Product
<b>NIH</b>	National Institutes of Health
<b>NIST</b>	National Institute of Standards and Technology
<b>NLP</b>	Natural Language Processing
<b>NSF</b>	National Science Foundation
<b>OAGM</b>	Office of Acquisition and Grants Management
<b>OC</b>	Office of Communications
<b>OHEI</b>	Office of Healthcare Experience & Interoperability
<b>OHI</b>	Office of Hearings & Inquiries
<b>OIT</b>	Office of Information Technology
<b>OMB</b>	Office of Management and Budget
<b>OMM</b>	Organizational Maturity Model
<b>OPOLE</b>	Office of Program Operations & Local Engagement
<b>OSPO</b>	Open Source Program Office
<b>PHI</b>	Protected Health Information
<b>PII</b>	Personally Identifiable Information
<b>PoC</b>	Proof of Concept
<b>QASP</b>	Quality Assurance Surveillance Plan
<b>RAG</b>	Retrieval-Augmented Generation
<b>RAI</b>	Responsible AI
<b>RCS</b>	Revision Control System
<b>ROC</b>	Receiver Operating Characteristic
<b>ROI</b>	return-on-investment
<b>SCCS</b>	Source Code Control System

## CMS Artificial Intelligence Playbook

---

Term	Full Form
<b>SMEs</b>	Subject Matter Experts
<b>STRIDE</b>	Spoofing, Tampering, Repudiation, Information Disclosure, Denial of Service, Elevation of Privilege
<b>SVN</b>	Subversion
<b>TAM</b>	Technology Acceptance Model
<b>TRB</b>	Technical Review Board
<b>TRL</b>	Technology Readiness Level
<b>TFVC</b>	Team Foundation Version Control
<b>UI</b>	User interface
<b>UX</b>	User experience
<b>VRAM</b>	Video Random Access Memory
<b>WCAG</b>	Web Content Accessibility Guidelines
<b>WR</b>	Workforce Resilience