



An Evaluation Roadmap for a More Effective Government

American Evaluation Association

May 2026 Update

An Evaluation Roadmap for a More Effective Government

Evaluation¹ is an essential function of good governance. It can be employed to: improve the effectiveness² and efficiency of programs³ and services, enhance the oversight and accountability⁴ of U.S. and Tribal governments,⁵ assess the extent to which programs or policies⁶ are working, and provide critical information needed for making difficult decisions about them.

In keeping with its mission,⁷ the American Evaluation Association (AEA) describes its vision of the role of evaluation in government through this document. The *Evaluation Roadmap for a More Effective Government* (hereafter the *Evaluation Roadmap*) provides guidance for improving good governance through evaluation, offering strategies and principles to strengthen the practice of evaluation throughout the life cycle of policies and programs.

The *Evaluation Roadmap* provides:

1. Why evaluation matters for effective government.
2. A framework for use in developing and implementing evaluation programs in U.S. and Tribal governments.
3. A set of principles to guide the implementation of the framework.
4. A description of key federal laws and policies affecting evaluation in government.

Background and History of the Evaluation Roadmap

Established in 1986, the AEA is a professional association of evaluators devoted to the application and exploration of approaches to evaluate programs, policies, and activities. Evaluation involves assessing the strengths and weaknesses of programs, policies, personnel, products, and organizations to improve their effectiveness. The AEA has thousands of members representing all 50 states in the United States and more than 80 countries.

The EPTF is a standing advisory body that provides current insights and recommended actions to the AEA Board of Directors on matters of evaluation policy and practice⁸ across the United States within federal, state, and local government; Tribal; and philanthropic contexts.

Established in September 2007, the EPTF serves as a strategic thought partner to the Board, helping AEA interpret, anticipate, and respond to developments in evaluation-specific policies that affect the field and its members. In 2009, AEA's Evaluation Policy Task Force (EPTF)⁹ developed the first *Evaluation Roadmap* to share the lessons learned in public agencies that

¹ See Glossary: "Evaluation."

² See Glossary: "Effectiveness."

³ See Glossary: "Program."

⁴ See Glossary: "Accountability."

⁵ See Glossary: "U.S. and Tribal governments."

⁶ See Glossary: "Policy."

⁷ American Evaluation Association. "AEA Mission, Vision, and Values." <https://www.eval.org/About/About-AEA/Mission-Vision-Values>.

⁸ See Glossary: "Evaluation policy."

⁹ American Evaluation Association. "The Evaluation Policy Task Force." <https://www.eval.org/Evaluation-Policy/Evaluation-Policy-Task-Force>.

have applied evaluation. It has been used by several organizations in shaping their evaluation policies. This update to the *Evaluation Roadmap* offers a framework to guide organizations in U.S. and Tribal governments as they strive to ensure consistency in standards and expectations on producing and using evaluation and evidence¹⁰ to inform policymaking.

Using this Roadmap

Consistent with the AEA's mission, this document describes AEA's vision of the role for evaluation in government. This *Evaluation Roadmap* provides guidance to build government capacity for strengthening the practice of evaluation, and should be adapted to context as needed.

The U.S. federal government, state and local governments, Tribal governments, and U.S. Territories continue to address major challenges to improve lives, protect the planet, and create efficiencies. With adequate resources; more thoughtful and systematic integration of evaluation into the planning, management, and oversight of programs; and increased application of evaluation results to planning and decision-making, we can improve the performance of today's programs and ensure that tomorrow's programs reap the lessons learned today. Institutionalizing evaluation can also help achieve a more accountable, open, and democratic system of governance for future generations.

The standards and guidance provided in this *Evaluation Roadmap* are intended to support good governance at any level of capacity for public agencies, organizations, and evaluation offices. By engaging with this *Roadmap*, it is hoped that public decision-makers will be able to understand the importance of evaluation as an essential function of good government, assess the maturity of the evaluation functions in their organizations, and successfully strengthen the practice of evaluation throughout their policies and programs.

The principles offered here are applicable and essential when setting up a new office, or when strengthening an existing office, regardless of its current capacity. To that end, the *Roadmap* promotes trust in the integrity and credibility of the evaluation team to implement the guidance for practice offered in this document. Trust-building is a key skill needed for evaluators that facilitates all of evaluation practice.

¹⁰ See Glossary: "Evidence."

Table of Contents

An Evaluation Roadmap for a More Effective Government	2
Background and History of the Evaluation Roadmap	2
Using this Roadmap	3
Table of Contents	4
The Value Proposition for Evaluation in Government	5
Standards for Evaluation Practice	7
RELEVANCE AND UTILITY	7
RIGOR	9
INDEPENDENCE	12
TRANSPARENCY	13
ETHICS AND INCLUSIVITY	14
Key Laws and Policies Affecting Evaluation in Government	17
Executive Branch Role	18
Legislative Branch Role	19
Collaboration Between Executive and Legislative Branches	21
Tribal Governments and Evaluation	21
Looking to the Future	24
Appendix: Glossary	26
Citations	31

The Value Proposition for Evaluation in Government

U.S. and Tribal governments, across all levels, consistently face an array of challenges in energy, the environment, health care, education, and the economy. Government decision-makers are responsible for identifying and understanding problems, and then developing solutions by designing programs, regulations, policies, activities, systems,¹¹ or initiatives to mitigate or resolve problems. To determine the merit, quality, appropriateness, and usefulness of interventions, credible systematic information is needed about the program or policy in question. This information includes knowledge about operations, what has been achieved, and at what cost. Such information is crucial if government officials want evidence to demonstrate that the chosen solutions are working, taxpayers' money is spent wisely, and governments are accountable to the public for producing results.

Evaluation involves assessing the strengths and weaknesses of programs, policies, and organizations to improve their effectiveness, efficiency, and worth. Evaluators employ systematic data collection and analysis to address questions about how well U.S. and Tribal government programs and policies are working, and whether they are achieving their objectives. Evaluation provides a useful and important tool to bring credible, well-grounded evidence to bear on a broad range of government decisions. It supports the development of new knowledge, innovation, and organizational learning in both the executive and legislative branches of government at every level.

Evidence from evaluations can be used to compare alternative programs or policy options, guide program development and decision-making, and reveal effective practices. More fundamentally, evaluation can contribute the evidence needed to inform program managers and policy makers at all levels of government. The courts, as well, may cite evaluation findings as evidence in their judgments. Evaluation can be used to enhance congressional oversight and executive accountability and to supply publicly accessible information that is at the heart of transparency and open government. In addition, evaluation can address contextual appropriateness, providing evidence that can help ensure equitable practices and resources, as well as appropriateness of programs across diverse settings.

Evaluation is an integral feature of good government, whether the goal is better performance, stronger oversight and accountability, or more data-informed, evidence-based and innovative decision making. U.S. and Tribal governments can benefit significantly from using relevant data from evaluations as evidence to inform program and policy planning, monitor program performance, and inform major decisions about program reform, expansion or termination.

In order to fully realize the potential benefits evaluation has to offer in pursuing a more effective government, executives, program managers, and other staff in U.S. and Tribal governments are faced with the challenge of building and effectively leveraging and enhancing evaluation

¹¹ See Glossary: "Systems."

capacity.¹² The key to getting value from evaluation is to ensure it is integral to managing government policies and programs at all stages, from planning and initial development through launch, ongoing implementation, appropriations, and reauthorization. **In short, evaluation is an essential function included in all U.S. and Tribal organizations.**

Exhibit 1. Examples of Possible Uses of Evaluation

- Develop and address questions about current and emerging problems
 - Assess the needs within communities for public action
 - Develop strategic plans and action plans, and set criteria for assessing success
- Improve programs and policies in a systematic manner
 - Articulate and test assumptions about the relevance, purpose, and impact of policies or programs
 - Identify program implementation and outcome failures and successes
 - Re-examine program relevance and effectiveness over time
 - Inform the development of new programs, where needed
 - Provide timely feedback to decision-makers, enabling them to make changes to programs and policies when needed
- Increase government accountability and transparency
 - Identify ways to reduce waste and enhance efficiency
 - Identify innovative solutions that work and in what contexts
 - Share information about effective practices across government programs and agencies
- Assess issues of equal access and equitable treatment across diverse communities
 - Examine the requirements for the transfer or tailoring of promising programs to new sites or cultural contexts

¹² See Glossary: "Evaluation capacity."

Standards for Evaluation Practice

We recommend that U.S. and Tribal governments broadly adhere to the following five evaluation standards:

1. **Relevance and Utility**
2. **Rigor**
3. **Independence and Objectivity**
4. **Transparency**
5. **Ethics and Inclusivity**

These standards, accompanied by guidance and resources for implementation and practice, provide a framework for building and maintaining high-quality evaluation capacity while allowing for appropriate flexibility in meeting organizational needs. Each standard requires the integration of all the others; at times, adherence to one standard must be judiciously balanced with adherence to others.

This *Roadmap* builds upon the standards developed and adopted by the federal government as authorized by the Foundations for Evidence-Based Policy Act of 2018 (Evidence Act).¹³ Under the Evidence Act, the government was charged with establishing “evaluation standards to guide agencies in developing and implementing evaluation activities, evaluation policies, and in hiring and retaining qualified staff.”¹⁴ The *Roadmap* adds to these standards in several ways: by emphasizing inclusivity alongside ethics, considering the role of emerging technologies such as artificial intelligence, and elevating culturally responsive evaluation methods, among other additions.

RELEVANCE AND UTILITY

Evaluations must address questions of importance and serve the information needs of interestholders¹⁵ in order to be useful resources. This standard requires that evaluations present findings in ways that are understandable, actionable, and available in a timely manner for use in agency activities.

To be relevant, evaluations should address important issues within, and sometimes across, agencies. Organizations should develop learning agendas¹⁶ and other documents that reflect their long- and short-term goals, link to their strategic plans, and reflect their missions. In these documents, they should identify priority evaluation questions and develop plans for evaluation that contribute valuable evidence and data to inform budget and policy decisions, program design and development, and regulatory actions. The evaluation questions should

¹³ Foundations for Evidence-Based Policymaking Act of 2018, Pub. L. No. 115-435, 132 Stat. 5529 (2019). <https://www.congress.gov/115/plaws/publ435/PLAW-115publ435.pdf>.

¹⁴ Office of Management and Budget. *Phase 4 Implementation of the Foundations for Evidence-Based Policymaking Act of 2018: Program Evaluation Standards and Practices*. Memorandum M-20-12. Washington, DC: Executive Office of the President, 2020. <https://www.whitehouse.gov/wp-content/uploads/2020/03/m-20-12.pdf>.

¹⁵ See Glossary: "Interestholder."

¹⁶ See Glossary: "Learning agenda."

focus on important agency issues, such as considerations of efficiency, effectiveness, improvement, learning, or accountability.

To promote the use of findings, evaluators should center interestholder engagement throughout the evaluation process, considering the cultural context of the programs and participants. What counts as a relevant and important question, a meaningful outcome, or a useful finding depends on whose perspectives are included in making those determinations. Evaluators should therefore engage a broad range of interestholders, including program participants and affected communities, in defining evaluation questions and interpreting findings, not only in reviewing results after the fact. This helps ensure that evaluation designs and methods are suited to the questions that matter most to those with an interest in the program, and facilitates use of findings across a wider range of decision-makers. Useful evaluations produce findings within a timeline that is appropriate to the questions under consideration, and they use dissemination strategies that enable relevant interestholders to know of and understand the findings in clear, concise, and actionable ways.

Guidance for Practice: Relevance and Utility

Leadership and Infrastructure

- Assign experienced, senior evaluation officials and managers to administer evaluation centers or coordinate evaluation functions at influential levels in U.S. and Tribal governments.
- Ensure that evaluation units and staff receive support from top leadership, including time to participate in planning and implementing evaluations, to reflect with others on evaluation findings, and to develop clear ways to move results into meaningful change.
- Provide sufficient, stable funding to support evaluations and professional capacity building.
- Coordinate and communicate about evaluation efforts across public organizations with overlapping or complementary missions.

Scope and Coverage

- Conduct evaluations of public operations, programs, and policies throughout their life cycles and use evaluation both to improve programs and to assess their effectiveness.
- Evaluate services, programs, and policies in a manner that is appropriate for program stewardship and useful for decision-making.
- Build into each new program and major policy initiative an appropriate theory of change¹⁷ to guide and assess the program or initiative at key points throughout its life.
- For existing programs and policies, synthesize prior evaluations and other evidence to assess what is already known and develop evaluation plans¹⁸ to inform future decision-making.

¹⁷ See Glossary: "Theory of change"

¹⁸ See Glossary: "Evaluation plans."

Evaluation Planning

- Prepare annual and long-term evaluation plans to ensure timely insights are available from evaluations to guide decision-making about initiatives, programs, and policies at all stages.
- Ensure evaluation planning takes into account the need for relevant data to inform program budgeting, reauthorization, agency strategic plans, ongoing program and regulation development and management, and responses to critical issues concerning program effectiveness, efficiency, and waste.
- Include an appropriate mix of short- and long-term studies to produce results of appropriate scope and rigor¹⁹ for short- or long-term policy or management decisions.
- Develop plans in consultation with diverse program interestholders, while allowing flexibility for evaluation questions that arise unexpectedly and urgently.

Tracking and Use

- Track and re-examine the use made of evaluation findings over time to determine if data were used by policymakers, program managers, and other interestholders. If so, how? If not, why not?
- Assess whether the program and its evaluation contributed over the longer term to resolving the public problem addressed.
- Share findings and recommendations from tracking the use of evaluation data with professionals and the public to support inclusivity in decision-making and promote continuous improvement.

RIGOR

Evaluators draw on an extensive array of analytical approaches and methods, including, but not limited to, those in **Exhibit 2**. Evaluation as a field is very interdisciplinary, encompassing many areas of expertise. For evaluations to produce credible findings that agencies and their interestholders can confidently rely upon, **they must be undertaken by qualified evaluators with relevant education, skills, and experience for the methods undertaken**. The choice of design and methods highly depends on what is most appropriate to answer the key questions at hand, balanced with the evaluation's goals, scale, timeline, feasibility, and available resources. Often, several analytic methods are needed in a single evaluation.

Programs and their environments are complex; no single analytic approach or method can decipher the complexities in the program environment. Furthermore, as a program matures, the right method or combination of methods may change. Some evaluation approaches are particularly helpful in a program's early developmental stages, whereas others are more suited to ongoing, fully implemented programs or to analysis of temporary programs upon their completion. How the evaluation is to be used in decision-making also can influence which approach or combination of approaches is best. When information is needed quickly, studies that can use existing data or rapid data-collection methods make the most sense. On

¹⁹ See Glossary: "Rigor."

the other hand, when the dynamics of the program and the behavior of its beneficiaries must be understood, then more sophisticated long-term studies may make more sense.

Regardless of the approaches taken, rigorous evaluations should be grounded in a theory of change and take into account existing relevant data and gaps in data. Furthermore, all evaluation methods should be context-sensitive, have cultural relevance, and adhere to widely accepted scientific principles. When planning evaluations, evaluators should solicit input and feedback from interestholders, technical working groups, and other consultation from independent, unbiased experts as possible.

Once data collection and analyses are complete, **transparency in documenting findings, methods, and developing reports are essential principles of good and rigorous evaluation.** Evaluators should document the populations, settings, or circumstances the evaluation examined and to which findings can be generalized; use precise language to characterize findings accurately based on the design and methods; and disclose limitations, assumptions, methods undertaken, data and sampling frames, and justifications for any changes in design and methods from the initial plan.

Guidance for Practice: Rigor

Professional Competence

- Ensure evaluation staff and contractors have the needed knowledge and skills, including technical competencies and cultural competence, for the scope of their work, according to recognized guiding principles and evaluator competencies from the AEA.^{20,21} The framework for assessing evaluator's experience and training was established in 2018 for use by its membership.
- Identify and leverage opportunities for expert staff to maintain and continue to build their skills and specializations relevant to evaluation through professional associations, conferences, and trainings.

Evaluation Policies and Procedures

- Develop written evaluation policies and procedures across and within public organizations that can guide evaluation efforts and help ensure their quality and appropriateness, consistent with the guiding principles for evaluators from the AEA and federal guidance pursuant to the Evidence Act.²²
- Ensure evaluation policies identify the criteria and administrative steps for developing evaluation plans and setting priorities, including:
 - selecting evaluation approaches and methods,
 - consulting subject matter experts,

²⁰ American Evaluation Association. "Guiding Principles for Evaluators." 2018. <https://www.eval.org/About/Guiding-Principles>.

²¹ American Evaluation Association. "Competencies & Standards." 2018. <https://www.eval.org/About/Competencies-Standards>.

²² Office of Management and Budget, *Phase 4 Implementation*.

- ensuring evaluation product quality, publishing evaluation reports,
- ensuring independence²³ of the evaluation function,
- using an appropriate mix of staff and outside consultants and contractors,
- including diverse representation of staff and interestholders, and
- promoting professional development of staff.

Technology and Methods

- Remain abreast of rapidly developing technologies, such as artificial intelligence and big data systems, that affect evaluation activities and continuously assess and reassess implications for program evaluations²⁴ and evaluation policies, communications, and analytic approaches.
- Incorporate into evaluation policies a set of standards for adopting and using new technological tools and methods, including artificial intelligence and big data analytics.
- Leverage the availability of big data sets to capture data for analyzing past trends, applying statistical models and algorithms, and delineating subgroup characteristics and outcomes.
- Use AI and generative AI tools appropriately and transparently for literature reviews, research summaries, meta-analyses, and analysis of qualitative information, while maintaining quality standards.
- Ensure evaluation policies and plans clearly indicate how AI tools and big data will, or were, be used to address key questions of interest.

Exhibit 2. Examples of Tools and Methods Used in Evaluation

- Case studies
- Cost-benefit and cost-effectiveness analyses
- Culturally and contextually responsive methods
- Implementation analyses
- Evaluation synthesis and meta-analysis
- Experimental designs and randomized control trials (RCTs)
- Logic models²⁵ and evaluability assessments
- Needs assessments
- Process tracing
- Surveys and interviews
- Systems and complexity-informed evaluation (SCIE) methods²⁶
- Quasi-experimental designs

²³ See Glossary: "Independence."

²⁴ See Glossary: "Program Evaluation."

²⁵ See Glossary: "Logic model."

²⁶ Margaret Hargreaves, Brandon Coffee-Borden, and Chandria Jones, "Evaluating Systems Change: Systemic Change Measurement Rubric," presentation, NORC at the University of Chicago, April 2026. <https://www.norc.org/content/dam/norc-org/pdf2026/systems-change-rubric-presentation-april-2026.pdf>.

INDEPENDENCE

Evaluations must be viewed as credible and legitimate in order for interestholders, experts, and the public to accept their findings. Acceptance depends on the perceived independence and trustworthiness of the evaluators.

Agencies should enable evaluators to, and evaluators should, operate with an appropriate level of independence from programmatic, regulatory, policymaking, and interestholder activities. While interestholders have an important role in identifying evaluation priorities, evaluators should be appropriately insulated from political and other undue influences that may affect their impartiality and professional judgement.

Evaluators should strive for transparency in the planning and conduct of evaluations and in the interpretation and dissemination of findings, avoiding conflicts of interest, bias, and other partiality. Credible information is impartial, unbiased, clear, and complete. Lack of credibility compromises an evaluation's findings and uses. Evaluators should acknowledge and attempt to mitigate conflicts of interest or bias and other partiality that may be introduced in how they frame evaluation questions, collect or analyze data, or interpret findings.

Independence must be balanced with appropriate interestholder engagement. Consultation with internal and external interestholders is generally desirable on such matters as the questions to be addressed in an evaluation and the reporting schedule for study results. The leadership of public organizations can, and in many cases should, play a role in establishing general evaluation agendas, budgets, schedules, and priorities. At the same time, the real and perceived independence of the evaluation office must be preserved. Especially for evaluations performed in the service of public accountability, the evaluation office must retain control of the evaluation's questions, design and methods, and most importantly, its results and distribution.

Guidance for Practice: Independence

Safeguarding Independence

- Establish and maintain policies and procedures to ensure that evaluation offices and staff have the authority to approve an evaluation's design and methods and release evaluation findings to preserve political independence.
- Safeguard the independence of evaluations with respect to study design, conduct, results, and recommendations, while allowing for an appropriate level of consultation with and input from organization staff personnel and other interested parties.

Interestholder Engagement

- Consult with internal and external interestholders on matters such as the questions to be addressed in an evaluation and the reporting schedule for study results, while preserving evaluator control over design, methods, and findings.
- Recognize that the degree of shared decision-making, and in some cases program staff participation in the conduct of an evaluation, may be greater for evaluations aimed at program improvement than for evaluations serving accountability purposes.

- Engage key interestholders in interpreting implications of findings and providing important context to help evaluators understand and meaningfully interpret findings, while safeguarding independence from undue influence and potential bias.

Tribal Consultation

- Engage relevant interestholders in Tribal governments in careful planning regarding evaluation, recognizing that consultations will vary across different policy arenas.
- Ensure evaluators working with Tribal communities possess relevant expertise, and understand and respect the unique legal frameworks, governance structures, and data sovereignty requirements of Tribal nations.

Evaluation Policies and Procedures

- Establish a routine basis for reviewing and updating evaluation policies and procedures as needed to remain current with technological developments while retaining key evaluation principles such as privacy, transparency and independence.
- Develop and execute policies and procedures that maintain a balance between incorporating important interestholder feedback in evaluation reports and the need to release findings in a timely manner.

TRANSPARENCY

Evaluation should be transparent in the planning, implementation, and reporting phases to enable accountability and help ensure that aspects of an evaluation are not tailored to generate specific findings. By following best practices of open science, evaluators can support credibility, trust, and the ongoing building of systemic knowledge about effective government programs.

Decisions about the evaluation's purpose and objectives should be clearly documented before conducting the evaluation, including the range of interestholders who will have access to details of the work and findings, the design and methods, and the timeline and strategy for releasing findings. These decisions should take into consideration any legal, ethical, national security, or other constraints for disclosing information. There should be a commitment established in an agency's evaluation policies to releasing all findings, including favorable, null, and unfavorable findings, in a timely manner to relevant interestholders.

After an evaluation is completed, the implemented design, methods, and findings should be reported in sufficient detail such that readers can determine whether the evaluation meets standards for rigor and independence, assess the credibility and objectivity of the findings, and replicate or reproduce the work. Whenever possible, the data collected during an evaluation should be made available to support replication and secondary use, consistent with applicable laws, regulations, and policies to ensure proper protection of privacy and data integrity.

Guidance for Practice: Transparency

Open Science Practices

- Release evaluation methods and results in a broadly accessible, plain language manner to the public and potential evaluation users, as feasible, with sufficient detail for transparency and reproducibility.
- Follow best practices of open science in disclosing evaluation methods, including making design and analysis plans, results, data, and analytic files and code available publicly and in a timely manner, except where inconsistent with relevant laws such as the Freedom of Information Act, Privacy Act, national security considerations, non-disclosure agreements, or other applicable requirements.
- Ensure results are broadly accessible through the Internet and with sufficient detail for reproducibility of findings and future potential design replication.

Strategic Communication of Findings

- Consult closely with internal and external interestholders in identifying program and policy objectives, critical operations, and definitions of success.
- Communicate evaluation findings and methods to policymakers, program managers, and the public in contextually appropriate ways.
- Create accessible information systems for publicly communicating and sharing evaluation findings about effective and ineffective program practices or policies.
- Tailor reports and briefings on evaluation findings and methods, as well as their limitations, so that various interestholder audiences can understand, interpret, and assess the validity, reliability, and credibility of the work.

Technological Transparency

- Ensure that evaluators and their reports and products clearly acknowledge use of AI, where AI tools have been used in analysis and in report preparation, and which AI tools were used.
- Develop standards for AI acknowledgment consistent with emerging policies in academic science and research units, journals, and research organizations.
- Maintain standards for quality and transparency when disseminating and communicating evaluation results, including when evaluation reports are placed into information systems and clearinghouses.

ETHICS AND INCLUSIVITY

Evaluations must be conducted to the highest ethical standards to protect the public and maintain public trust in the government's efforts. Evaluations should be planned and implemented to safeguard the dignity, rights, safety, and privacy of participants and other interestholders and affected entities. Evaluators should abide by current professional standards pertaining to treatment of participants, including informed consent,²⁷ confidentiality, and maximizing benefits while minimizing risks to participants.

²⁷ See Glossary: "Informed consent."

An ethical commitment to fairness and equity for interestholders requires the intentional effort of evaluators to produce work that is valid, honest, respectful of interestholders, and considerate of the general public welfare. This applies to all aspects of evaluation design, including providing information that is accessible, intelligible, and usable to meet the needs of interestholders. To ensure fair, just, and equitable treatment of all persons and affected entities, evaluators should gain an understanding of the range of perspectives and interests that individuals and groups bring to the evaluation, including those not usually represented.

Evaluations should be equitable, fair, and just, and should take into account cultural and contextual factors that could influence the findings or their use. This includes accounting for cultural and contextual factors such as languages spoken, political and social climate, power and privilege, or economic conditions. In addition to protecting participants and evaluated entities, evaluators should take into account those affected by evaluation findings, including individuals, communities, and organizations. Ethical evaluators identify and inform interestholders of the evaluation prior to its start and communicate findings to affected individuals and entities upon completion.

Guidance for Practice: Ethics and Inclusivity

Informed Consent and Participant Protection

- Establish and regularly update protocols, procedures, and accountability mechanisms that ensure individuals contributing data to evaluations are accurately informed of their rights and potential risks associated with the provision of such data, as well as additional information relevant to their informed decision making about participation.
- Provide clear information about who will have access to the data acquired, how the data will be stored and maintained, and when or if the data will be deleted.
- Ensure an adequate balance between risks and benefits by weighing the utility of reducing data collection burden on participants with the potential benefits through additional assessments with respect to government efficiency, effectiveness, and preventing harm.

Data Privacy and Governance

- Develop mechanisms to ensure that internal and external evaluators are aware of and adhere to the rights, rules, and procedures regarding data collection, data ownership, and communications of the resulting evaluation insights within each unique context.
- Create mechanisms for regularly reviewing and updating privacy-related operational level processes and protocols to ensure they remain robust means for adequately protecting data privacy amid rapidly advancing changes in technology.
- Uphold applicable laws, policies, and protocols for assessing and mitigating the risk of disclosure when evaluation data includes personally identifiable information or other sensitive information.
- Ensure alignment of evaluation policies and procedures with applicable federal, state, and local rules and regulations regarding data privacy and governance.

- Take advantage of rapid advances in methods for protecting privacy while allowing for ongoing research and evaluation, including data access enclaves using statistical algorithms, synthetic data matching, and other approaches.

Data Sovereignty and Inclusivity

- Ensure Indigenous data sovereignty is recognized and upheld for all communities.²⁸ Incorporate principles and relevant U.S. and Tribal law on Indigenous data²⁹ ownership, data governance,³⁰ and data sovereignty into evaluation policies, plans, and guidance documents.³¹
- Assess the extractive nature of data collection within relevant communities and take meaningful steps to ensure reciprocity and mutual benefit in the evaluation process.

²⁸ See Glossary: “Indigenous data sovereignty”

²⁹ See Glossary: “Indigenous data”

³⁰ See Glossary: “Indigenous data governance”

³¹ Nicole Bowman and Claire Bjork, "Indigenous Systems of Evaluation," *Oxford Bibliographies in Education*, October 2025, <https://doi.org/10.1093/obo/9780199756810-0329>.

Key Laws and Policies Affecting Evaluation in Government

Significant progress has been made in establishing evaluation as an integral component of government program management. However, additional steps are needed before agencies are consistently using program evaluation and systematic analysis to improve program design, implementation, and effectiveness and sharing what works, does not work, and why. There are historical precedents to build on and some new steps worth taking. The federal laws that have relevance are discussed here, but it is important to recognize that Tribal laws and state laws also affect the goals and implementation of evaluation.

Evaluation has taken a long trajectory of development in the U.S. federal government, shaped by many laws, policies, and administrations.³² Starting in the 1960s, some federal departments established evaluation offices and used a variety of funding mechanisms to provide resources and expertise for evaluation. In the 1970s, independent oversight on program effectiveness was established through federal agencies' inspectors general.³³ During the same period, the then-called U.S. General Accountability Office (now Government Accountability Office) began leading more government-wide performance audits and evaluations to promote agency economy, efficiency, and effectiveness.

The Government Performance and Results Act (GPRA) of 1993,³⁴ modified by the GPRA Modernization Act of 2010,³⁵ was one of the first government-wide efforts to encourage the use of evidence and evaluation by civil servants and policymakers. GPRA requires each agency to develop a strategic plan and mission statement, and to determine whether their programs achieve their goals and objectives. In practice, most government agencies have implemented GPRA by using performance measurement³⁶ to determine whether they have reached a goal, why programs do or do not meet their goals and objectives, and how to improve programs.

While the practice of evaluation has continued to grow rapidly within federal agencies following the enactment of GPRA, the capacity to effectively implement evaluation activities varied greatly across the federal government. In 2017, the U.S. Commission on Evidence-Based Policymaking's members unanimously concluded that the federal government needed to improve its capacity to generate useful evidence and then use that information in decision-making.³⁷

³² Esther Nolton, "Mapping the Institutionalization of Evaluation in the U.S. Federal Government" (PhD diss., George Mason University, 2020), <https://hdl.handle.net/1920/12476>.

³³ Charles A. Johnson and Kathryn E. Newcomer, *U.S. Inspectors General: Truth Tellers in Turbulent Times* (Washington, DC: Brookings Institution Press, 2019).
<https://www.brookings.edu/books/u-s-inspectors-general/>.

³⁴ Government Performance and Results Act of 1993, Pub. L. No. 103-62, 107 Stat. 285 (1993).
<https://www.govinfo.gov/content/pkg/BILLS-103s20enr/pdf/BILLS-103s20enr.pdf>.

³⁵ GPRA Modernization Act of 2010, Pub. L. No. 111-352, 124 Stat. 3866 (2011).
<https://www.congress.gov/111/plaws/publ352/PLAW-111publ352.pdf>.

³⁶ See Glossary: "Performance measurement."

³⁷ U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation. "Commission on Evidence-Based Policymaking."
<https://acf.gov/opre/project/commission-evidence-based-policymaking-cep>.

Following the Commission's final report,³⁸ the Evidence Act was passed by Congress in 2018 and signed in 2019, outlining an expectation that federal agencies establish a core infrastructure to support and implement an evaluation function. The Evidence Act aims to increase both agencies' production and their use of a portfolio of evaluation and other systematic evidence. The Evidence Act's required infrastructure for agencies' evaluation activities includes a senior responsible leader or evaluation officer, strategic planning for research and evaluation, written evaluation policies and procedures, support for workforce expertise to implement the function, and reporting on agency capabilities for analysis and evaluation.

The law directs agencies to develop evidence-building plans, establish Evaluation Officers, produce written evaluation policies, periodically report on their capacity to engage in statistics, evaluation, and policy analyses, and use such evidence for day-to-day governance. The law also established Chief Data Officers who manage data quality and governance and Statistical Officials who coordinate and oversee statistical activities. Together, the leadership, technical expertise, and strategic infrastructure established under the law has been a driving force generating demand for evaluation in government.

To build on this legacy and continue to improve the capabilities for evaluation in government, and the ensuing benefits, both the executive branch and Congress need to continue to refine the role of evaluation moving forward.

Executive Branch Role

Many public agencies now have a legal obligation to develop evaluation infrastructure and facilitate the practice of evaluation, particularly within the U.S. federal executive branch. Under the Foundations for Evidence-Based Policymaking Act, this legal expectation for well-developed and stable evaluation practices only extends to 24 of the largest federal agencies. Other agencies, departments, and bureaus, to which some provisions of the law do not apply, can also use the direction and expectations of the law in conjunction with GPRA to develop robust evaluation capabilities.

With or without a legal requirement for evaluation infrastructure, public agencies and programs will have different evaluation needs, which affects the design and organization of supporting operations. No single best practice or strategy exists for organizing evaluation offices and functions. While the 24 largest federal agencies are directed by law to develop certain activities department-wide, including the appointment of an evaluation officer, there are multiple strategies for envisioning what an effective or useful operation may look like in a U.S. or Tribal agency at any level of governance.

Central Evaluation Leader and Office

One option is to develop a central evaluation center at an agency, headed by an evaluation officer with support staff, the financial capability, and staff expertise to support evaluation

³⁸ Commission on Evidence-Based Policymaking. *The Promise of Evidence-Based Policymaking: Report of the Commission on Evidence-Based Policymaking*. Washington, DC: CEP, 2017. <https://www2.census.gov/adrm/fesac/2017-12-15/Abraham-CEP-final-report.pdf>.

activities across the entire agency. Employing this approach would likely require direct funding for a central evaluation unit as well as budgetary transfer authorities, such as in place at many federal agencies (e.g. the U.S. Department of Labor's Chief Evaluation Office³⁹), to support central evaluation activities.

Evaluation Leadership in a Federated System

In a federated system, different programs in an agency establish evaluation expertise and representatives that work in collaboration with a central evaluation officer at the agency. Individual agencies coordinate evaluation expenditures and resources with a chief evaluation officer, but the bulk of expertise and resources are retained by the individual programs. The U.S. Department of Health and Human Services has essentially operated a federated system since the 1970s, with some leadership provided by the Assistant Secretary for Planning and Evaluation,⁴⁰ but decision-making about evaluation activities conducted within individual operating divisions like the Centers for Disease Control and the Administration for Children and Families.

Evaluation Leadership with Decentralized Units

A decentralized evaluation approach in an agency recognizes that a central evaluation leader offers nominal guidance to support evaluation activities dispersed throughout the agency. This type of an approach may be preferable in agencies with a vast array of different types of activities and programs, such as the U.S. Department of Commerce. Decentralized approaches can also be useful in ensuring that the evaluation function remains strong and present during leadership turnovers or unexpected changes in resource allocations.

Whatever model is chosen, evaluation office(s) must include the functions and possess the attributes described above under the evaluation standards. Any of these approaches could be used to plan, conduct, and procure evaluation.

Legislative Branch Role

Building on the existing structures and organizations that support decision-making in legislative bodies such as the U.S. Congress and state and Tribal legislatures, a stronger connection between evaluations and legislative activities can emerge. This can be done by building evaluation expectations into authorizing legislation, explicitly setting aside adequate resources for evaluation, and signaling the importance of evaluation activities to agency officials.

Authorizations

Program authorization and periodic reauthorization provide opportunities for legislatures to establish frameworks for systematic evaluation of new and continuing programs. Legislative committees can, through authorizing legislation, provide guidance or mandates relevant to evaluation. Related activities could include:

³⁹ U.S. Department of Labor, Office of the Assistant Secretary for Policy. "About the Chief Evaluation Office." <https://www.dol.gov/agencies/oasp/evaluation/about>.

⁴⁰ U.S. Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation. "About ASPE." <https://aspe.hhs.gov/about>.

- Clearly identifying the intended goals of desired outcomes in legislation
- Encouraging early implementation reviews to identify start-up problems and correct them before they become more serious
- Requiring evaluation planning at the outset of program design
- Providing adequate authority to collect meaningful data relevant for program analysis, or the sharing of data between relevant government agencies and programs
- Establishing or expanding ongoing surveys or other data-collection mechanisms to become relevant sources of reliable data
- Monitoring relevant performance indicators
- Facilitating studies reviewing the efficiency of program management and fidelity of implementation
- Encouraging targeted studies assessing program effects and identifying why programs are or are not effective
- Requiring evaluation about topics of interest to the legislative body in support of authorization activities

Appropriations

During appropriations processes, legislatures should ensure that agencies have sufficient resources to meet the evidentiary and evaluation needs of decision-makers in the legislature as well as executive agencies. Related activities could include:

- Funding for evaluation activities, including strengthening evaluation capacity in terms of personnel and professional development
- Developing funding set-aside authorities to enable agencies to transfer funding across budget accounts for evaluation activities
- Establishing flexible funding approaches and incentive funds to encourage innovation in research and evaluation
- Providing procurement flexibilities for long-term evaluations that may extend beyond the statutory period of availability for appropriated funds
- Posing questions to agencies about evaluation activities and insights of completed evaluations during budget and appropriations hearings
- Requiring evaluations about topics of interest to the legislature in support of appropriation activities

Oversight

The routine role of legislative oversight could lead to countless suggestions about facilitating strong evaluation activities in government. Activities could include:

- Holding hearings or less formal meetings to discuss evaluation findings, capacity, infrastructure, plans, or learning agendas with agency evaluation officers
- Monitoring agency implementation of the evaluation function, including establishment of evaluation units, appointment of qualified evaluation leaders, and transparency about evaluation activities
- Asking questions to individuals seeking legislative confirmation about their support and leadership interest in promoting strong evaluation policy in the agency

- Requiring evaluation about topics of interest to the legislature in support of oversight activities

Collaboration Between Executive and Legislative Branches

The usefulness of evaluation results could be maximized if legislative and executive branches of government jointly specify broad evaluation expectations and concerns in authorizing statutes, appropriations, and other activities. Such collaboration will not always be easy or even possible to achieve. Nevertheless, when possible, a partnership of this kind can help increase the benefits that evaluation provides.

Legislatures and public agencies can jointly work together to identify evaluation priorities, funding needs, and capabilities. The process of developing learning agendas in agencies should include input about legislative informational needs, as well as other interestholders. The allocation of resources to evaluations should result from a dialogue about respective priorities, acknowledging timelines and knowledge gaps.

Tribal Governments and Evaluation

Tribal governments occupy a distinct position in the federal legal order. Recognized Tribal nations are sovereign governments with inherent authority that predates the U.S. Constitution, and the federal government's relationship with them is a government-to-government relationship grounded in treaties, statutes, and the federal trust responsibility.^{41,42} This relationship has direct consequences for how evaluations are planned, conducted, and governed in Tribal contexts. Tribes are not a demographic category or a type of program interestholder; they are governing authorities with legal rights over their citizens, territories, resources, and data.

The Indian Self-Determination and Education Assistance Act (ISDEAA)⁴³ and the Tribal Self-Governance Act of 1994⁴⁴ are the principal federal vehicles through which Tribes exercise operational authority over program data. ISDEAA, enacted in 1975, codified Tribal self-determination by authorizing Tribes and Tribal organizations to contract directly with the federal government for health, education, and other programs previously administered by the Bureau of Indian Affairs and the Indian Health Service. Under self-determination contracts and self-governance compacts, Tribes generate, own, and manage program data, with federal access negotiated through the agreement rather than imposed by agency policy. Many Tribes

⁴¹ University of Alaska Fairbanks, Department of Tribal Governance, "General Principles of Federal Indian Law," *Tribal Governance*, <https://www.uaf.edu/tribal/academics/112/unit-4/generalprinciplesoffederalindianlaw.php>.

⁴² U.S. Department of the Interior, Bureau of Indian Affairs, "What Is the Federal Indian Trust Responsibility?," Indian Affairs, November 8, 2017, <https://www.bia.gov/faqs/what-federal-indian-trust-responsibility>.

⁴³ Indian Self-Determination and Education Assistance Act, Public Law 93-638, 88 Stat. 2203 (1975), codified at 25 U.S.C. § 5301 et seq. <https://www.govinfo.gov/content/pkg/STATUTE-88/pdf/STATUTE-88-Pg2203.pdf>

⁴⁴ Tribal Self-Governance Act of 1994, Public Law 103-413, Title II, 108 Stat. 4270 (1994), codified at 25 U.S.C. §§ 5361–5368. <https://www.govinfo.gov/content/pkg/BILLS-103hr3508eh/pdf/BILLS-103hr3508eh.pdf>

and Tribal organizations have built on this foundation by establishing their own policies, research review processes, and data governance protocols governing research and evaluation with their community members.^{45,46}

Evaluators working with federally funded programs in Tribal contexts should treat Tribal ownership of program data as a structural feature of the relationship, not as a preference to be accommodated. The historical context for these protections matters: Indigenous communities have experienced research that harmed rather than benefited them, and distrust of outside researchers and federal agencies remains well-documented.⁴⁷ That history is part of what Tribal data governance structures exist to address, and evaluators should approach it accordingly.

The federal government's consultation obligations derive from Executive Order 13175, signed in 2000, making it a durable anchor in federal Tribal consultation policy.⁴⁸ EO 13175 requires federal agencies to consult with Tribal governments before taking actions that have substantial direct effects on Tribes, their rights, or their resources. More recently, the 2022 Memorandum on Uniform Standards for Tribal Consultation established minimum procedural standards for that consultation and explicitly recognized Tribal data privacy as a consultation consideration.⁴⁹ In the same month in 2022, the Office of Science and Technology Policy and the Council on Environmental Quality issued the first government-wide guidance directing federal agencies to identify mechanisms to protect sensitive Indigenous Knowledge in their programs and activities.⁵⁰ Together, these policies and guidance, among others, establish frameworks for engagement and data-protection that directly apply to evaluation activities, including decisions about data collection and dissemination, where those activities have meaningful implications for Tribal governments or their members.

⁴⁵ Nicole Bowman and Carolee Dodge-Francis, "Culturally Responsive Indigenous Evaluation and Tribal Governments: Understanding the Relationship," *New Directions for Evaluation* 2018, no. 159 (2018): 17–31, <https://doi.org/10.1002/ev.20329>

⁴⁶ Stephanie Russo Carroll, Desi Rodriguez-Lonebear, and Andrew Martinez, "Indigenous Data Governance: Strategies from United States Native Nations," *Data Science Journal* 18, no. 1 (2019): 31, <https://doi.org/10.5334/dsj-2019-031>.

⁴⁷ Anna Harding, Barbara Harper, Dave Stone, Catherine O'Neill, Patricia Berger, Stuart Harris, and Jamie Donatuto, "Conducting Research with Tribal Communities: Sovereignty, Ethics, and Data-Sharing Issues," *Environmental Health Perspectives* 120, no. 1 (2012): 6–10, <https://doi.org/10.1289/ehp.1103904>.

⁴⁸ Executive Order 13175, "Consultation and Coordination with Indian Tribal Governments," 65 Fed. Reg. 67249 (November 9, 2000). <https://www.federalregister.gov/documents/2000/11/09/00-29003/consultation-and-coordination-with-indian-tribal-governments>.

⁴⁹ Executive Office of the President, Memorandum on Uniform Standards for Tribal Consultation, 87 Fed. Reg. 74479 (December 5, 2022). <https://www.federalregister.gov/documents/2022/12/05/2022-26555/uniform-standards-for-tribal-consultation>.

⁵⁰ Executive Office of the President, Guidance for Federal Departments and Agencies on Indigenous Knowledge (Washington, DC: Office of Science and Technology Policy and Council on Environmental Quality, November 30, 2022). <https://bidenwhitehouse.archives.gov/ceq/news-updates/2022/12/01/white-house-releases-first-of-a-kind-indigenous-knowledge-guidance-for-federal-agencies/>.

Many Tribal nations have enacted formal research approval requirements, including Tribal Institutional Review Boards (IRBs) and research review processes, that apply to evaluation activities conducted within their jurisdictions or using data about their citizens.⁵¹ These processes operate in addition to federal Common Rule protections under 45 CFR Part 46. The Common Rule's 2018 revisions expressly preserve the force of Tribal law in research governance, providing that federal human-subjects protections do not preempt Tribal laws that require additional protections for research participants or that mandate separate Tribal review of cooperative research projects.⁵² Evaluators should identify and engage Tribal review processes early in evaluation planning, recognizing that Tribal research codes frequently require community partnership agreements, prepublication review, and the return of results to the community.

The Evidence Act and its implementing guidance do not address Tribal data sovereignty or government-to-government consultation obligations as they apply to evaluation. The *Roadmap* recognizes this silence as a gap. Frameworks such as the CARE Principles for Indigenous Data Governance⁵³ — covering Collective Benefit, Authority to Control, Responsibility, and Ethics — and the First Nations OCAP® Principles⁵⁴ provide practical guidance for evaluators seeking to implement data sovereignty in evaluation design and data stewardship.

State and local governments engaged in evaluation with Tribal implications face a parallel set of considerations. Many states have codified government-to-government consultation requirements, and some have enacted Tribal data protections that apply to state-administered programs serving Tribal citizens. Evaluators working at the state and local level should identify applicable state consultation statutes and any Tribal data-sharing agreements in effect before designing evaluations that involve Tribal communities or their data.

⁵¹ Deana Around Him, Temana Andalcio Aguilar, Anita Frederick, Heather Larsen, Michaela Seiber, and Jyoti Angal, "Tribal IRBs: A Framework for Understanding Research Oversight in American Indian and Alaska Native Communities," *American Indian and Alaska Native Mental Health Research* 26, no. 2 (2019): 71–95.
https://coloradosph.cuanschutz.edu/docs/librariesprovider205/journal_files/vol26/26_2_2019_71_around-him.pdf?sfvrsn=cffe0b9_2

⁵² Protection of Human Subjects, 45 C.F.R. § 46.101(f) (2018).

⁵³ Carroll, S. R., Herczog, E., Hudson, M., Russell, K., & Stall, S. (2021). Operationalizing the CARE and FAIR principles for Indigenous data futures. *Scientific Data*, 8(108) 1–6.
<https://doi.org/10.1038/s41597-021-00892-0>.

⁵⁴ First Nations Information Governance Center. (n.d.). The First Nations principles of ownership, control, access, and possession (OCAP). <https://fnigc.ca/ocap-training/>.

Looking to the Future

The rapid development of artificial intelligence, machine learning, generative AI, predictive analytics, privacy enhancing technologies, and big data systems presents both substantial opportunities and significant challenges for evaluation practice. These technologies are transforming how evaluations can be designed, conducted, and used, and the pace of change requires continuous attention from evaluation leaders and practitioners.

Open access to data and technologically-based data privacy tools can broaden the scope of evaluations and the diversity of researchers conducting evaluations, potentially enabling smaller and newer evaluators to participate in large-scale studies that previously required high levels of staff and resources.⁵⁵ The availability of big data sets allows evaluations to capture information for analyzing past trends, applying statistical models and algorithms to analyze various scenarios, and delineating subgroup characteristics and outcomes that are often of interest in evaluations. Multiple large electronic data sources provide opportunities to link and merge data items and track outcomes and trends over time, for example in large-scale longitudinal evaluations of the impact of social programs and services on economic and social wellbeing of participants.

As machine learning technologies have developed, AI integrations with big data sources are becoming widely adopted across industries to enable faster, less costly insights, potentially opening the door to many opportunities for evaluators to enhance their analyses.^{56,57} These new technologies place new requirements on evaluation management and leadership. Evaluation managers and leaders should engage with experts to continuously monitor the most appropriate tools for different evaluation activities and methods, and incorporate assessments into evaluation plans and study designs. There are many AI tools already available for performing regular tasks in evaluation, including analyzing and synthesizing literature, summarizing discussions from focus groups or staff interviews, and conducting meta-analyses of results from multiple evaluations.^{58,59}

Likewise, emerging benefits of rapid developments in big data include new methods for protecting privacy while allowing for ongoing research and evaluation. Technologies such as data access enclaves using statistical algorithms, synthetic data matching, and other emerging approaches allow information from multiple sources to be accessed and merged while

⁵⁵ Commission on Evidence-Based Policymaking. *The Promise of Evidence-Based Policymaking: Report of the Commission on Evidence-Based Policymaking*.

⁵⁶ Steffen Bohni Nielsen, "Disrupting Evaluation? Emerging Technologies and Their Implications for the Evaluation Industry," *New Directions for Evaluation* 2023, no. 178–179 (2023): 47–57, <https://doi.org/10.1002/ev.20558>.

⁵⁷ Cari Beth Head, Paul Jasper, Matthew McConnachie, Linda Raftree, and Grace Higdon, "Large Language Model Applications for Evaluation: Opportunities and Ethical Implications," *New Directions for Evaluation* 2023, no. 178–179 (2023): 33–46, <https://doi.org/10.1002/ev.20556>.

⁵⁸ Nina R. Sabarre, Blake Beckmann, Sahiti Bhaskara, and Kathleen Doll, "Using AI to Disrupt Business as Usual in Small Evaluation Firms," *New Directions for Evaluation* 2023, no. 178–179 (2023): 59–71, <https://doi.org/10.1002/ev.20562>.

⁵⁹ Head et al., "Large Language Model Applications for Evaluation."

restricting ongoing access to private and confidential data.⁶⁰ Evaluation managers and leaders of public agencies should seek out opportunities to understand, apply, and evaluate the effectiveness of these technologies for supporting learning activities.

The true potential of AI and big data opportunities in evaluation is still not fully understood, but the standards and practices outlined in this document provide a way to harness these opportunities by incorporating new technological contexts into existing evaluation management, policies and procedures, guidelines, and usage. The principles of relevance, rigor, independence, transparency, and ethics apply to evaluations using these new technologies just as they apply to traditional methods. High-quality standards for the design and conduct of evaluations should require complete transparency where AI is applied in data and analysis, with clear acknowledgment of when and how AI was used and appropriate documentation to enable assessment and replication of findings.

As these technologies continue to evolve, evaluation policies and procedures should be continuously reviewed and updated to remain current while retaining key evaluation principles. Frameworks have already emerged that provide structured approaches for characterizing AI systems, managing AI-related risks, and integrating trustworthy AI into research and evaluation practice.^{61,62,63} The evaluation community must remain abreast of technological developments and their implications for evaluation practice, ensuring that the benefits of new technologies are realized while protecting against their potential risks. Many academic science and research units, news media, journals, and research organizations are establishing policies for the ethical and transparent use of AI, and evaluation offices should do likewise.

Additional information about the American Evaluation Association is available at www.eval.org. The AEA Evaluation Policy Task Force can be contacted at evaluationpolicy@eval.org.

⁶⁰ Federal Committee on Statistical Methodology. *Data Protection Toolkit*. Washington, DC: Office of Management and Budget. <https://nces.ed.gov/fcsm/dpt>.

⁶¹ Bianca Montrosse-Moorhead, "Evaluation Criteria for Artificial Intelligence," *New Directions for Evaluation* 2023, no. 178–179 (2023): 123–134, <https://doi.org/10.1002/ev.20566>.

⁶² National Institute of Standards and Technology, *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*, NIST AI 100-1 (Gaithersburg, MD: U.S. Department of Commerce, January 2023), <https://doi.org/10.6028/NIST.AI.100-1>.

⁶³ OECD, *OECD Framework for the Classification of AI Systems*, OECD Digital Economy Papers, No. 323 (Paris: OECD Publishing, 2022), <https://doi.org/10.1787/cb6d9eca-en>.

Appendix: Glossary

Accountability	The obligation of public officials and agencies to answer in a timely manner about actions, decisions, and use of public resources under their authority, demonstrating that programs achieve intended results and provide value to taxpayers.
Culturally responsive evaluation	A holistic framework that includes the centrality of, and attunement to, culture in the theory and practice of evaluation. CRE recognizes that demographic, sociopolitical, and contextual dimensions, locations, perspectives, and characteristics of culture matter fundamentally in evaluation. CRE gives particular attention to groups that have been historically marginalized, seeking to bring balance and equity into the evaluation process. ⁶⁴
Effectiveness	The extent to which a program achieves its stated goals, objectives, and intended outcomes. Measures whether a program produces desired results and impacts for its target population, comparing actual outcomes against performance targets. An effective program successfully addresses the problem it was designed to solve and generates meaningful benefits for intended beneficiaries.
Evaluation	<p>A systematic process to determine merit, worth, value or significance. Documentation of the process to memorialize activities is critical for transparency, interpretation, external validation, and future replication.</p> <p>Program evaluation is the application of systematic methods to address questions about program operations and results. It may include ongoing monitoring of a program, as well as one-off studies of program processes or program impact. The approaches used are based on social science research methodologies and professional standards.</p> <ul style="list-style-type: none"> • Formative evaluation – Evaluation during program development or early implementation to improve design and operations, providing real-time feedback for refinement before full-scale rollout. • Impact evaluation – Assessment of net effects directly attributable to a program using rigorous methods to establish causality, typically comparing participants against control or comparison groups. • Outcome evaluation – Assessment of whether a program achieves intended results and objectives, measuring changes in participant knowledge, behaviors, or conditions. • Process evaluation – Examination of program implementation, service delivery, and operations to determine whether the program functions as designed and reaches intended populations. • Economic evaluation – Analysis of program costs relative to benefits or outcomes, including cost-benefit, cost-effectiveness, and return on

⁶⁴ Stafford Hood, Rodney K. Hopson, and Karen E. Kirkhart, "Culturally Responsive Evaluation: Theory, Practice, and Future Implications," in *Handbook of Practical Program Evaluation*, 4th ed., ed. Kathryn E. Newcomer, Harry P. Hatry, and Joseph S. Wholey (Hoboken, NJ: Wiley, 2015), 281–317, <https://doi.org/10.1002/9781119171386.ch12>.

	investment calculations.
Evaluation capacity	An organization's ability to conduct, commission, interpret, and apply results from program evaluations effectively. Includes necessary resources (funding, staff), expertise (trained evaluators), technical infrastructure (data systems, tools), and organizational cultures that value evidence-based decision-making. Encompasses both the skills needed to design rigorous evaluations and the institutional mechanisms needed to integrate findings into program management and policy decisions.
Evaluation plans	A structured document outlining the framework and methodology for conducting a program evaluation. Specifies the evaluation's purpose, key questions, measures, data collection methods, timelines, roles and responsibilities, and reporting procedures. Serves as a roadmap for systematic assessment that addresses interestholder's information needs and supports evidence-based decision-making.
Evaluation policy	Any rule or principle that a group or organization uses to guide its decisions and actions when doing evaluation. ⁶⁵
Evidence	Data that are used with appropriate analysis to support a claim being made. Typically, the claims describe conditions targeted or affected by public policies and programs, and the data used may include administrative or survey data to support performance management or evaluation activities. ⁶⁶
Independence	Freedom from perceived or actual conflicts of interest and undue influence that could compromise an evaluation's objectivity and credibility. Involves structural separation from the program being evaluated and professional autonomy to design evaluations, analyze findings, and report results without interference. Independent evaluations are conducted by entities with no stake in program outcomes, ensuring impartial, evidence-based assessments that enhance interestholder confidence in findings.
Indigenous data	Information or knowledge, in any format or medium, which is about and may affect Indigenous peoples both collectively and individually. ⁶⁷
Indigenous data governance	The right of Indigenous peoples to autonomously decide what, how, and why Indigenous data are collected, accessed and used. It ensures that data on or about Indigenous peoples reflects our priorities, values, cultures, worldviews, and diversity. ⁶⁸

⁶⁵ Trochim, William M. K. "Evaluation Policy and Evaluation Practice." *New Directions for Evaluation* 123 (2009): 13–32. <https://doi.org/10.1002/ev.303>.

⁶⁶ Kathryn E. Newcomer and Nicholas Hart, *Evidence-Building and Evaluation in Government*, Evaluation in Practice Series, vol. 7 (Thousand Oaks, CA: Sage, 2022). <https://collegepublishing.sagepub.com/products/evidence-building-and-evaluation-in-government-1-27309>

⁶⁷ Nicole Bowman and Leanne Bremner, "Indigenous Data Sovereignty: Applying It By, With, For, and Through Indigenous Evaluators and Evaluations," *Canadian Journal of Program Evaluation* 39, no. 2 (2025): 265–287, <https://doi.org/10.3138/cjpe-2024-0039>.

⁶⁸ Bowman and Bremner, "Indigenous Data Sovereignty."

Indigenous data sovereignty	The right of Indigenous peoples to exercise ownership over Indigenous data. Ownership of data can be expressed through the creation, collection, access, analysis, interpretation, management, dissemination, and reuse of Indigenous data. ⁶⁹
Informed consent	<p>The voluntary agreement to participate in evaluation activities after receiving clear information about the evaluation's purpose, procedures, risks, benefits, and participant rights. Ensures individuals understand what participation involves, how information will be used and protected, and their right to refuse or withdraw without penalty. Protects participant autonomy and ensures evaluations meet ethical standards and human subjects research regulations.</p> <p>In Tribal and Indigenous contexts, informed consent may also operate at the community or Tribal government level. Many Tribal nations require community-level consent through Tribal Council authorization or Tribal IRB approval in addition to individual participant consent, reflecting the collective nature of research and evaluation impacts on Indigenous communities.</p>
Interestholder	<p>Any person or organization having an investment in the evaluation, such as those served or affected by the program, those planning or implementing the program, those who might use the evaluation findings, and those who are skeptical about the program.</p> <p>This <i>Roadmap</i> adopts the term “interestholder” in place of the commonly used “stakeholder” in alignment with the 2024 CDC Program Evaluation Framework.⁷⁰ The term “stakeholder” can “indicate a power differential between groups and that is recognized as having a violent connotation for certain American Indian or Alaska Native Tribes and Tribal members.”⁷¹ The term “interestholder” advances equity through inclusive and respectful language that welcomes all persons with an interest in programs and policies to be engaged in the collaborative process of evaluation.</p>
Learning agenda	A strategic framework identifying priority questions that an organization seeks to answer through evaluation and research activities over a defined period. Outlines what an organization needs to learn to improve program effectiveness and inform decisions. Guides coordinated evaluation planning and ensures evaluation efforts align with organizational priorities. Promotes evidence-based management and continuous improvement through systematic inquiry.
Logic model	A visual representation depicting logical relationships between a program's resources, activities, outputs, and intended outcomes. Maps the causal pathway from inputs through the activities and outputs of a program, to the short-term and long-term outcomes and impacts of the program. Clarifies

⁶⁹ Bowman and Bremner, "Indigenous Data Sovereignty."

⁷⁰ Kidder, Daniel P., Lisa A. Fierro, Erica Luna, et al. "CDC Program Evaluation Framework, 2024." MMWR Recommendations and Reports 73, no. RR-6 (2024): 1–37. <https://doi.org/10.15585/mmwr.rr7306a1>.

⁷¹ Kidder et al., "CDC Program Evaluation Framework, 2024."

	program theory, articulates assumptions about how a program works, identifies performance indicators, and facilitates interestholder communication about program design and expected results.
Performance measurement	The ongoing, systematic collection and reporting of data to track program progress toward goals using quantitative and qualitative indicators. Monitors inputs, activities, outputs, outcomes, and efficiency over time through routine metrics and reporting systems. Enables managers to assess whether targets are met, identify trends, and make timely adjustments. Provides continuous data that inform management decisions and support more in-depth evaluations.
Policy	A deliberate course of action or set of principles adopted by an organization to address issues and achieve specific objectives. Provides the framework and authority for programs by defining the problems to address, goals to pursue, and resource allocation. In evaluation, policies are both the subject of assessment (whether they achieve intended effects) and the context for programs. Policies are expressed through legislation, regulations, executive orders, or agency directives.
Program	A set of resources and activities directed toward one or more common goals, typically under the direction of a single manager or management team. Programs may also span multiple agencies, levels of government, or sectors in pursuit of shared objectives. This includes systemic change initiatives that address social problems embedded in multiple, interconnected systems, such as efforts to improve how service systems affect individual and community well-being.
Rigor	The degree to which an evaluation employs methodologically sound approaches that produce credible, reliable, and valid findings. Involves appropriate research designs, systematic data collection, valid measurement instruments, robust analysis, and transparent documentation.
Systems	Interdependent entities that, through their interactions, function as a whole and have some boundary that can be used to distinguish the system from its larger environment. ⁷²
Systemic Change	A way of seeing and acting in the world; a way of working that centers interconnectedness and uses systems and methods in ways that are seeking and responsive to changes occurring in systems. ⁷³
Theory of change	A model that describes and illustrates how and why a specific policy or program is expected to produce desired outcomes, and maps out the various

⁷² Gates, Emily F. and Pablo Vidueira, *Evaluative Inquiry for Systemic Change: A Guide to Shift Beyond Fixes to Lasting Value* (Thousand Oaks, CA: Sage, 2026).
<https://uk.sagepub.com/en-gb/eur/evaluative-inquiry-for-systemic-change/book290068>

⁷³ Gates and Vidueira, *Evaluative Inquiry for Systemic Change*.

	assumptions and mediating factors that may facilitate or impede the policy or program from achieving espoused goals. ⁷⁴
U.S. and Tribal Governments	Governmental and quasi-governmental entities that deliver public services and manage publicly funded programs in the United States of America and Tribal nations therein. Includes federal, state, and local agencies, publicly funded institutions, and the sovereign Tribal governments ⁷⁵ with their agencies and enterprises serving Native American and Alaska Native communities. In evaluation, these organizations are both subjects of assessment and users of evaluation findings to improve services, and demonstrate accountability under their unique legal frameworks and governance structures.

⁷⁴ Newcomer and Hart, *Evidence-Building and Evaluation in Government*.

⁷⁵ Legal Information Institute. "2 U.S.C. § 658 - Definitions." Cornell Law School. <https://www.law.cornell.edu/uscode/text/2/658>.

Citations

- American Evaluation Association. "Competencies & Standards." 2018. <https://www.eval.org/About/Competencies-Standards>.
- American Evaluation Association. "Guiding Principles for Evaluators." 2018. <https://www.eval.org/About/Guiding-Principles>.
- American Evaluation Association. "AEA Mission, Vision, and Values." <https://www.eval.org/About/About-AEA/Mission-Vision-Values>.
- American Evaluation Association. "The Evaluation Policy Task Force." <https://www.eval.org/Evaluation-Policy/Evaluation-Policy-Task-Force>.
- Around Him, Deana, Temana Andalco Aguilar, Anita Frederick, Heather Larsen, Michaela Seiber, and Jyoti Angal. "Tribal IRBs: A Framework for Understanding Research Oversight in American Indian and Alaska Native Communities." *American Indian and Alaska Native Mental Health Research* 26, no. 2 (2019): 71–95. https://coloradosph.cuanschutz.edu/docs/librariesprovider205/journal_files/vol26/26_2_2_019_71_aroundhim.pdf?sfvrsn=cffe0b9_2.
- Bowman, Nicole, and Carolee Dodge-Francis. "Culturally Responsive Indigenous Evaluation and Tribal Governments: Understanding the Relationship." *New Directions for Evaluation* 2018, no. 159 (2018): 17–31. <https://doi.org/10.1002/ev.20329>.
- Bowman, Nicole, and Claire Bjork. "Indigenous Systems of Evaluation." *Oxford Bibliographies in Education*. October 2025. <https://doi.org/10.1093/obo/9780199756810-0329>.
- Bowman, Nicole, and Leanne Bremner. "Indigenous Data Sovereignty: Applying It By, With, For, and Through Indigenous Evaluators and Evaluations." *Canadian Journal of Program Evaluation* 39, no. 2 (2025): 265–287. <https://doi.org/10.3138/cjpe-2024-0039>.
- Carroll, Stephanie Russo, Edit Herczog, Maui Hudson, Keith Russell, and Shelley Stall. "Operationalizing the CARE and FAIR Principles for Indigenous Data Futures." *Scientific Data* 8, no. 108 (2021): 1–6. <https://doi.org/10.1038/s41597-021-00892-0>.
- Carroll, Stephanie Russo, Desi Rodriguez-Lonebear, and Andrew Martinez. "Indigenous Data Governance: Strategies from United States Native Nations." *Data Science Journal* 18, no. 1 (2019): 31. <https://doi.org/10.5334/dsj-2019-031>.
- Commission on Evidence-Based Policymaking. *The Promise of Evidence-Based Policymaking: Report of the Commission on Evidence-Based Policymaking*. Washington, DC: CEP, 2017. <https://www2.census.gov/adrm/fesac/2017-12-15/Abraham-CEP-final-report.pdf>.

- Executive Office of the President. *Guidance for Federal Departments and Agencies on Indigenous Knowledge*. Washington, DC: Office of Science and Technology Policy and Council on Environmental Quality, November 30, 2022.
<https://bidenwhitehouse.archives.gov/ceq/news-updates/2022/12/01/white-house-releases-first-of-a-kind-indigenous-knowledge-guidance-for-federal-agencies/>.
- Executive Order 13175. "Consultation and Coordination with Indian Tribal Governments." 65 Fed. Reg. 67249 (November 9, 2000).
<https://www.federalregister.gov/documents/2000/11/09/00-29003/consultation-and-coordination-with-indian-tribal-governments>.
- Federal Committee on Statistical Methodology. *Data Protection Toolkit*. Washington, DC: Office of Management and Budget. <https://nces.ed.gov/fcsm/dpt>.
- First Nations Information Governance Centre. "The First Nations Principles of Ownership, Control, Access, and Possession (OCAP)." <https://fnigc.ca/ocap-training/>.
- Foundations for Evidence-Based Policymaking Act of 2018, Pub. L. No. 115-435, 132 Stat. 5529 (2019). <https://www.congress.gov/115/plaws/publ435/PLAW-115publ435.pdf>.
- Gates, Emily F. and Pablo Vidueira, *Evaluative Inquiry for Systemic Change: A Guide to Shift Beyond Fixes to Lasting Value* (Thousand Oaks, CA: Sage, 2026).
<https://uk.sagepub.com/en-gb/eur/evaluative-inquiry-for-systemic-change/book290068>.
- Government Performance and Results Act of 1993, Pub. L. No. 103-62, 107 Stat. 285 (1993).
<https://www.govinfo.gov/content/pkg/BILLS-103s20enr/pdf/BILLS-103s20enr.pdf>.
- GPRA Modernization Act of 2010, Pub. L. No. 111-352, 124 Stat. 3866 (2011).
<https://www.congress.gov/111/plaws/publ352/PLAW-111publ352.pdf>.
- Harding, Anna, Barbara Harper, Dave Stone, Catherine O'Neill, Patricia Berger, Stuart Harris, and Jamie Donatuto. "Conducting Research with Tribal Communities: Sovereignty, Ethics, and Data-Sharing Issues." *Environmental Health Perspectives* 120, no. 1 (2012): 6–10. <https://doi.org/10.1289/ehp.1103904>.
- Hargreaves, Margaret, Brandon Coffee-Borden, and Chandria Jones. "Evaluating Systems Change: Systemic Change Measurement Rubric." Presentation, NORC at the University of Chicago, April 2026.
<https://www.norc.org/content/dam/norc-org/pdf2026/systems-change-rubric-presentation-april-2026.pdf>.
- Head, Cari Beth, Paul Jasper, Matthew McConnachie, Linda Raftree, and Grace Higdon. "Large Language Model Applications for Evaluation: Opportunities and Ethical Implications."

- New Directions for Evaluation* 2023, no. 178–179 (2023): 33–46.
<https://doi.org/10.1002/ev.20556>.
- Hood, Stafford, Rodney K. Hopson, and Karen E. Kirkhart. "Culturally Responsive Evaluation: Theory, Practice, and Future Implications." In *Handbook of Practical Program Evaluation*, 4th ed., edited by Kathryn E. Newcomer, Harry P. Hatry, and Joseph S. Wholey, 281–317. Hoboken, NJ: Wiley, 2015. <https://doi.org/10.1002/9781119171386.ch12>.
- Indian Self-Determination and Education Assistance Act, Public Law 93-638, 88 Stat. 2203 (1975), codified at 25 U.S.C. § 5301 et seq.
<https://www.govinfo.gov/content/pkg/STATUTE-88/pdf/STATUTE-88-Pg2203.pdf>.
- Johnson, Charles A., and Kathryn E. Newcomer. *U.S. Inspectors General: Truth Tellers in Turbulent Times*. Washington, DC: Brookings Institution Press, 2019. [U.S. Inspectors General | Brookings](https://www.brookings.edu/research/u-s-inspectors-general/)
- Kidder, Daniel P., Lisa A. Fierro, Erica Luna, et al. "CDC Program Evaluation Framework, 2024." *MMWR Recommendations and Reports* 73, no. RR-6 (2024): 1–37.
<https://doi.org/10.15585/mmwr.rr7306a1>.
- Legal Information Institute. "2 U.S.C. § 658 - Definitions." Cornell Law School.
<https://www.law.cornell.edu/uscode/text/2/658>.
- Memorandum on Uniform Standards for Tribal Consultation. 87 Fed. Reg. 74479 (December 5, 2022).
<https://www.federalregister.gov/documents/2022/12/05/2022-26555/uniform-standards-for-tribal-consultation>.
- Montrosse-Moorhead, Bianca. "Evaluation Criteria for Artificial Intelligence." *New Directions for Evaluation* 2023, no. 178–179 (2023): 123–134. <https://doi.org/10.1002/ev.20566>.
- National Institute of Standards and Technology. *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. NIST AI 100-1. Gaithersburg, MD: U.S. Department of Commerce, January 2023. <https://doi.org/10.6028/NIST.AI.100-1>.
- Newcomer, Kathryn E., and Nicholas Hart. *Evidence-Building and Evaluation in Government*. Evaluation in Practice Series, vol. 7. Thousand Oaks, CA: Sage, 2022.
<https://collegepublishing.sagepub.com/products/evidence-building-and-evaluation-in-government-1-273097>.
- Nielsen, Steffen Bohni. "Disrupting Evaluation? Emerging Technologies and Their Implications for the Evaluation Industry." *New Directions for Evaluation* 2023, no. 178–179 (2023): 47–57. <https://doi.org/10.1002/ev.20558>.

- Nolton, Esther. "Mapping the Institutionalization of Evaluation in the U.S. Federal Government." PhD diss., George Mason University, 2020. <https://hdl.handle.net/1920/12476>.
- OECD. *OECD Framework for the Classification of AI Systems*. OECD Digital Economy Papers, No. 323. Paris: OECD Publishing, 2022. <https://doi.org/10.1787/cb6d9eca-en>.
- Office of Management and Budget. *Phase 4 Implementation of the Foundations for Evidence-Based Policymaking Act of 2018: Program Evaluation Standards and Practices*. Memorandum M-20-12. Washington, DC: Executive Office of the President, 2020. <https://www.whitehouse.gov/wp-content/uploads/2020/03/m-20-12.pdf>.
- Protection of Human Subjects. 45 C.F.R. § 46.101(f) (2018). <https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-A/part-46>
- Rogers, Patricia. *Theory of Change*. Methodological Briefs: Impact Evaluation No. 2. Florence: UNICEF Office of Research–Innocenti, 2014. <https://reliefweb.int/report/world/methodological-briefs-impact-evaluation-no-2-theory-change>.
- Sabarre, Nina R., Blake Beckmann, Sahiti Bhaskara, and Kathleen Doll. "Using AI to Disrupt Business as Usual in Small Evaluation Firms." *New Directions for Evaluation* 2023, no. 178–179 (2023): 59–71. <https://doi.org/10.1002/ev.20562>.
- Tribal Self-Governance Act of 1994, Public Law 103-413, Title II, 108 Stat. 4270 (1994), codified at 25 U.S.C. §§ 5361–5368. <https://www.govinfo.gov/content/pkg/BILLS-103hr3508eh/pdf/BILLS-103hr3508eh.pdf>.
- Trochim, William M. K. "Evaluation Policy and Evaluation Practice." *New Directions for Evaluation* 123 (2009): 13–32. <https://doi.org/10.1002/ev.303>.
- University of Alaska Fairbanks, Department of Tribal Governance. "General Principles of Federal Indian Law." Tribal Governance. <https://www.uaf.edu/tribal/academics/112/unit-4/generalprinciplesoffederalindianlaw.php>.
- U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation. "Commission on Evidence-Based Policymaking." <https://acf.gov/opre/project/commission-evidence-based-policymaking-cep>.
- U.S. Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation. "About ASPE." <https://aspe.hhs.gov/about>.

U.S. Department of the Interior, Bureau of Indian Affairs. "What Is the Federal Indian Trust Responsibility?" Indian Affairs, November 8, 2017.
<https://www.bia.gov/fags/what-federal-indian-trust-responsibility>.

U.S. Department of Labor, Office of the Assistant Secretary for Policy. "About the Chief Evaluation Office." <https://www.dol.gov/agencies/oasp/evaluation/about>.