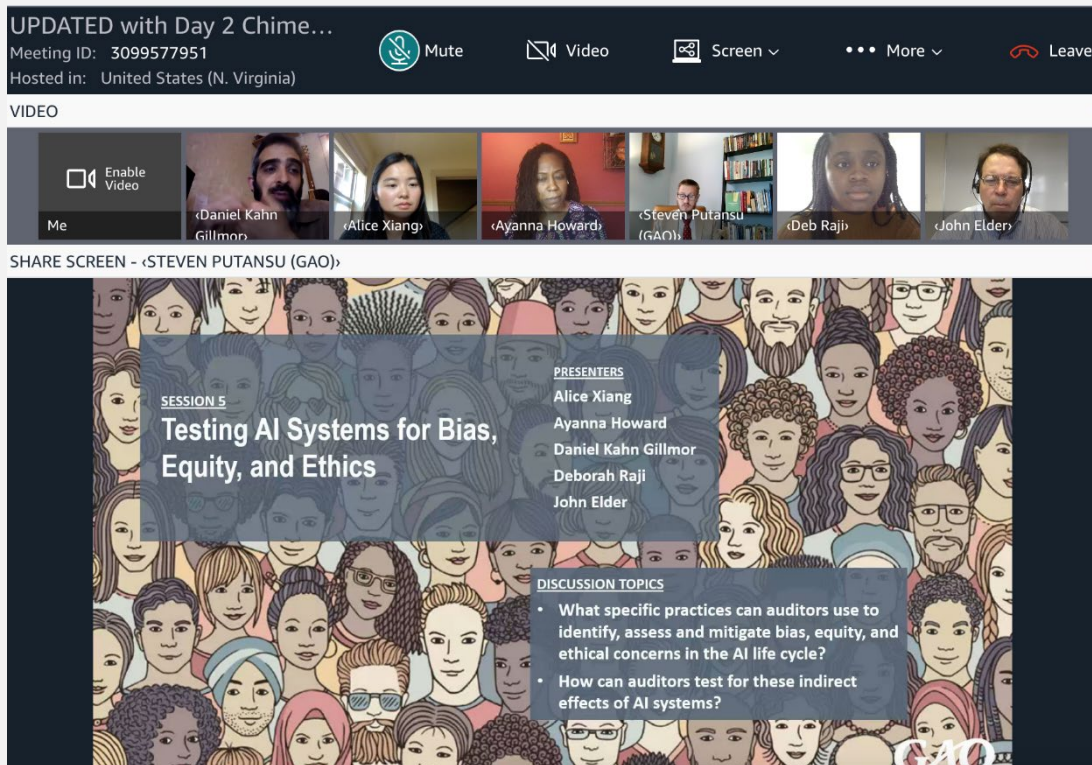# ARTIFICIAL INTELLIGENCE ACCOUNTABILITY FRAMEWORK

Farahnaaz Khakoo-Mausel
Assistant Director
Science Technology
Assessment & Analytics

Technology is advancing every day. We believe oversight should, too.

# Comptroller General Forum on AI Oversight: September 2020

- ➢ Factors to consider when auditing AI

- ➢ Criteria and challenges associated with auditing AI systems

- ➢ Challenges with using and auditing AI systems in the public sector

- ➢ Possible sources of evidence for auditing AI systems

- ➢ Testing AI systems for bias and equity

Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities

GAO: Oversight of federal Artificial Intelligence (AI) systems

The Phases in the AI Life Cycle

Design
involves articulating the system's concept and objectives, underlying assumptions, context and requirements, and potentially building a prototype.

Development
involves planning and design, including establishing technical requirements, data collection and processing, model building and interpretation, and system verification and validation.

Deployment
involves piloting, checking compatibility with legacy systems, ensuring regulatory compliance, managing organizational change, and evaluating user experience.

Continuous monitoring
involves operating the AI system and continuously assessing its recommendations and impacts (both intended and unintended) in light of objectives and ethical considerations. This phase identifies problems and adjusts by reverting to other phases or, if necessary, retiring the AI system from production.

We are still early in the AI journey. It is important that oversight is integrated into AI development now.
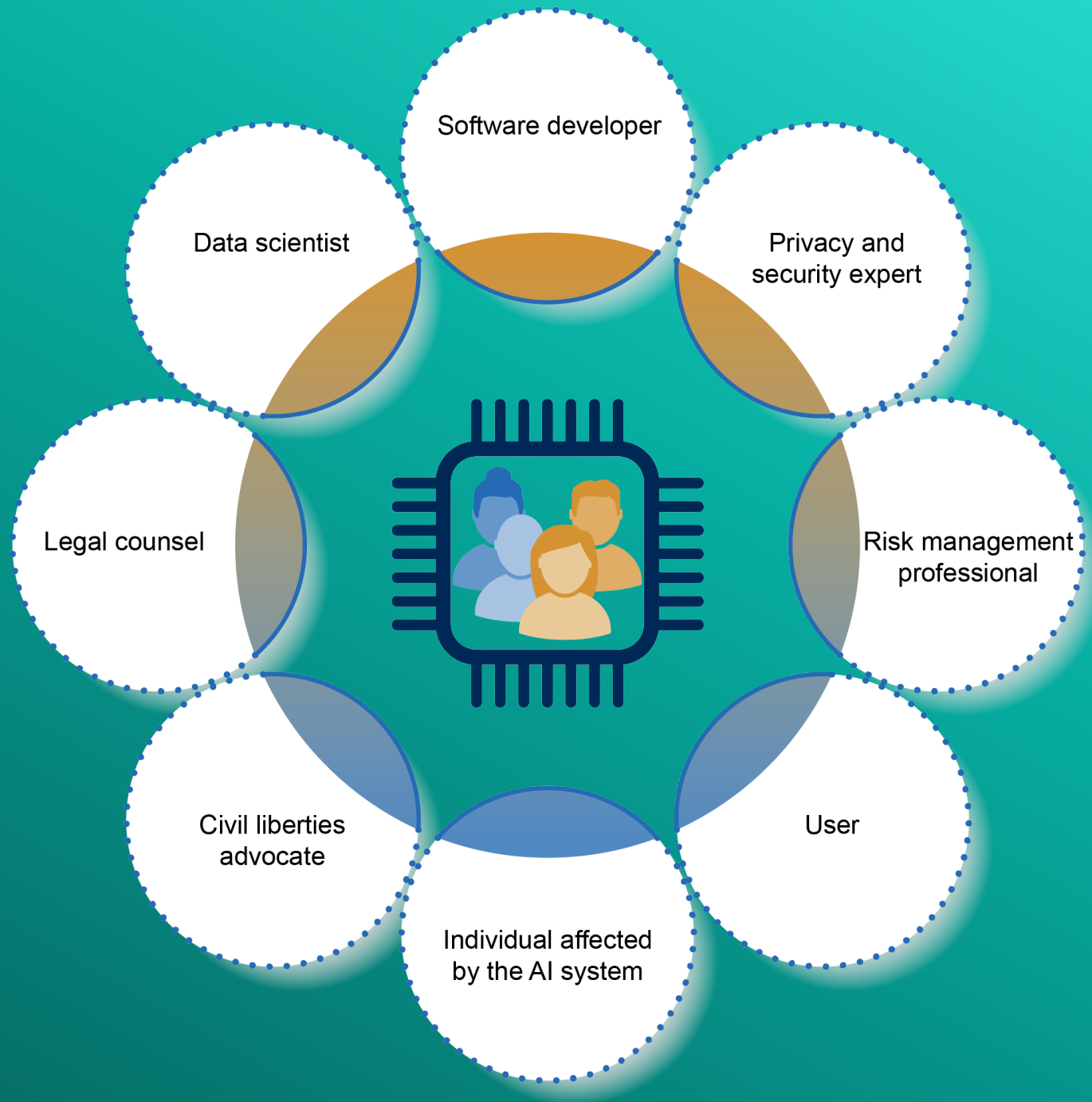
Data
Ensure quality, reliability, and representativeness of data sources and processing.

Data Used to Develop an AI Model
Entities should document sources and origins of data, ensure the reliability of data, and assess data attributes, variables, and augmentation/enhancement for appropriateness.

Data Used to Operate an AI System
Entities should assess the interconnectivities and dependencies of data streams that operationalize an AI system, identify potential biases, and assess data security and privacy.

Monitoring
Ensure reliability and relevance over time.

Continuous Monitoring of Performance
Entities should develop plans for continuous or routine monitoring of the AI system and document results and corrective actions taken to ensure the system produces desired results.

Assessing Sustainment and Expanded Use
Entities should assess the utility of the AI system to ensure its relevance and identify conditions under which the AI system may or may not be scaled or expanded beyond its current use.

Governance
Promote accountability by establishing processes to manage, operate, and oversee implementation.

Governance at the Organizational Level
Entities should define clear goals, roles, and responsibilities, demonstrate values and principles to foster trust, develop a competent workforce, engage stakeholders with diverse perspectives to mitigate risks, and implement an AI-specific risk management plan.

Governance at the System Level
Entities should establish technical specifications to ensure the AI system meets its intended purpose and complies with relevant laws, regulations, standards, and guidance. Entities should promote transparency by enabling external stakeholders to access information on the AI system.
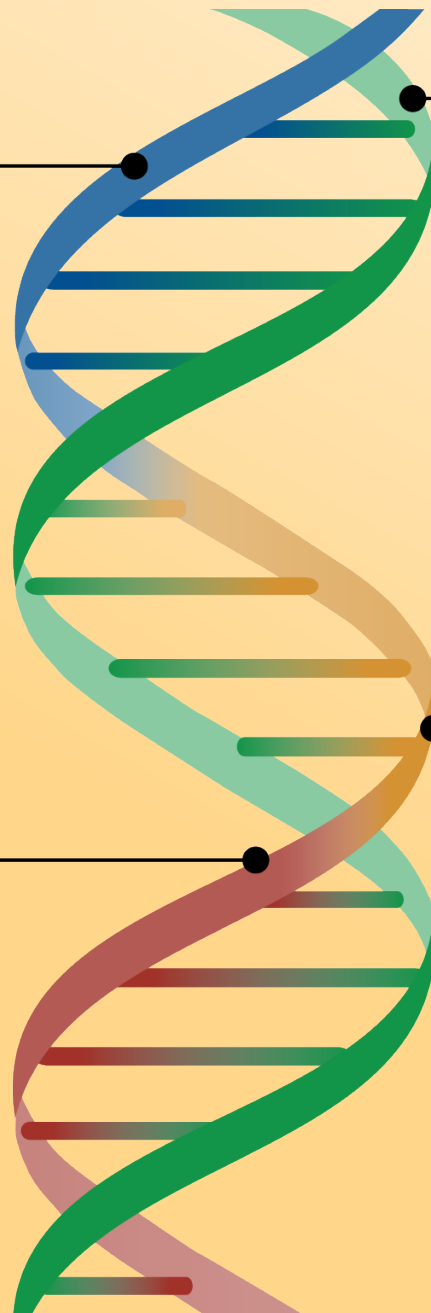
Performance
Produce results that are consistent with program objectives.
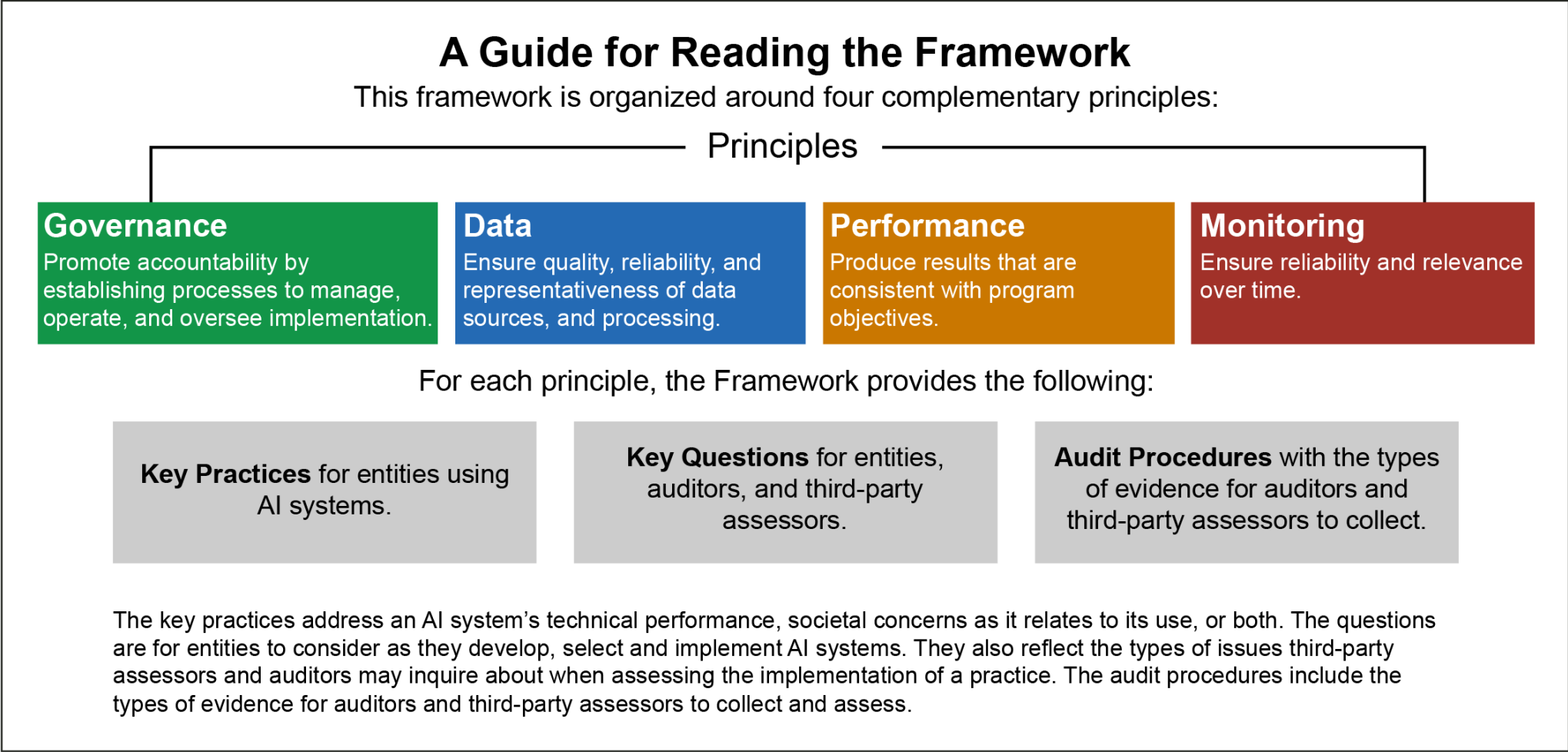
Performance at the Component Level
Entities should catalog model and non-model components that make up the AI system, define metrics, and assess performance and outputs of each component.

Performance at the System Level
Entities should define metrics and assess performance of the AI system. In addition, entities should document methods for assessment, performance metrics, and outcomes; identify potential biases; and define and develop procedures for human supervision of the AI system.
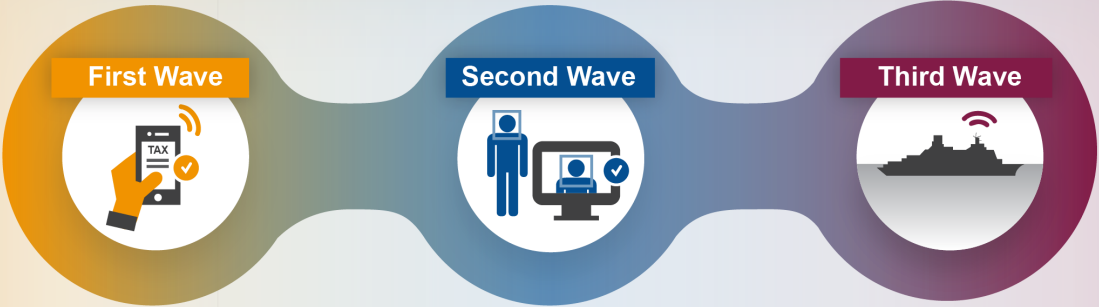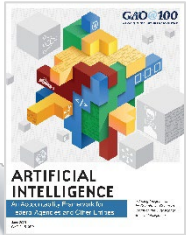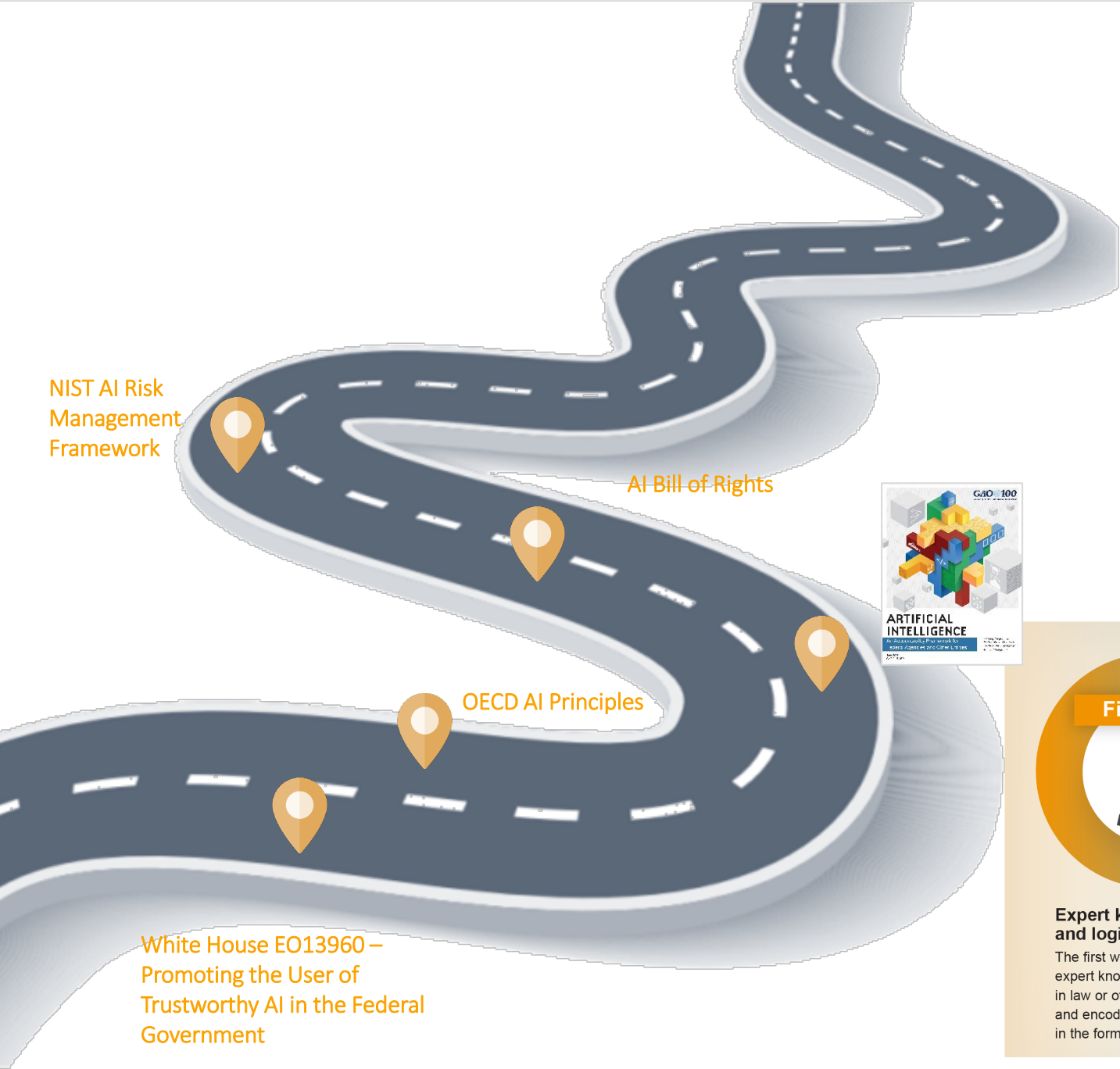
GAO

How to Use the AI Accountability Framework

A Guide for Reading the Framework

This framework is organized around four complementary principles:

Principles

Governance
Promote accountability by establishing processes to manage, operate, and oversee implementation.

Data
Ensure quality, reliability, and representativeness of data sources, and processing.

Performance
Produce results that are consistent with program objectives.

Monitoring
Ensure reliability and relevance over time.

For each principle, the Framework provides the following:

Key Practices for entities using AI systems.

Key Questions for entities, auditors, and third-party assessors.

Audit Procedures with the types of evidence for auditors and third-party assessors to collect.

The key practices address an AI system's technical performance, societal concerns as it relates to its use, or both. The questions are for entities to consider as they develop, select and implement AI systems. They also reflect the types of issues third-party assessors and auditors may inquire about when assessing the implementation of a practice. The audit procedures include the types of evidence for auditors and third-party assessors to collect and assess.

Source: GAO. | GAO-21-519SP

Where do we go from here?

NIST AI Risk Management Framework

AI Bill of Rights

OECD AI Principles

White House EO13960 – Promoting the User of Trustworthy AI in the Federal Government

ARTIFICIAL INTELLIGENCE

First Wave

Expert knowledge or criteria and logical reasoning
The first wave of AI is represented by expert knowledge or criteria developed in law or other authoritative sources and encoded into a computer program in the form of an expert system.

Second Wave

Machine/Statistical learning
Second-wave AI technology is based on machine learning, or statistical learning, and includes voice recognition, natural-language processing, and computer-vision technologies, among others.

Third Wave

Contextual adaptation
Third-wave AI technology combines the strengths of first- and second-wave AI, and is also capable of contextual sophistication, abstraction, and explanation.
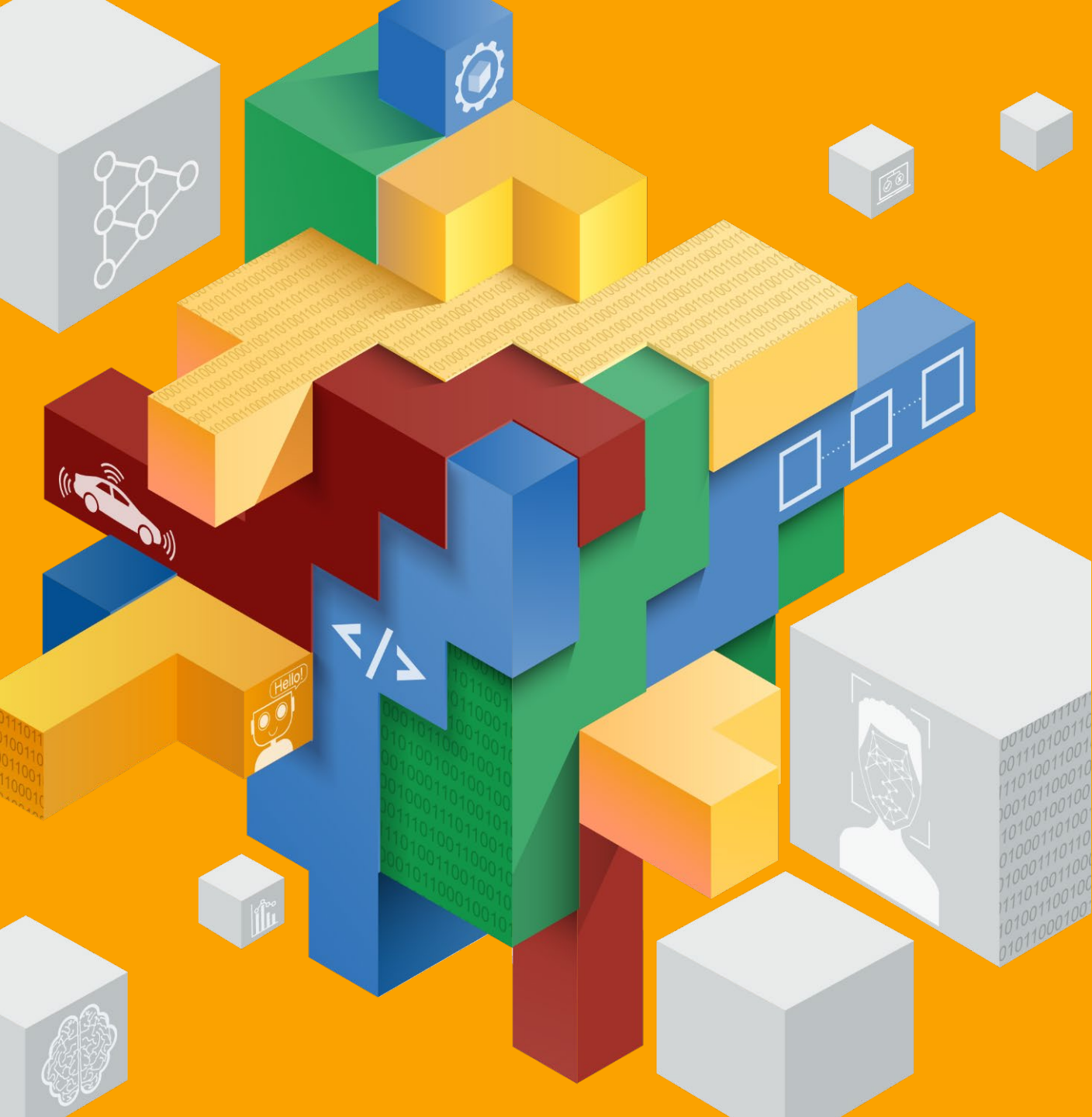
ONGOING OVERSIGHT, INSIGHT, AND FORESIGHT WORK RELATED TO AI

A growing portfolio of in-depth GAO work related to AI

Q & A