NIST AIRC - Playbook

Type	Title	AI Actors	Topics	Description
Govern	Govern 1.1	Governance and Oversight	Legal and Regulatory, Governance	Legal and regulatory requirements involving AI are understood, managed, and documented.
Govern	Govern 1.2	Governance and Oversight	Trustworthy Characteristics, Governance, Validity and Reliability, Safety, Secure and Resilient, Accountability and Transparency, Explainability and Interpretability, Privacy, Fairness and Bias	The characteristics of trustworthy AI are integrated into organizational policies, processes, and procedures.
Govern	Govern 1.3	Governance and Oversight	Risk Tolerance, Governance	Processes and procedures are in place to determine the needed level of risk management activities based on the organization's risk tolerance.
Govern	Govern 1.4	Governance and Oversight	Risk Management, Governance, Documentation	The risk management process and its outcomes are established through transparent policies, procedures, and other controls based on organizational risk priorities.
Govern	Govern 1.5	Governance and Oversight, Operation and Monitoring	Monitoring, Governance	Ongoing monitoring and periodic review of the risk management process and its outcomes are planned, organizational roles and responsibilities are clearly defined, including determining the frequency of periodic review.
Govern	Govern 1.6	Governance and Oversight	Risk Management, Governance, Data, Documentation	Mechanisms are in place to inventory AI systems and are resourced according to organizational risk priorities.
Govern	Govern 1.7	AI Deployment, Operation and Monitoring	Decommission, Governance	Processes and procedures are in place for decommissioning and phasing out of AI systems safely and in a manner that does not increase risks or decrease the organization’s trustworthiness.
Govern	Govern 2.1	Governance and Oversight	Governance, Risk Culture	Roles and responsibilities and lines of communication related to mapping, measuring, and managing AI risks are documented and are clear to individuals and teams throughout the organization.
Govern	Govern 2.2	Governance and Oversight	Governance, Training	The organization’s personnel and partners receive AI risk management training to enable them to perform their duties and responsibilities consistent with related policies, procedures, and agreements.
Govern	Govern 2.3	Governance and Oversight	Governance, Risk Tolerance	Executive leadership of the organization takes responsibility for decisions about risks associated with AI system development and deployment.
Govern	Govern 3.1	Governance and Oversight, AI Design	Diversity, Interdisciplinarity, Governance	Decision-makings related to mapping, measuring, and managing AI risks throughout the lifecycle is informed by a diverse team (e.g., diversity of demographics, disciplines, experience, expertise, and backgrounds).
Govern	Govern 3.2	AI Design	Human-AI teaming, Human oversight, Governance	Policies and procedures are in place to define and differentiate roles and

Type	Title	AI Actors	Topics	Description
				responsibilities for human-AI configurations and oversight of AI systems.
Govern	Govern 4.1	AI Design, AI Development, AI Deployment, Operation and Monitoring	Risk Culture, Governance, Adversarial	Organizational policies, and practices are in place to foster a critical thinking and safety-first mindset in the design, development, deployment, and uses of AI systems to minimize negative impacts.
Govern	Govern 4.2	AI Design, AI Development, AI Deployment, Operation and Monitoring	Risk Culture, Governance, Impact Assessment	Organizational teams document the risks and potential impacts of the AI technology they design, develop, deploy, evaluate and use, and communicate about the impacts more broadly.
Govern	Govern 4.3	TEVV, Operation and Monitoring, Governance and Oversight, Fairness and Bias	Risk Culture, Governance, AI Incidents, Impact Assessment, Drift, Fairness and Bias	Organizational practices are in place to enable AI testing, identification of incidents, and information sharing.
Govern	Govern 5.1	AI Design, Governance and Oversight, AI Impact Assessment, Affected Individuals and Communities	Participation, Governance, Impact Assessment	Organizational policies and practices are in place to collect, consider, prioritize, and integrate feedback from those external to the team that developed or deployed the AI system regarding the potential individual and societal impacts related to AI risks.
Govern	Govern 5.2	AI Impact Assessment, Governance and Oversight, Operation and Monitoring	Participation, Governance, Impact Assessment	Mechanisms are established to enable AI actors to regularly incorporate adjudicated feedback from relevant AI actors into system design and implementation.
Govern	Govern 6.1	Third-party entities, Operation and Monitoring, Procurement	Third-party, Legal and Regulatory, Procurement, Supply Chain, Governance	Policies and procedures are in place that address AI risks associated with third-party entities, including risks of infringement of a third party's intellectual property or other rights.
Govern	Govern 6.2	AI Deployment, TEVV, Operation and Monitoring, Third-party entities	Third-party, Governance, Risk Management, Supply Chain	Contingency processes are in place to handle failures or incidents in third-party data or AI systems deemed to be high-risk.
Manage	Manage 1.1	AI Deployment, Operation and Monitoring, AI Impact Assessment	AI Deployment, Risk Assessment	A determination is made as to whether the AI system achieves its intended purpose and stated objectives and whether its development or deployment should proceed.
Manage	Manage 1.2	AI Deployment, Operation and Monitoring, AI Impact Assessment	Risk Tolerance	Treatment of documented AI risks is prioritized based on impact, likelihood, or available resources or methods.
Manage	Manage 1.3	AI Deployment, Operation and Monitoring, AI Impact Assessment	Legal and Regulatory, Risk Tolerance	Responses to the AI risks deemed high priority as identified by the Map function, are developed, planned, and documented. Risk response options can include mitigating, transferring, avoiding, or accepting.
Manage	Manage 1.4	AI Deployment, Operation and Monitoring, AI Impact Assessment	Risk Response	Negative residual risks (defined as the sum of all unmitigated risks) to both downstream acquirers of AI systems and end users are documented.
Manage	Manage	AI Deployment,	Risk Tolerance, Trade-offs	Resources required to manage AI risks are

Type	Title	AI Actors	Topics	Description
	2.1	Operation and Monitoring, AI Impact Assessment, Governance and Oversight		taken into account, along with viable non-AI alternative systems, approaches, or methods – to reduce the magnitude or likelihood of potential impacts.
Manage	Manage 2.2	AI Deployment, Operation and Monitoring, AI Impact Assessment, Governance and Oversight	AI Deployment, Drift, Societal Values	Mechanisms are in place and applied to sustain the value of deployed AI systems.
Manage	Manage 2.3	AI Deployment, Operation and Monitoring	Risk Response	Procedures are followed to respond to and recover from a previously unknown risk when it is identified.
Manage	Manage 2.4	AI Deployment, Operation and Monitoring, Governance and Oversight	Risk Response, Decommission, Risky Emergent Behavior	Mechanisms are in place and applied, responsibilities are assigned and understood to supersede, disengage, or deactivate AI systems that demonstrate performance or outcomes inconsistent with intended use.
Manage	Manage 3.1	Third-party entities, Operation and Monitoring, AI Deployment	Third-party, Supply Chain	AI risks and benefits from third-party resources are regularly monitored, and risk controls are applied and documented.
Manage	Manage 3.2	Third-party entities, Operation and Monitoring, AI Deployment	Pre-trained models, Monitoring	Pre-trained models which are used for development are monitored as part of AI system regular monitoring and maintenance.
Manage	Manage 4.1	AI Deployment, Operation and Monitoring, End-Users, Human Factors, Domain Experts, Affected Individuals and Communities	Monitoring, Participation, AI Deployment, AI Incidents, Risk Response, Adversarial, Risky Emergent Behavior	Post-deployment AI system monitoring plans are implemented, including mechanisms for capturing and evaluating input from users and other relevant AI actors, appeal and override, decommissioning, incident response, recovery, and change management.
Manage	Manage 4.2	TEVV, AI Design, AI Development, AI Deployment, Operation and Monitoring, End-Users, Affected Individuals and Communities	Monitoring, Impact Assessment, Risk Assessment, Continual Improvement	Measurable activities for continual improvements are integrated into AI system updates and include regular engagement with interested parties, including relevant AI actors.
Manage	Manage 4.3	AI Deployment, Operation and Monitoring, End-Users, Human Factors, Domain Experts, Affected Individuals and Communities	AI Incidents, Monitoring	Incidents and errors are communicated to relevant AI actors including affected communities. Processes for tracking, responding to, and recovering from incidents and errors are followed and documented.
Map	Map 1.1		Socio-technical systems, Societal Values, Context of Use, Impact Assessment, TEVV, Trustworthy Characteristics, Validity and Reliability, Safety, Secure	Intended purpose, potentially beneficial uses, context-specific laws, norms and expectations, and prospective settings in which the AI system will be deployed are understood and documented. Considerations include: specific set or

Type	Title	AI Actors	Topics	Description
			and Resilient, Accountability and Transparency, Explainability and Interpretability, Privacy, Fairness and Bias	types of users along with their expectations; potential positive and negative impacts of system uses to individuals, communities, organizations, society, and the planet; assumptions and related limitations about AI system purposes; uses and risks across the development or product AI lifecycle; TEVV and system metrics.
Map	Map 1.2		Diversity, Interdisciplinarity, Socio-technical systems	Inter-disciplinary AI actors, competencies, skills and capacities for establishing context reflect demographic diversity and broad domain and user experience expertise, and their participation is documented. Opportunities for interdisciplinary collaboration are prioritized.
Map	Map 1.3		Socio-technical systems, Societal Values	The organization's mission and relevant goals for the AI technology are understood and documented.
Map	Map 1.4		Context of Use	The business value or context of business use has been clearly defined or – in the case of assessing existing AI systems – re-evaluated.
Map	Map 1.5		Risk Tolerance	Organizational risk tolerances are determined and documented.
Map	Map 1.6		Socio-technical systems, Impact Assessment, Documentation	System requirements (e.g., "the system shall respect the privacy of its users") are elicited from and understood by relevant AI actors. Design decisions take socio-technical implications into account to address AI risks.
Map	Map 2.1		Socio-technical systems	The specific task, and methods used to implement the task, that the AI system will support is defined (e.g., classifiers, generative models, recommenders).
Map	Map 2.2		Limitations, Human oversight, Impact Assessment, Documentation	Information about the AI system's knowledge limits and how system output may be utilized and overseen by humans is documented. Documentation provides sufficient information to assist relevant AI actors when making informed decisions and taking subsequent actions.
Map	Map 2.3	AI Development, TEVV, Domain Experts	TEVV, Data, Impact Assessment, Limitations	Scientific integrity and TEVV considerations are identified and documented, including those related to experimental design, data collection and selection (e.g., availability, representativeness, suitability), system trustworthiness, and construct validation.
Map	Map 3.1	AI Development, AI Deployment, AI Impact Assessment	Socio-technical systems, Documentation	Potential benefits of intended AI system functionality and performance are examined and documented.
Map	Map 3.2	AI Design, AI Development, Operation and Monitoring, AI Design, AI Impact Assessment	Impact Assessment, Trustworthy Characteristics, Validity and Reliability, Safety, Secure and Resilient, Accountability and Transparency, Explainability	Potential costs, including non-monetary costs, which result from expected or realized AI errors or system functionality and trustworthiness - as connected to organizational risk tolerance - are examined and documented.

Type	Title	AI Actors	Topics	Description
			and Interpretability, Privacy, Fairness and Bias	
Map	Map 3.3	AI Design, AI Development, Human Factors	Context of Use, Documentation	Targeted application scope is specified and documented based on the system's capability, established context, and AI system categorization.
Map	Map 3.4	AI Design, AI Development, Human Factors, End-Users, Domain Experts, Operation and Monitoring	Human-AI teaming	Processes for operator and practitioner proficiency with AI system performance and trustworthiness – and relevant technical standards and certifications – are defined, assessed and documented.
Map	Map 3.5	Human Factors, End-Users, Domain Experts, Operation and Monitoring, AI Design	Human oversight	Processes for human oversight are defined, assessed, and documented in accordance with organizational policies from GOVERN function.
Map	Map 4.1	Third-party entities, Procurement, Operation and Monitoring, Governance and Oversight	Legal and Regulatory, Third-party, Pre-trained models, Supply Chain, Risk Tolerance, Risky Emergent Behavior	Approaches for mapping AI technology and legal risks of its components – including the use of third-party data or software – are in place, followed, and documented, as are risks of infringement of a third-party's intellectual property or other rights.
Map	Map 4.2	AI Deployment, TEVV, Operation and Monitoring, Third-party entities	Third-party, Pre-trained models	Internal risk controls for components of the AI system including third-party AI technologies are identified and documented.
Map	Map 5.1	AI Design, AI Development, AI Deployment, AI Impact Assessment, Operation and Monitoring, Affected Individuals and Communities, End-Users	Participation, Impact Assessment	Likelihood and magnitude of each identified impact (both potentially beneficial and harmful) based on expected use, past uses of AI systems in similar contexts, public incident reports, feedback from those external to the team that developed or deployed the AI system, or other data are identified and documented.
Map	Map 5.2	AI Design, Human Factors, AI Deployment, AI Impact Assessment, Operation and Monitoring, Domain Experts, Affected Individuals and Communities, End-Users	Participation, Impact Assessment	Practices and personnel for supporting regular engagement with relevant AI actors and integrating feedback about positive, negative, and unanticipated impacts are in place and documented.
Measure	Measure 1.1	AI Development, TEVV, Domain Experts	Trustworthy Characteristics, Risk Assessment, Risky Emergent Behavior, TEVV, Validity and Reliability, Safety, Secure and Resilient, Accountability and Transparency, Explainability and Interpretability, Privacy, Fairness and Bias	Approaches and metrics for measurement of AI risks enumerated during the Map function are selected for implementation starting with the most significant AI risks. The risks or trustworthiness characteristics that will not – or cannot – be measured are properly documented.
Measure	Measure 1.2	TEVV, AI Impact Assessment, AI	Impact Assessment, TEVV, Context of Use	Appropriateness of AI metrics and effectiveness of existing controls is

Type	Title	AI Actors	Topics	Description
		Development, AI Deployment, Affected Individuals and Communities		regularly assessed and updated including reports of errors and impacts on affected communities.
Measure	Measure 1.3	TEVV, AI Impact Assessment, AI Development, AI Deployment, Affected Individuals and Communities, Domain Experts, End-Users, Operation and Monitoring	Participation, Impact Assessment, Context of Use	Internal experts who did not serve as front-line developers for the system and/or independent assessors are involved in regular assessments and updates. Domain experts, users, AI actors external to the team that developed or deployed the AI system, and affected communities are consulted in support of assessments as necessary per organizational risk tolerance.
Measure	Measure 2.1	TEVV	TEVV, Documentation, Validity and Reliability	Test sets, metrics, and details about the tools used during test, evaluation, validation, and verification (TEVV) are documented.
Measure	Measure 2.2	TEVV, Human Factors, AI Development	Data, Human Subjects Protection	Evaluations involving human subjects meet applicable requirements (including human subject protection) and are representative of the relevant population.
Measure	Measure 2.3	TEVV, AI Deployment	TEVV, Impact Assessment	AI system performance or assurance criteria are measured qualitatively or quantitatively and demonstrated for conditions similar to deployment setting(s). Measures are documented.
Measure	Measure 2.4	AI Deployment, TEVV	TEVV, Monitoring, Drift	The functionality and behavior of the AI system and its components – as identified in the MAP function – are monitored when in production.
Measure	Measure 2.5	TEVV, Domain Experts	TEVV, Validity and Reliability, Trustworthy Characteristics, Data	The AI system to be deployed is demonstrated to be valid and reliable. Limitations of the generalizability beyond the conditions under which the technology was developed are documented.
Measure	Measure 2.6	TEVV, Domain Experts, Operation and Monitoring, AI Impact Assessment, AI Deployment	TEVV, Safety, Trustworthy Characteristics, Context of Use	AI system is evaluated regularly for safety risks – as identified in the MAP function. The AI system to be deployed is demonstrated to be safe, its residual negative risk does not exceed the risk tolerance, and can fail safely, particularly if made to operate beyond its knowledge limits. Safety metrics implicate system reliability and robustness, real-time monitoring, and response times for AI system failures.
Measure	Measure 2.7	TEVV, Domain Experts, Operation and Monitoring, AI Impact Assessment, AI Deployment	TEVV, Secure and Resilient, Trustworthy Characteristics, Adversarial, Risky Emergent Behavior	AI system security and resilience – as identified in the MAP function – are evaluated and documented.
Measure	Measure 2.8	TEVV, Domain Experts, Operation and Monitoring, AI Impact Assessment, AI Deployment	TEVV, Accountability and Transparency, Trustworthy Characteristics	Risks associated with transparency and accountability – as identified in the MAP function – are examined and documented.

Type	Title	AI Actors	Topics	Description
Measure	Measure 2.9	TEVV, Domain Experts, Operation and Monitoring, AI Impact Assessment, AI Deployment, End-Users	TEVV, Explainability and Interpretability, Trustworthy Characteristics	The AI model is explained, validated, and documented, and AI system output is interpreted within its context – as identified in the MAP function – and to inform responsible use and governance.
Measure	Measure 2.10	TEVV, Domain Experts, Operation and Monitoring, AI Impact Assessment, AI Deployment, End-Users	TEVV, Privacy, Trustworthy Characteristics	Privacy risk of the AI system – as identified in the MAP function – is examined and documented.
Measure	Measure 2.11	TEVV, Domain Experts, Operation and Monitoring, AI Impact Assessment, AI Deployment, End-Users, Affected Individuals and Communities	TEVV, Fairness and Bias, Trustworthy Characteristics	Fairness and bias – as identified in the MAP function – is evaluated and results are documented.
Measure	Measure 2.12	TEVV, Domain Experts, Operation and Monitoring, AI Impact Assessment, AI Deployment	TEVV, Environmental Impact	Environmental impact and sustainability of AI model training and management activities – as identified in the MAP function – are assessed and documented.
Measure	Measure 2.13	TEVV, AI Deployment, Operation and Monitoring	TEVV, Effectiveness	Effectiveness of the employed TEVV metrics and processes in the MEASURE function are evaluated and documented.
Measure	Measure 3.1	TEVV, AI Impact Assessment, Operation and Monitoring	TEVV, Monitoring, Continual Improvement	Approaches, personnel, and documentation are in place to regularly identify and track existing, unanticipated, and emergent AI risks based on factors such as intended and actual performance in deployed contexts.
Measure	Measure 3.2	TEVV, Domain Experts, AI Impact Assessment, Operation and Monitoring	Monitoring	Risk tracking approaches are considered for settings where AI risks are difficult to assess using currently available measurement techniques or where metrics are not yet available.
Measure	Measure 3.3	TEVV, AI Deployment, Operation and Monitoring, End-Users, Affected Individuals and Communities	Participation, Contestability, TEVV, Impact Assessment	Feedback processes for end users and impacted communities to report problems and appeal system outcomes are established and integrated into AI system evaluation metrics.
Measure	Measure 4.1	TEVV, AI Deployment, Operation and Monitoring, End-Users, Affected Individuals and Communities	TEVV, Participation, Context of Use	Measurement approaches for identifying AI risks are connected to deployment context(s) and informed through consultation with domain experts and other end users. Approaches are documented.
Measure	Measure 4.2	TEVV, AI Deployment,	TEVV, Participation, Trustworthy Characteristics,	Measurement results regarding AI system trustworthiness in deployment context(s)

Type	Title	AI Actors	Topics	Description
		Domain Experts, Operation and Monitoring, End-Users	Validity and Reliability, Safety, Secure and Resilient, Accountability and Transparency, Explainability and Interpretability, Privacy, Fairness and Bias	and across AI lifecycle are informed by input from domain experts and other relevant AI actors to validate whether the system is performing consistently as intended. Results are documented.
Measure	Measure 4.3	TEVV, AI Deployment, Operation and Monitoring, End-Users, Affected Individuals and Communities	TEVV, Participation, Trustworthy Characteristics, Validity and Reliability, Safety, Secure and Resilient, Accountability and Transparency, Explainability and Interpretability, Privacy, Fairness and Bias	Measurable performance improvements or declines based on consultations with relevant AI actors including affected communities, and field data about context-relevant risks and trustworthiness characteristics, are identified and documented.