National Security Commission on AI Recommended Practices

		DoD Principles of AI Ethics				
		Responsible	Equitable	Traceable	Reliable	Governable
Core Values	**A1.** Employ technologies and operational policies for privacy, fairness, inclusion, human rights, and LOAC	×	×			×
	B1. Consider and document value considerations based on how tradeoffs with accuracy are handled	×	×	×	×	
	B2. Consider and document value considerations in systems that rely on representations of objective or utility functions	×	×	×	×	
	B3. Conduct documentation, reviews, and set limits per derived disallowed outcomes	×	×	×	×	×
Engineering	**1.** Refine design and development requirements, informed by the concept of operations and risk assessment	×	×	×	×	×
	2. Produce documentation of the AI lifecycle			×		
	3. Leverage infrastructure to support traceability, including auditability and forensics		×	×		
	4. For security and robustness, address intentional and unintentional failures				×	×
	5. Conduct red teaming				×	
System Performance	**A1.** Use regularly updated standards for testing and reporting system performance	×	×	×	×	
	A2. Consider and document the representativeness of data and model for the context at hand	×	×	×	×	
	A3. Evaluate an AI system's performance relative to current benchmarks			×		×
	A4. Evaluate aggregate performance of human-machine teams	×				
	A5. Provide sustained attention to reliability and robustness	×		×	×	
	A6. For systems of systems, test machine-machine/multi-agent interaction	×			×	
	B1. Specify maintenance requirements	×	×	×	×	
	B2. Continuously monitor and evaluate AI system performance	×	×	×	×	×
	B3. Conduct iterative and sustained testing and validation	×			×	×
	B4. Monitor and mitigate emergent behavior	×			×	×
Human-AI Interaction & Teaming	**A1.** Define the tasks of humans and the goals and mission of the human-machine team across the AI lifecycle	×		×		
	A2. Define functions and responsibilities of humans during system operation and assign them to individuals	×				
	B1. Extend human-AI design methodologies and guidelines	×	×	×		×
	B2. Employ algorithms and functions to support interpretability and explanation	×		×		×
	B3. Design systems to provide cues to operators about system confidence in its results or behaviors	×		×		×
	B4. Refine policies for machine-human handoff and control of initiative	×		×		×
	B5. Leverage traceability to assist with system development and understanding	×		×	×	×
	B6. Conduct training	×	×	×	×	×
Accountability & Governance	**1.** Appoint full-time responsible AI leads to join senior leadership	×		×		×
	2. Identify responsible actors	×		×		×
	3. Require technology to strengthen accountability processes and goals	×		×		×
	4. Adopt policies to strengthen accountability and governance	×		×		×
	5. Support external oversight			×		

National Security Commission on AI Recommended Practices

Category	Practice	Respect the Law; Act with Integrity	Transparent & Accountable	Objective & Equitable	Human-Centered Development & Use	Secure & Resilient	Informed by Science & Technology
Core Values	A1. Employ technologies and operational policies for privacy, fairness, inclusion, human rights, and LOAC	X		X	X		X
	B1. Consider and document value considerations based on how tradeoffs with accuracy are handled	X	X	X	X	X	X
	B2. Consider and document value considerations in systems that rely on representations of objective or utility functions	X	X	X	X	X	X
	B3. Conduct documentation, reviews, and set limits per derived disallowed outcomes	X	X	X	X	X	X
Engineering	1. Refine design and development requirements, informed by the concept of operations and risk assessment	X	X	X	X	X	X
	2. Produce documentation of the AI lifecycle	X			X	X	X
	3. Leverage infrastructure to support traceability, including auditability and forensics		X	X		X	X
	4. For security and robustness, address intentional and unintentional failures				X	X	
	5. Conduct red teaming				X	X	X
System Performance	A1. Use regularly updated standards for testing and reporting system performance	X	X	X		X	X
	A2. Consider and document the representativeness of data and model for the context at hand	X	X	X	X	X	X
	A3. Evaluate an AI system's performance relative to current benchmarks		X		X		X
	A4. Evaluate aggregate performance of human-machine teams		X		X		
	A5. Provide sustained attention to reliability and robustness		X			X	
	A6. For systems of systems, test machine-machine/multi-agent interaction					X	X
	B1. Specify maintenance requirements		X			X	X
	B2. Continuously monitor and evaluate AI system performance	X	X	X		X	X
	B3. Conduct iterative and sustained testing and validation		X			X	X
	B4. Monitor and mitigate emergent behavior					X	
Human-AI Interaction & Teaming	A1. Define the tasks of humans and the goals and mission of the human-machine team across the AI lifecycle		X		X		X
	A2. Define functions and responsibilities of humans during system operation and assign them to individuals		X		X		X
	B1. Extend human-AI design methodologies and guidelines		X		X		X
	B2. Employ algorithms and functions to support interpretability and explanation		X		X		X
	B3. Design systems to provide cues to operators about system confidence in its results or behaviors		X		X		X
	B4. Refine policies for machine-human handoff and control of initiative		X		X		X
	B5. Leverage traceability to assist with system development and understanding		X		X		X
	B6. Conduct training	X	X	X	X	X	X
Accountability & Governance	1. Appoint full-time responsible AI leads to join senior leadership	X	X				X
	2. Identify responsible actors	X	X				X
	3. Require technology to strengthen accountability processes and goals	X	X				X
	4. Adopt policies to strengthen accountability and governance	X	X				X
	5. Support external oversight	X	X				X