Responsible AI Guidelines

www.social-dynamics.net/rai-guidelines

CATEGORY	RESPONSIBLE AI GUIDELINE	ISO	AI ACT	PHASE	ROLE
Intended uses	**#1** Work with relevant parties to identify intended uses (e.g., identify the system's usage, deployment, and contextual conditions)	5338, 38507, 23894, 24027, 24368, 42001	Art. 6, 9	1-3	D, E, M
	#2 Obtain approval from an Ethics Committee or similar body for intended uses (e.g., obtain Ethics Committee approval for the intended use, aligned with sustainability goals)	38507, 5338, 23894, 42001	Art. 11, 69	1-3	D, E, M
Harms	**#3** Identify potential harms and risks associated with the intended uses (e.g., prevent privacy violation, discrimination, and adversarial attacks)	23894, 24028, 38507, 24368, 42001, 25059	Art. 9, 65	1-3	D, E, M
	#4 Provide mechanism(s) for incentivizing reporting of system harms (e.g., provide contact emails and feedback form to raise concerns)	38507, 23894, 42001	Art. 9, 60-63	1	D, E
	#5 Develop strategies to mitigate identified harms or risks for each intended use (e.g., use stratified sampling and safeguards against adversarial attacks during training)	24368, 23894, 42001, 25059	Art. 9, 67	1-3	D, E, M
System	**#6** Document all system components, including the AI models, to enable reproducibility and scrutiny (e.g., create UML diagrams and specify model types, versions, hardware architecture)	5338, 23894, 24027, 42001, 25059	Art. 11, 12, 16-18, 50	1-3	D, E
	#7 Review the code for reliability (e.g., manage version control using software)	5338, 25059	Art. 17	1-3	D, E
	#8 Report evaluation metrics for various groups based on factors such as age and ethnicity (e.g., evaluate false positive/negative and feature importance across protected attributes)	23894, 5338, 24028, 24027, 42001	Art. 10, 13	1-3	D, E, M
	#9 Provide mechanisms for interpretable outputs and auditing (e.g., output feature importance and provide human-understandable explanations)	38507, 23894, 42001	Art. 12-14	1-3	D, E, M
	#10 Document the security of all system components in consultation with experts (e.g., guard against adversarial attacks and unauthorized access)	24028, 24368, 42001, 25059	Art. 12, 13, 15, 17	1-3	E, M
	#11 Provide an environmental assessment of the system (e.g., report the number of GPU hours used in training and deployment)	38507, 23894, 5338, 24368, 42001, 25059	Art. 69	1-3	E
	#12 Develop feedback mechanisms to update the system (e.g., provide contact email, feedback form, and notification of new knowledge extracted)	24028, 42001	Art. 61	1-3	D, E
	#13 Ensure safe system decommissioning (e.g., ensure decommissioned data is either deleted or restricted to authorized personnel)	38507, 24368, 42001	Art. 9	3	E
	#14 Redocument model information and contractual requirements at every system update (e.g., update the model information when re-training the system)	23894, 5338, 24368, 42001	Art. 11, 12, 17, 61	3	E
Data	**#15** Ensure compliance with agreements and legal requirements when handling data (e.g., create data sharing and non-disclosure agreements and secure servers)	38507, 23894, 5338, 42001	Art. 10, 17, 61	1-3	D, E, M
	#16 Compare the quality, representativeness, and fit of training and testing datasets with the intended uses (e.g., report dataset details such as public/private, personal information)	38507, 5338, 24028, 24027, 42001, 25059	Art. 10, 13, 17, 64	1-3	E
	#17 Identify any measurement errors in input data and their associated assumptions (e.g., account for potential input errors in the input device, text data, audio, and video)	38507, 42001, 25059	Art. 10, 13, 17, 64	1-3	3
	#18 Protect sensitive variables in training/testing datasets (e.g., protect sensitive data and use techniques such as k-anonymity and differential privacy)	38507, 24028, 42001	Art. 10, 13, 17	1-3	D, E, M
Oversight	**#19** Continuously monitor metrics and utilize guardrails or rollbacks to ensure the system's output stays within a desired range (e.g., validate against concept drift and test)	38507, 5338, 24028, 24027, 24368, 42001	Art. 12, 20, 29, 61	1-3	D, E
	#20 Ensure human control over the system, particularly for designers, developers, and end-users (e.g., include human in the loop with the ability to inspect data, models, and training methods)	38507, 5338, 24028, 24368, 25059	Art. 13, 14	1-3	D, E, M
Team	**#21** Ensure team diversity (e.g., consider diversity in neurotypes and thinking styles)	38507, 5338, 24028, 24368, 42001	Art. 69	1-3	D, E, M
	#22 Train team members on ethical values and regulations (e.g., train on privacy regulations, ethical issues, and raising concerns)	38507, 24368, 42001	Art. 69	1-3	D, E, M

Responsible AI Guidelines are actionable items applied across the three phases of the AI development lifecycle: development (1), deployment (2), and use (3). These guidelines are grounded in scientific literature, verified for compliance with eight ISO standards, and cross-referenced with the EU AI Act. Each guideline specifies the responsible role: designer (D), engineer/researcher (E), or manager/executive (M).