

PUBBLICAZIONE, AI SENSI DELL'ART. 19 DEL D.LGS N. 33 DEL 14 MARZO 2013,
MODIFICATO DALL'ART. 18 DEL D.LGS N. 97 DEL 25 MAGGIO 2016 COME INTEGRATO
DALL'ART.1 C. 145 DELLA LEGGE 27 DICEMBRE 2019 N. 160,
DELLE TRACCE D'ESAME STABILITI DALLA COMMISSIONE ESAMINATRICE
DELLA SELEZIONE DI SEGUITO INDICATA
NELLA RIUNIONE IN DATA 29/04/2025

TRACCE DELLE PROVE D'ESAME – PROVA ORALE

BANDO N. 331.4 RIC ISTI

Selezione per titoli e colloquio ai sensi dell'art. 8 del *"Disciplinare concernente le assunzioni di personale con contratto di lavoro a tempo determinato"*, per l'assunzione, ai sensi dell'art. 141 del CCNL del Comparto "Istruzione e Ricerca" 2019-2021, sottoscritto in data 18 gennaio 2024, di una unità di personale con profilo professionale di **Ricercatore III livello**, presso l'Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo" di Pisa nell'ambito del Progetto "FAITH - Fostering Artificial Intelligence Trust for Humans towards the optimization of trustworthiness through large-scale pilots in critical domains" (CUP B53C23005500005).

BUSTA 1

PRIMA DOMANDA: Il candidato illustri cosa si intende per robustezza in un modello di intelligenza artificiale e perché è importante.

SECONDA DOMANDA: Il candidato descriva alcune tecniche per condurre un'analisi della sensitività su un modello di apprendimento automatico?

TERZA DOMANDA: Il candidato illustri quali pratiche consiglierebbe per testare e migliorare la generalizzabilità di un modello di apprendimento automatico?

QUARTA DOMANDA: Lettura e traduzione dalla lingua inglese dal testo "AI Risks and Trustworthiness" – dal Link <https://airc.nist.gov/airmf-resources/airmf/3-sec-characteristics/>

For AI systems to be trustworthy, they often need to be responsive to a multiplicity of criteria that are of value to interested parties. Approaches which enhance AI trustworthiness can reduce negative AI risks. This Framework articulates the following characteristics of trustworthy AI and offers guidance for addressing them. Characteristics of trustworthy AI systems include: valid and reliable, safe, secure and resilient, accountable and transparent, explainable and interpretable, privacy-enhanced, and fair with harmful bias managed. Creating trustworthy AI requires balancing each of these characteristics based on the AI system's context of use. While all characteristics are socio-technical system attributes, accountability and transparency also relate to the processes and activities internal to an AI system and its external setting. Neglecting these characteristics can increase the probability and magnitude of negative consequences.

BUSTA 2

PRIMA DOMANDA: Il candidato discuta quali metodi o tecniche possono essere utilizzati per valutare la robustezza di un modello di apprendimento automatico o intelligenza artificiale.

SECONDA DOMANDA: Il candidato esponga come definirebbe la sensitività in un contesto di apprendimento automatico. Può illustrare qualche esempio pratico?

TERZA DOMANDA: Il candidato descriva cosa significa che un modello di apprendimento automatico è generalizzabile e quali fattori influenzano la sua generalizzabilità.

QUARTA DOMANDA: Lettura e traduzione dalla lingua inglese dal testo "AI Risks and Trustworthiness" – dal Link <https://airc.nist.gov/airmf-resources/airmf/3-sec-characteristics/>

Trustworthiness characteristics are inextricably tied to social and organizational behavior, the datasets used by AI systems, selection of AI models and algorithms and the decisions made by those who build them, and the interactions with the humans who provide insight from and oversight of such systems. Human judgment should be employed when deciding on the specific metrics related to AI trustworthiness characteristics and the precise threshold values for those metrics.

Understanding and treatment of trustworthiness characteristics depends on an AI actor's particular role within the AI lifecycle. For any given AI system, an AI designer or developer may have a different perception of the characteristics than the deployer.

BUSTA 3

PRIMA DOMANDA: Il candidato discuta quali sono alcune delle principali sfide nel garantire la robustezza dei modelli di apprendimento automatico e come possono essere affrontate.

SECONDA DOMANDA: Il candidato illustri in che modo la presenza di outlier influisce sulla sensitività di un modello AI e come si possono gestire.

TERZA DOMANDA: Il candidato discuta come è possibile bilanciare la generalizzabilità con l'accuratezza su un set di dati di training specifico e quali sono le trade-off comuni.

QUARTA DOMANDA: Lettura e traduzione dalla lingua inglese dal testo "AI Risks and Trustworthiness" – dal Link <https://airc.nist.gov/airmf-resources/airmf/3-sec-characteristics/>

Addressing AI trustworthiness characteristics individually will not ensure AI system trustworthiness; tradeoffs are usually involved, rarely do all characteristics apply in every setting, and some will be more or less important in any given situation. Ultimately, trustworthiness is a social concept that ranges across a spectrum and is only as strong as its weakest characteristics. When managing AI risks, organizations can face difficult decisions in balancing these characteristics. For example, in certain scenarios tradeoffs may emerge between optimizing for interpretability and achieving privacy. In other cases, organizations might face a tradeoff between predictive accuracy and interpretability.

IL PRESIDENTE

Prof. Mario Giovanni Cosimo Antonio Cimino

IL SEGRETARIO

Sig. Riccardo Cini