Progetto cofinanziato da “Patti territoriali dell’alta formazione per le imprese” - CUP: F61B23000370006
Borsa cofinanziata dalla società BTNKEENG s.r.l.

Corso di Dottorato in Patrimoni Storici e Filosofici per un’Innovazione Sostenibile (XL ciclo) Università degli Studi di Bari “Aldo Moro”

Dottorando: Alessio Donvito **Tutor**: Prof. Francesco Marrone **Co-tutor:** Prof. Giovanni Vaia (UniVe)

TEORIE DELLA COMPLESSITÀ, RETI NEURALI E CERTIFICAZIONE DEGLI ALGORITMI NELL’INTELLIGENZA ARTIFICIALE

CONTESTO

Passaggio delle Tecnologie di Intelligenza Artificiale (**AIT**) dall’accademia all’industria (Maslej et al. 2024)

Corpora normativi per lo sviluppo di mercato e ambiente digitali orientati a **trasparenza** e ***fairness*** algoritmiche, ***privacy***, **responsabilità**, **sicurezza**

Obblighi di ***auditing*** dei sistemi di IA. Crescente richiesta di garanzie di sostenibilità per l’impiego delle AIT

CRITICITÀ

Generalità delle norme

↓

Traducibilità dei princìpi nell’effettiva analisi e certificazione delle AIT

↓

Variabilità applicativa, diversificazione tecnica interna

↓

Impatto etico-sociale AIT valutato frammentariamente. **Mancanza *framework*** analitici **standardizzati** e generalizzabili

Approcci di esclusiva natura normativa e tecnica

↓

Mancano il riconoscimento della **natura sociotecnica e culturale** degli algoritmi (***bias* umani e statistici**, **sistemi valoriali**, **stereotipi, distorsioni rappresentazionali)**

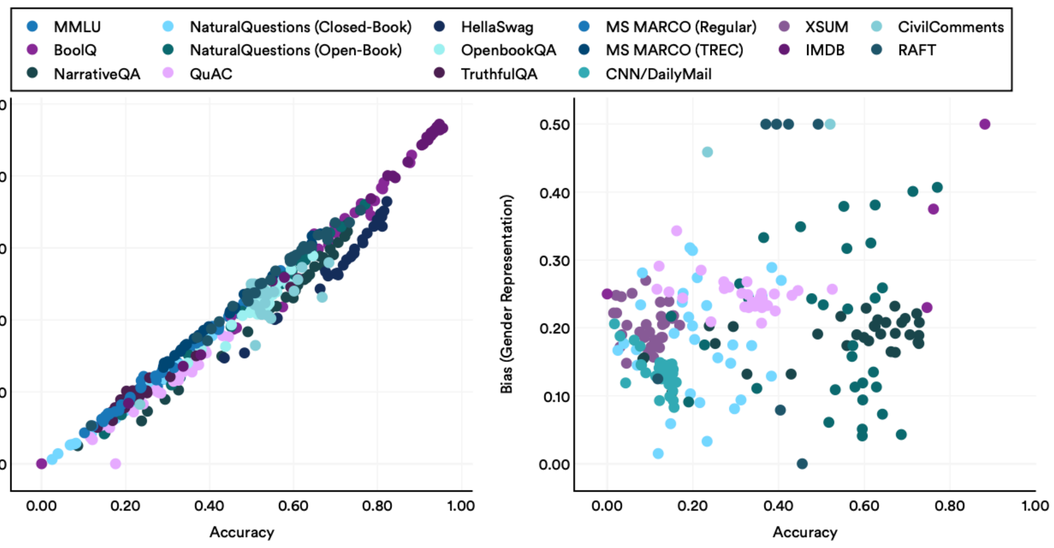

Fig. 1. Correlazione tra *performance* nelle metriche di *fairness* e *bias* rappresentazionali, *Maslej N. et al. (2024).*

OBIETTIVI DI RICERCA

1. Integrare le esistenti tassonomie di *bias* algoritmici: **studio critico della struttura matematico-computazionale** di un campione delle tecniche modellistiche di maggiore impiego applicativo.
2. **Operazionalizzare i princìpi** delineati nei *corpora* etico-normativi: sviluppo di un ***framework* di valutazione** del rischio algoritmico con caratteristiche di **sistematicità** e **robustezza contestuale**.

METODOLOGIA

1. **Studio comparativo** delle proposte di ***governance*** e delle **opzioni regolatorie** per le AIT; raccolta dei princìpi per la ***compliance* legale ed etica** dei sistemi.
2. Analisi con strumenti e concetti di **epistemologia dell’informazione** delle tecniche modellistiche: A) *Machine Learning* (Regressione Lineare, Alberi Decisionali, *Support Vector Machines* (SVM), *k-nearest neighbors* (KNN), Riduzione Dimensionale, *Principal Component Analysis*); B) *Deep Learning* (Reti Neurali Artificiali (ANN), Reti Neurali Convoluzionali (CNN), *Transformers-GPT.*
3. **Caso studio** per l’applicazione del *framework* di certificazione.

CRONOPROGRAMMA

1. **Ott. 2024 - Feb. 2025:** Ricognizione letteratura su trasparenza, *explainability* e *fairness*. Prima indagine su insufficienza delle metriche tecniche nella mitigazione dei *bias* (Fig. 1).

   Mar. - Dic. 2025: Analisi disegni di regolamentazione/standard di *auditing*: quanto si allineano ai princìpi (*digital ethics*) che inseguono? Studio critico strumenti modellistici (prima fase): quali ripercussioni sociotecniche? Presentazione risultati.

2. **Primo Semestre:** Periodo presso Btinkeeng. Sviluppo del *framework* (prima fase). Caso di studio. Presentazione risultati.

   Secondo Semestre: Soggiorno di studio presso l’Università di Granada. Analisi strumenti modellistici (seconda fase).

3. **Primo Semestre:** Completamento del *framework*. Scrittura tesi. Presentazione risultati.

   Secondo Semestre: Scrittura tesi. *Testing* del *framework.*

BIBLIOGRAFIA ESSENZIALE

- Fazelpour S., Danks D. (2021). *Algorithmic Bias*. Philosophy Compass, 16 (8).
- Floridi L. (2011), *The Philosophy of Information*. Oxford University Press.
- Floridi L. et al. (2018). *AI4People – An Ethical Framework for Good AI Society: Opportunities, Risks, Principles and Recommendations*. Minds and Machines, Vol. 28, pp. 689-707.
- Friedman B. et al. (2013). *Value Sensitive Design and Information Systems, Early engagement and new technologies: Opening up the laboratory*, Philosophy of Engineering and Technology, Springer Netherlands, pp. 55–95.
- Koshiyama A et al. (2024) *Towards algorithm auditing: managing legal, ethical and technological risks of AI, ML and associated algorithms*. R. Soc. Open Sci. 11: 230859.
- Maslej N. et al. (2024). *The AI Index 2024 Annual Report*, AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA.
- Mehrabi N. et al. (2021). *A survey on bias and fairness in Machine Learning*. ACM Computing Surveys, 54 (6).
- Russel S.J, Norvig P. (2021), *Artificial Intelligence. A Modern Approach*, 4a edizione, Pearson.