Intelligenza Artificiale: fiducia, trasparenza, regolamentazione, governance

Francesca Rossi

IBM Fellow e AI Ethics Global Leader

AAAI President-elect

Membro del board della Partnership on AI

Membro del comitato esecutivo della iniziativa IEEE sull'etica dell'AI

Esperto della Global Partnership on AI

Stato dell'arte in AI: capacita', limiti, e temi etici

Capacita'

Machine Learning

- Learning from data (Deep, Reinforced, Supervised/Unsupervised/Self Supervised)
- Hidden patterns in huge amounts of data
  - Prediction, perception tasks
  - Correlation, pattern discovery, data mining
- Flexible, can handle uncertainty

Approcci simbolici, logici, basati su regole

- Explicit procedure to solve a problem
- Reasoning, planning, scheduling, optimization for complex problems
- Symbolic, traceable, explainable

Limiti

- Generalizability and Abstraction
- Robustness and Resiliency
- Contextual awareness
- Multi-agent cooperation
- Resource efficiency (examples, energy, computing power)
- Adaptability
- Causality

Temi etici

- Trust
  - Fairness, robustness, explainability, causality, transparency
- Data governance, privacy, liability, human agency, impact on work and society
- AI autonomy vs augmented intelligence
- Real vs online life, metrics of success/goals

Libro bianco AI della Commissione Europea: rischi e regole

Documenti dell'HLEG

Approccio basato sul rischio

- Applicazioni (non tecnologie) ad alto rischio richiedono regole precise
- Imposte sugli attori rilevanti (i piu' adatti a contrastare il rischio)
- Non basato su settori ma sui casi di uso
- Ex-ante self-assessment + ex-post auditing ed enforcement

Applicazioni specifiche ad alto rischio

- Riconoscimento facciale / dati biometrici
- Face detection/ authentication/recognition: non lo stesso livello di rischio
- IBM non fornisce piu' riconoscimento facciale general purpose

Trustworthy AI

- Definita da Gruppo di Esperti AI della Comunita' Europea (HLEG)
  - Human agency e oversight, robustness e safety, privacy e data governance, trasparenza, fairness, wellbeing, accountability

Linee guida, regole, e trasparenza

- Assessment list
- Proposte di regole e investimenti
  - Generali e per tre settori: manifatturiero, pubblico, e sanita'

Standard e sand boxes

- Best practices condivise a livello globale
- Ambienti safe per sperimentare regole e processi multi-stakeholderper uno sviluppo e uso responsabile dell'AI

Armonizzazione e chiarezza della regolamentazione in Europa

- Importante per ridurre incertezza e aiutare le piccole/medie imprese

Trasparenza e self-assessment

Assessment list dell'HLEG

- Documentazione riguardo lo sviluppo e le caratteristiche di un sistema di trustworthy AI
- Impostazione e uso di processi aggiuntivi per bias/spiegazioni/ecc.
- Comunicazione trasparente a clienti, policy makers, utenti, cittadini

Uso dell'assessment list

- Self-assessment
- Maggiore consapevolezza/educazione degli sviluppatori
- Impostazione e uso di processi aggiuntivi per bias/spiegazioni/ecc.
- Comunicazione trasparente a clienti, policy makers, utenti, cittadini

IBM AI factsheets

- Trasparenza attraverso la documentazione
  - Bias, spiegazioni, robustezza, accuratezza, usi appropriati e non, …

Educazione e collaborazione

- Educazione/reskilling dei progettisti e sviluppatori
- Team diversificati
- Consultazione con stakeholders rilevanti

Iniziative open-source

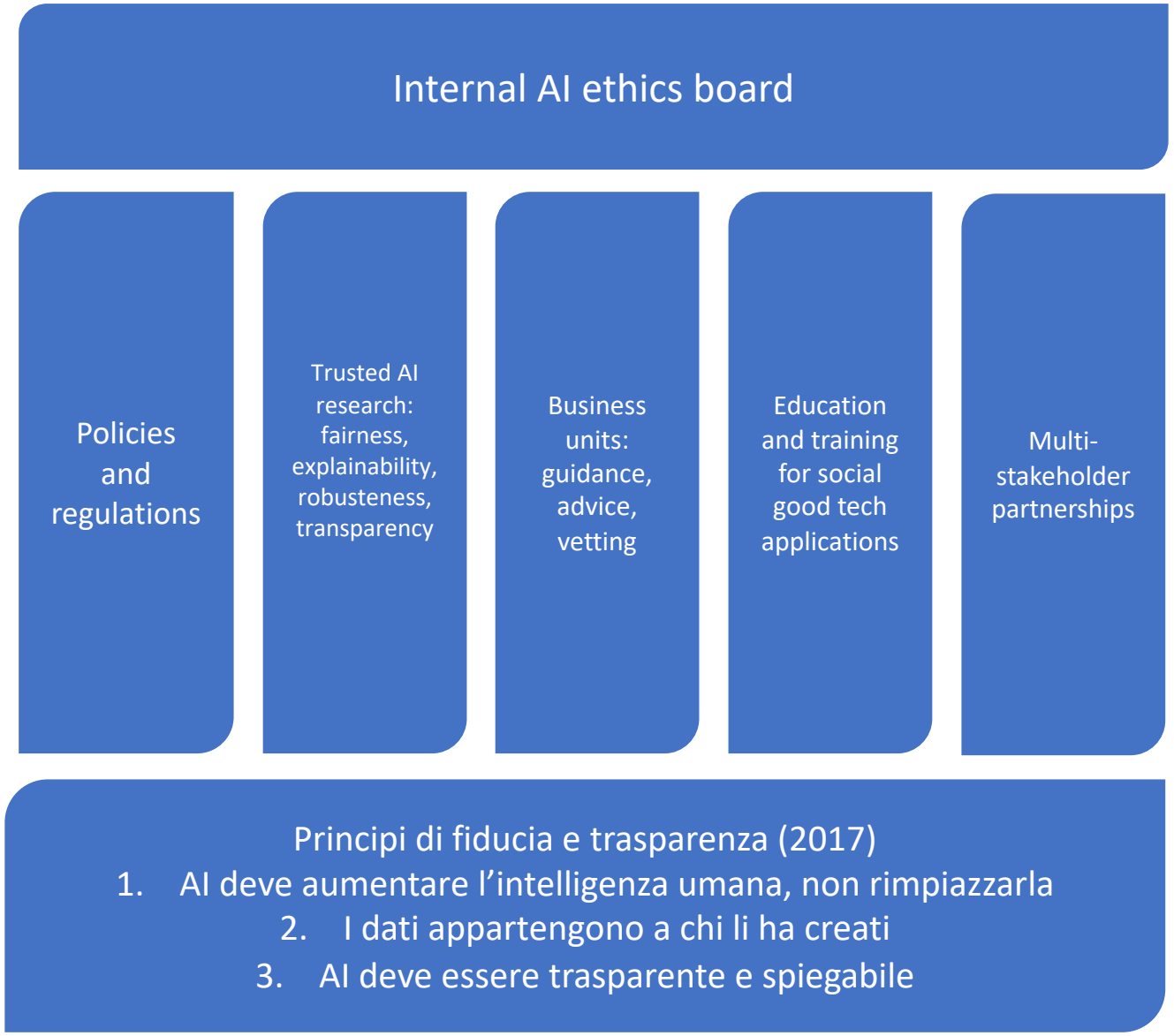
- Per educare e coinvolgere comunita' di "AI producers"
- Es.: IBM AI fairness 360, AI explainability 360, AI factsheet 360

Dai principi alla pratica: la governance dell'AI in una azienda

Principi per sviluppo e uso etico e responsabile dell'AI

- Educazione e reskilling
- Precise line guida
- Disseminazione e scaling
- Potere decisionale e incentivi
- Coinvolgimento e coordinamento
- Ricerca, sviluppo, servizi, piattaforme, soluzioni
- Collaborazione con stakeholder esterni
- Formazione e accompagnamento dell'intera azienda

Struttura della governance dell'etica dell'AI all'IBM



Link utili

- IBV study on “Advancing AI ethics beyond compliance”: https://www.ibm.com/thought-leadership/institute-business-value/report/ai-ethics
- Trusted AI for business: https://www.ibm.com/watson/ai-ethics/
- AI precision regulation: https://www.ibm.com/blogs/policy/ai-precision-regulation/
- Risposta a libro bianco AI della Commissione Europea: https://www.ibm.com/blogs/policy/ai-consultation-europe/
- Facial recognition: https://www.ibm.com/blogs/policy/facial-recognition/
- Response to COVID-19: https://www.ibm.com/thought-leadership/covid19/
- AI factsheet
  - website; https://aifs360.mybluemix.net/
  - paper: https://arxiv.org/pdf/2006.13796.pdf
- European Commission High Level Expert Group on AI on Trustworthy AI: https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence
- Partnership on AI: https://www.partnershiponai.org
- IEEE Global Initiative on Ethical Consideration on AI Systems: https://standards.ieee.org/industry-connections/ec/autonomous-systems.html
- Principled AI: https://cyber.harvard.edu/publication/2020/principled-ai