Corso di Laurea (Triennale) in Informatica

Fairness, Privacy, Ethics in sistemi di Machine Learning

Prof. Fabio Palomba
Dott. Carmine Ferrara

Thomas De Palma
Mat.:0512109541

t.depalma@studenti.unisa.it
https://github.com/andesrule
https://www.linkedin.com/in/thomas-de-palma-4459a1266

PDF Tesi —>

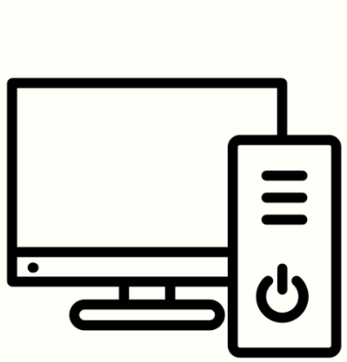
Introduzione e Background

I computer, che dal loro avvento hanno digitalizzato e migliorato le nostre vite, sono gli strumenti principali che hanno dato il via alla cosiddetta terza rivoluzione industriale.

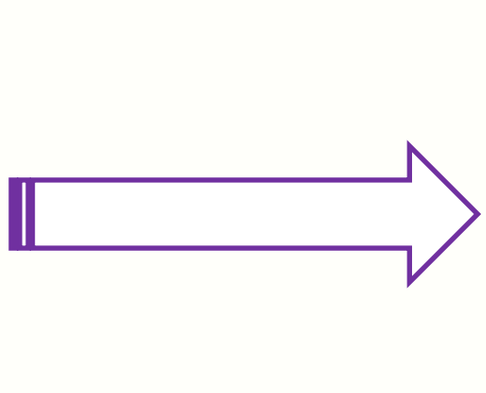
t.depalma@studenti.unisa.it
https://github.com/andesrule
https://www.linkedin.com/in/thomas-de-palma-4459a1266

Introduzione e Background

I computer, che dal loro avvento hanno digitalizzato e migliorato le nostre vite, sono gli strumenti principali che hanno dato il via alla cosiddetta terza rivoluzione industriale.



Sistemi informatici



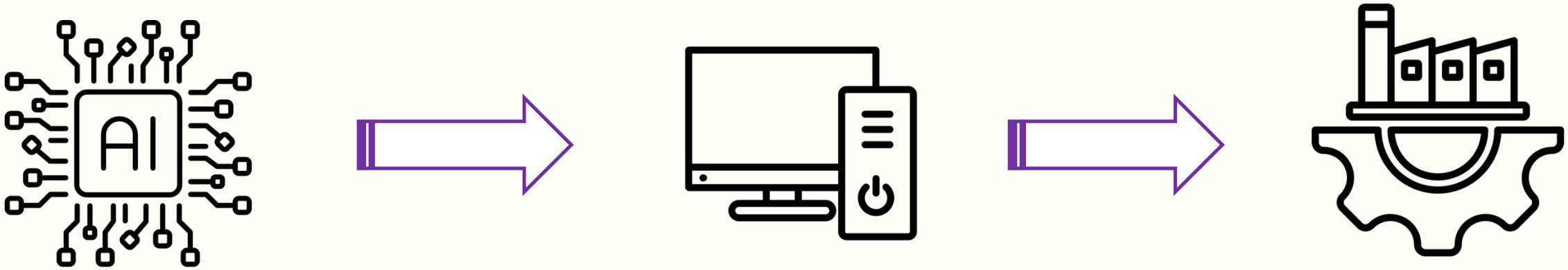
Rivoluzione digitale

t.depalma@studenti.unisa.it
https://github.com/andesrule
https://www.linkedin.com/in/thomas-de-palma-4459a1266

Introduzione e Background

I computer, che dal loro avvento hanno digitalizzato e migliorato le nostre vite, sono gli strumenti principali che hanno dato il via alla cosidetta terza rivoluzione industriale.

Un notevole contributo è stato dato dall’impiego sempre piú frequente di sistemi di intelligenza artificiale



Intelligenza Artificiale

Sistemi informatici

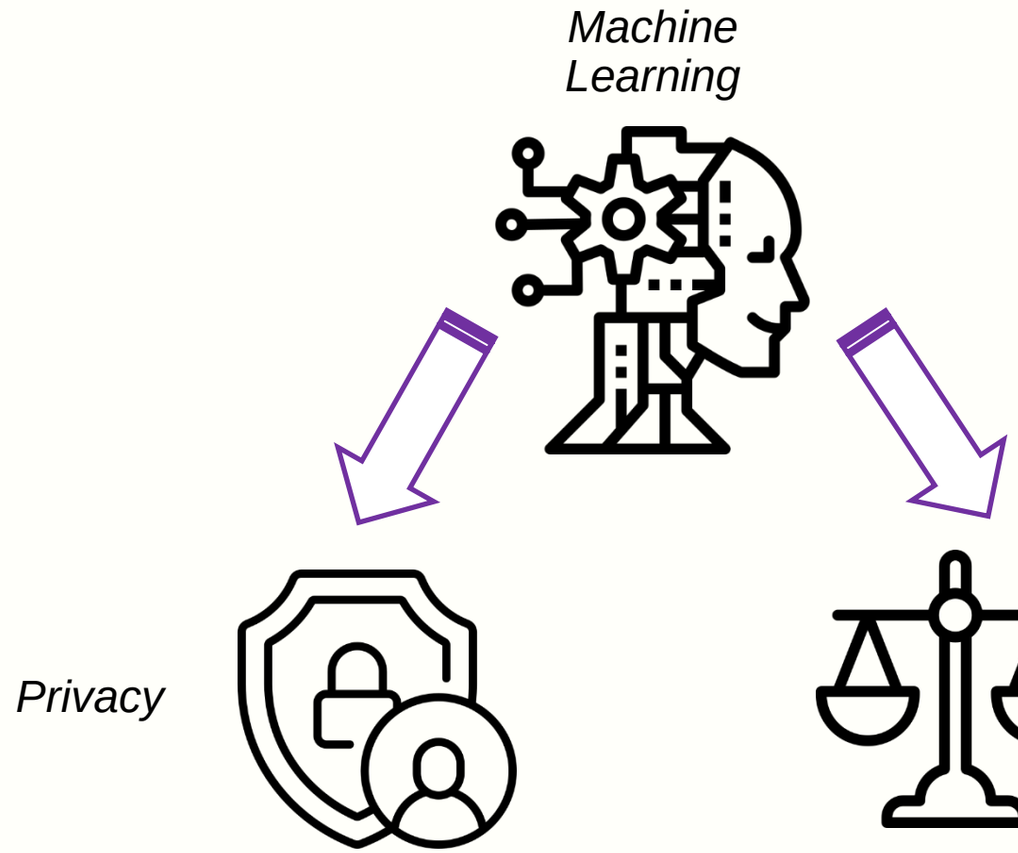
Rivoluzione digitale

t.depalma@studenti.unisa.it
https://github.com/andesrule
https://www.linkedin.com/in/thomas-de-palma-4459a1266

Fairness, Privacy, Ethics in sistemi di Machine Learning

Thomas De Palma
Università degli Studi di Salerno

Introduzione e Background



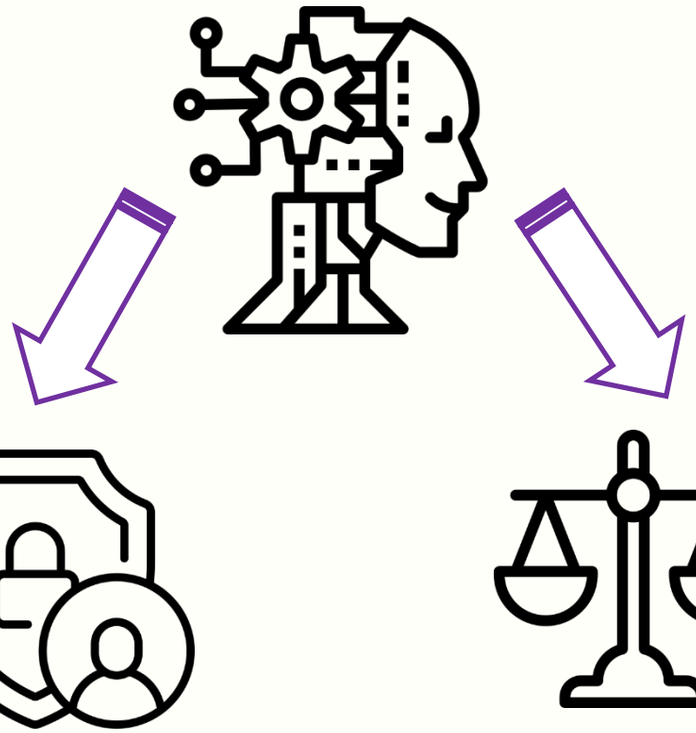
L’analisi di grandi quantità di dati tramite sistemi intelligenti ha aperto la società a nuove prospettive di sviluppo…

…Ma anche a nuove criticità da risolvere.

t.depalma@studenti.unisa.it
https://github.com/andesrule
https://www.linkedin.com/in/thomas-de-palma-4459a1266

Introduzione e Background

Machine Learning



Tutela
delle informazioni
sensibili degli utenti

Imparzialitá nelle predizioni

Privacy e Fairness sono due importanti vincoli qualitativi nella progettazione di un modulo di Machine Learning.

t.depalma@studenti.unisa.it
https://github.com/andesrule
https://www.linkedin.com/in/thomas-de-palma-4459a1266

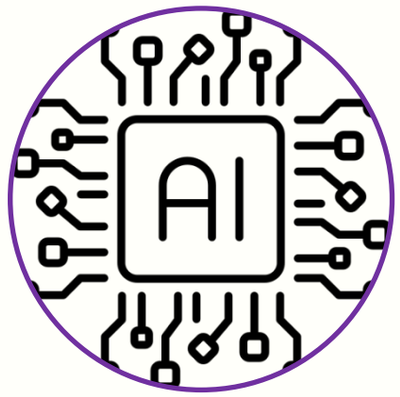
Introduzione e Background

La privacy è il diritto di ogni individuo di poter controllare e fornire le autorizzazioni relative alle proprie informazioni personali affinché possano essere raccolte, archiviate, elaborate e distribuite.

t.depalma@studenti.unisa.it
https://github.com/andesrule
https://www.linkedin.com/in/thomas-de-palma-4459a1266

Introduzione e Background

Tutela della Privacy



Intelligenza Artificiale

Sistemi di Machine Learning Privati

t.depalma@studenti.unisa.it
https://github.com/andesrule
https://www.linkedin.com/in/thomas-de-palma-4459a1266

Thomas De Palma
Università degli Studi di Salerno

Introduzione e Background

Sistemi di Machine Learning Privati

Crittografia (FHE,SMC)

Privacy Differenziale

t.depalma@studenti.unisa.it
https://github.com/andesrule
https://www.linkedin.com/in/thomas-de-palma-4459a1266

Fairness, Privacy, Ethics in sistemi di Machine Learning

Thomas De Palma
Università degli Studi di Salerno

Introduzione e Background

Fairness è un concetto che viene generalmente compreso trattando le persone in egual modo, senza pregiudizi o discriminazioni, affinché ad ognuno vengano date le stesse opportunitá.

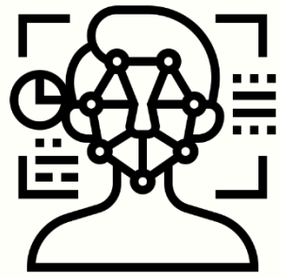
t.depalma@studenti.unisa.it
https://github.com/andesrule
https://www.linkedin.com/in/thomas-de-palma-4459a1266

Introduzione e Background

Celebri esempi di discriminazioni effettuate da sistemi di intelligenza artificiale

COMPAS è un software utilizzato negli Stati Uniti che misura la tendenza di una persona con precedenti penali a commettere di nuovo un crimine. COMPAS ha piú probabilitá di prevedere che un criminale Afro-Americano sia recidivo rispetto ad uno Caucasico

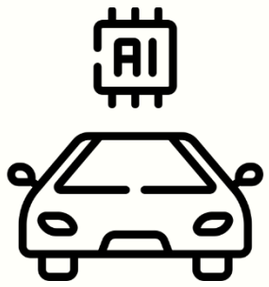
Il software di riconoscimento facciale di Amazon, nel 2018 ha effettuato pregiudizi significativi contro persone di colore in particolare verso donne con tonalitá della pelle piú scura.



t.depalma@studenti.unisa.it
https://github.com/andesrule
https://www.linkedin.com/in/thomas-de-palma-4459a1266

Introduzione e Background

Celebri esempi di discriminazioni effettuate da sistemi di intelligenza artificiale



Nel 2018 uno studio condotto dal Georgia Institute of Technology ha dimostrato come le auto con guida autonoma avevano piú probabilitá di colpire un pedone di colore rispetto ad un pedone di altra etnia.

Un algoritmo utilizzato nel settore sanitario americano per determinare quali pazienti hanno bisogno di piú cure mediche e trattamenti speciali, ha dato piú prioritá a pazienti di etnia caucasica rispetto a pazienti di altre etnie.

t.depalma@studenti.unisa.it
https://github.com/andesrule
https://www.linkedin.com/in/thomas-de-palma-4459a1266

Introduzione e Background

L’idea di Fairness puó essere declinata in significati differenti a seconda della persona a cui lo si chiede, queste variazioni di interpretazioni e definizioni possono essere dovute a ció che è piú o meno giusto a seconda
delle influenze socio-culturali e dei contesti storici dei soggetti coinvolti.

t.depalma@studenti.unisa.it
https://github.com/andesrule
https://www.linkedin.com/in/thomas-de-palma-4459a1266

Introduzione e Background

Definizioni di Fairness

Definizioni di Gruppo

- Equalized Odds
- Equal Opportunity
- Demographic Parity
- Disparate Impact

Definizioni Individuali

Individui simili devono essere trattati in modo simile

t.depalma@studenti.unisa.it
https://github.com/andesrule
https://www.linkedin.com/in/thomas-de-palma-4459a1266

Thomas De Palma
Università degli Studi di Salerno

Metodologia

È stata effettuata una revisione sistematica della letteratura. Sono state individuate le seguenti research questions:

> **RQ$_1$.** *Esistono relazioni di dipendeza tra fairness e privacy nello sviluppo di soluzioni di machine learning?*

> **RQ$_2$.** *In quali applicativi machine learning specific, le implicazioni e le dipendenze tra fairness e privacy sono particolarmente rilevanti?*

> **RQ$_3$.** *Esistono, ad oggi, strumenti automatici atti a misurare, trattare in maniera congiunta le implicazioni dirette tra privacy e fairness nello sviluppo ML?*

t.depalma@studenti.unisa.it
https://github.com/andesrule
https://www.linkedin.com/in/thomas-de-palma-4459a1266

Metodologia

Query di Ricerca

🔍 ("Privacy" ∨ "Private") ∧ ("Fairness" ∨ "Fair") ∧ ("Machine Learning" ∨ "ML")

Sottomissione query di ricerca nei database

Query di ricerca

- Google Scholar: 109.000 risultati
- IEEEXplore: 122 risultati
- ArXiv: 236 risultati
- ACM Digital Library: 9100 risultati

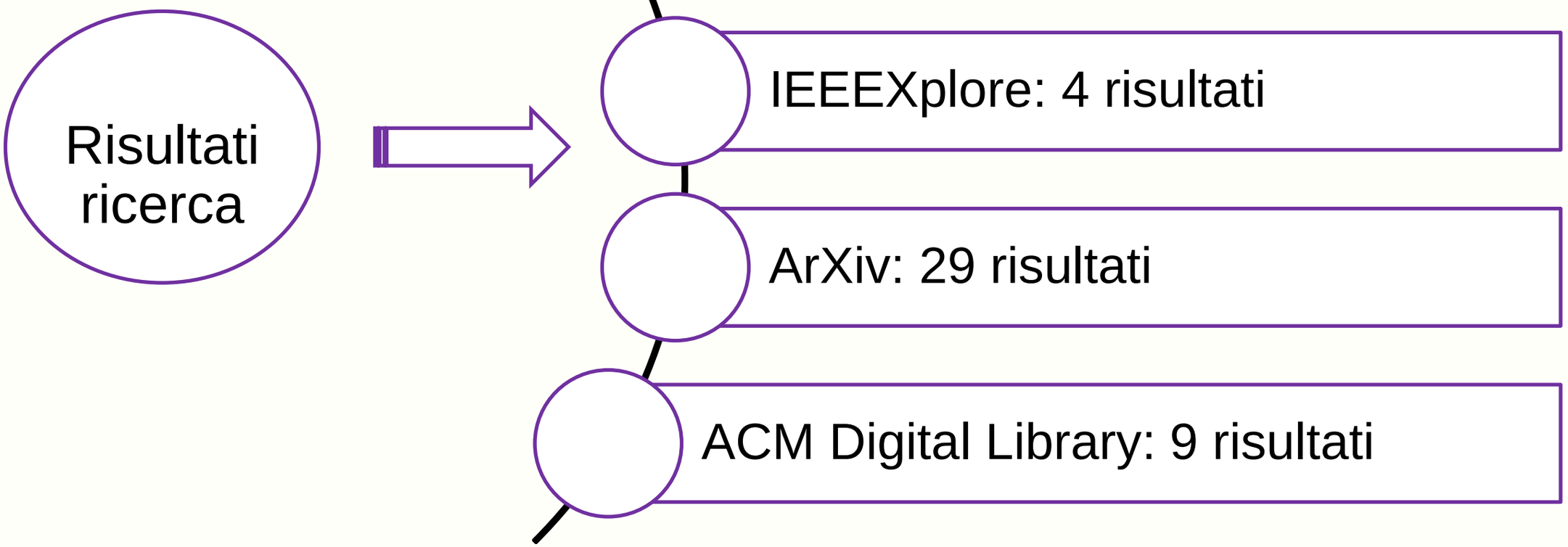
t.depalma@studenti.unisa.it
https://github.com/andesrule
https://www.linkedin.com/in/thomas-de-palma-4459a1266

Metodologia

Query di Ricerca

🔍 ("Privacy" ∨ "Private") ∧ ("Fairness" ∨ "Fair") ∧ ("Machine Learning" ∨ "ML")

Applicazione criteri di esclusione



t.depalma@studenti.unisa.it
https://github.com/andesrule
https://www.linkedin.com/in/thomas-de-palma-4459a1266

Metodologia

Applicazione criteri di inclusione

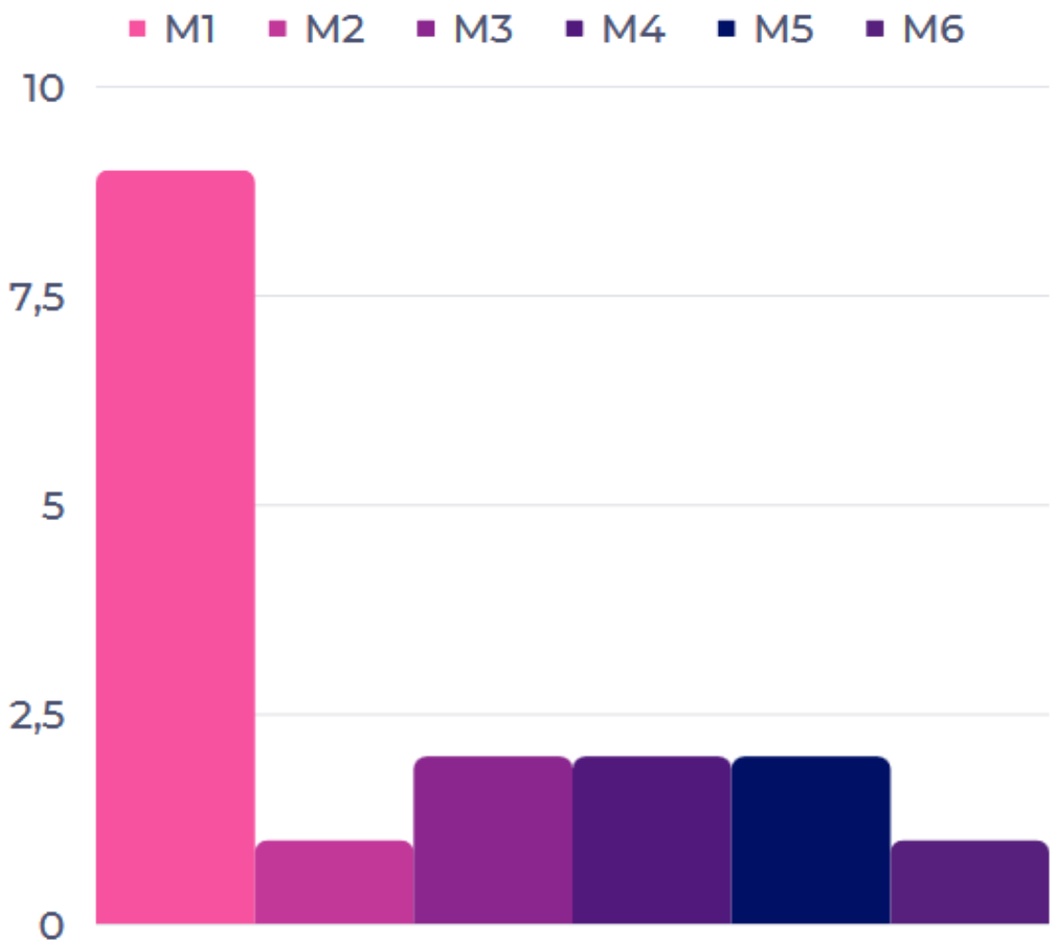
Prima selezione

15 documenti totali presi in esame

t.depalma@studenti.unisa.it
https://github.com/andesrule
https://www.linkedin.com/in/thomas-de-palma-4459a1266

Risultati

RQ$_1$. *Esistono relazioni di dipendeza tra fairness e privacy nello sviluppo di soluzioni di machine learning?*



M1: DP e Fairness Gruppo

M2: PPA e Fairness Individuale

M3: PPA e Fairness Generica

M4: DP e Fairness Generica

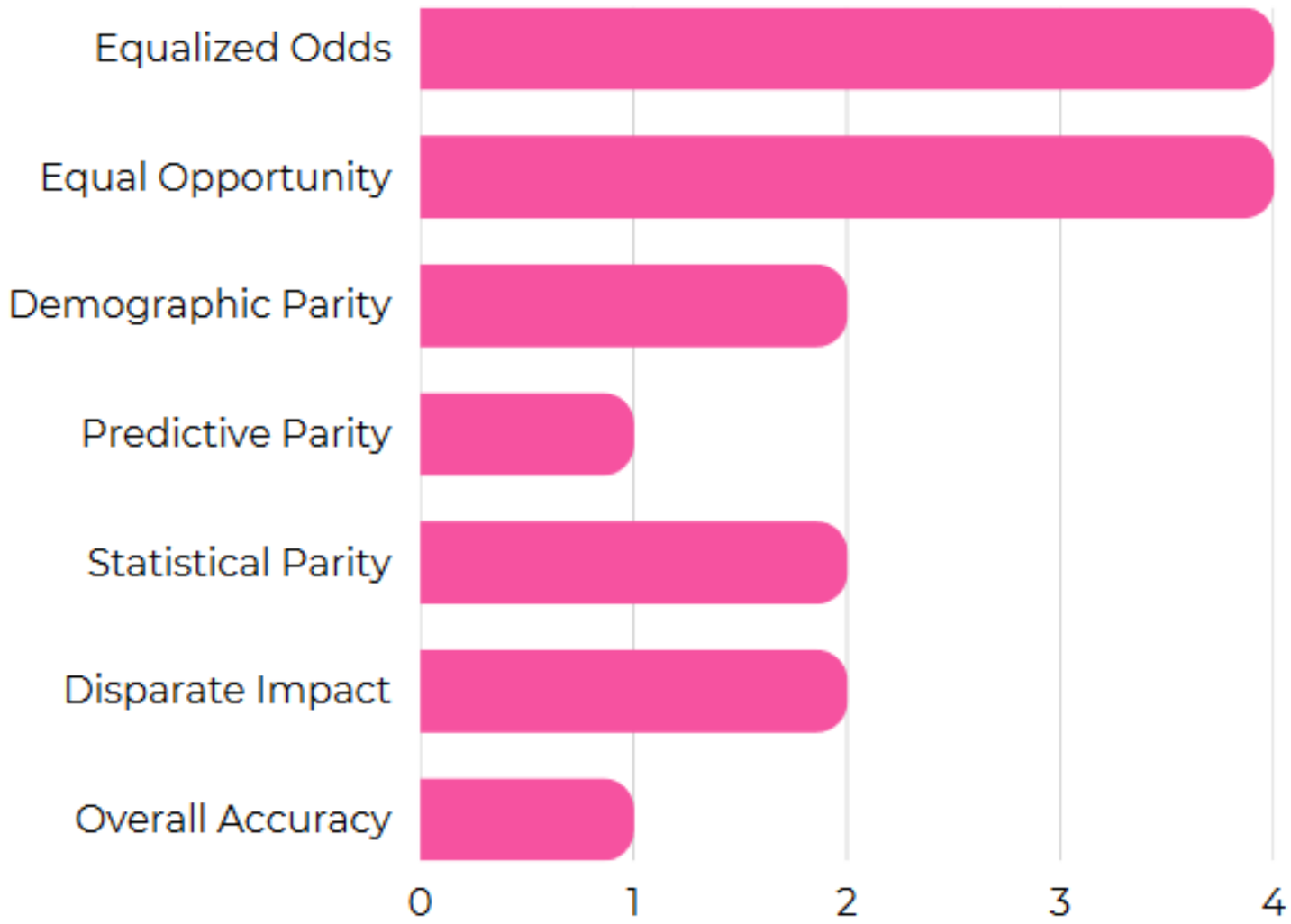
M5: Crittografia e Fairness Generica

M6: Crittografia e Fairness Gruppo

t.depalma@studenti.unisa.it
https://github.com/andesrule
https://www.linkedin.com/in/thomas-de-palma-4459a1266

Risultati

RQ$_1$. *Esistono relazioni di dipendeza tra fairness e privacy nello sviluppo di soluzioni di machine learning?*

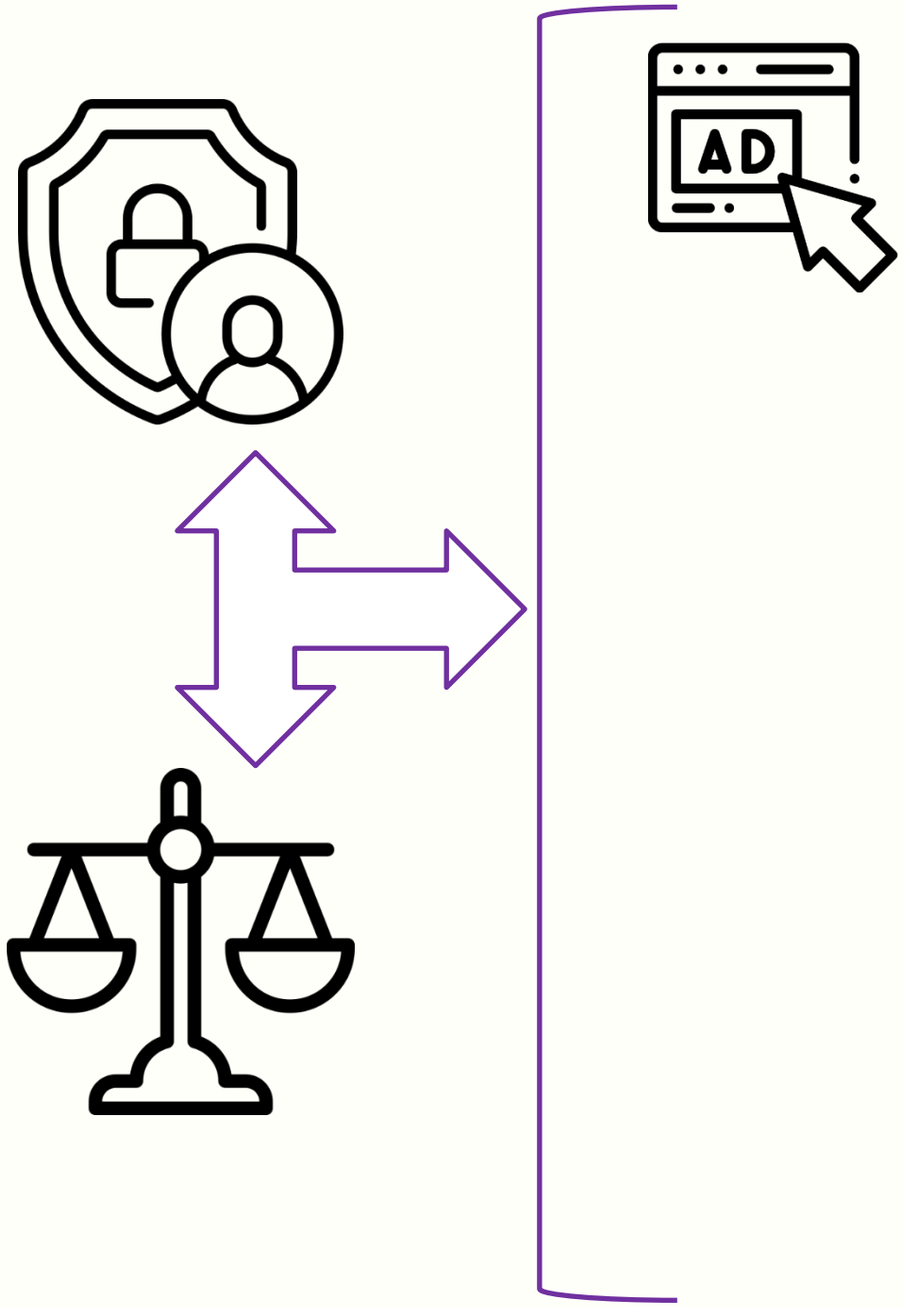


t.depalma@studenti.unisa.it
https://github.com/andesrule
https://www.linkedin.com/in/thomas-de-palma-4459a1266

Fairness, Privacy, Ethics in sistemi di Machine Learning

Thomas De Palma
Università degli Studi di Salerno

Risultati



RQ$_2$. *In quali applicativi machine learning specific, le implicazioni e le dipendenze tra fairness e privacy sono particolarmente rilevanti?*

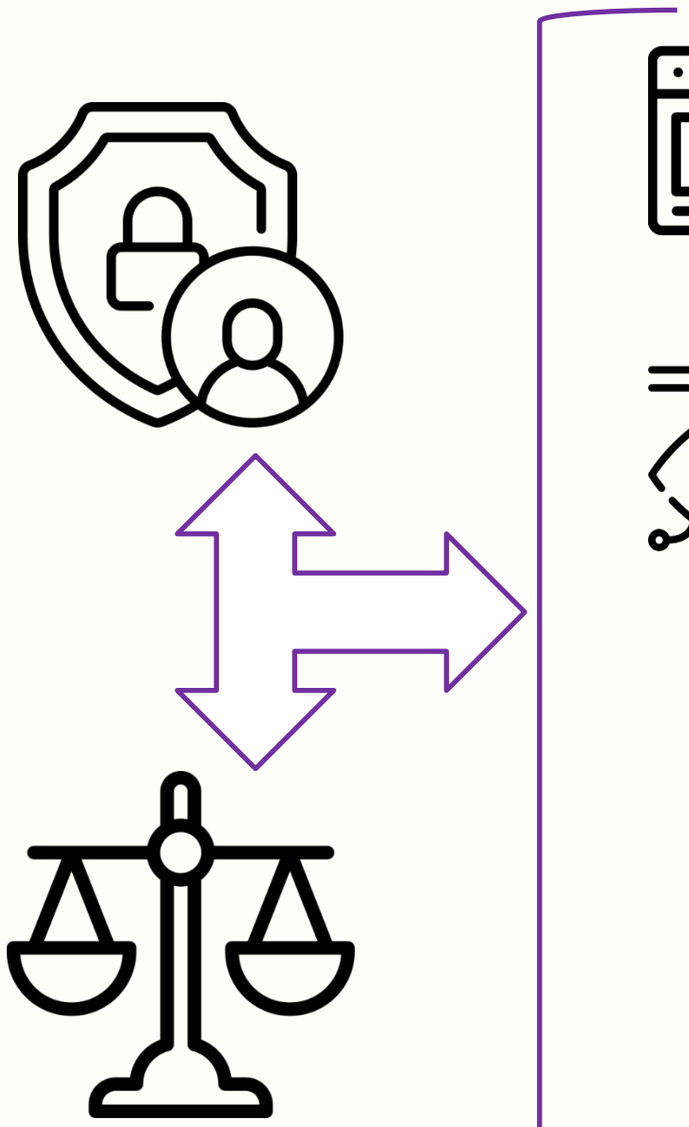
Sistemi pubblicitari personalizzati

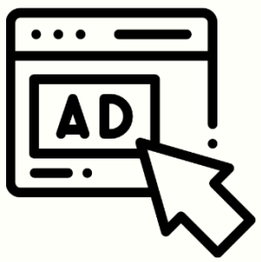
t.depalma@studenti.unisa.it
https://github.com/andesrule
https://www.linkedin.com/in/thomas-de-palma-4459a1266

Fairness, Privacy, Ethics in sistemi di Machine Learning

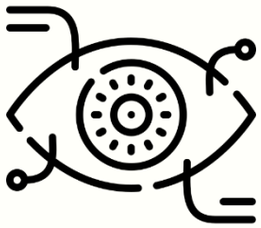
Thomas De Palma
Università degli Studi di Salerno

Risultati





Sistemi pubblicitari personalizzati



Computer Vision

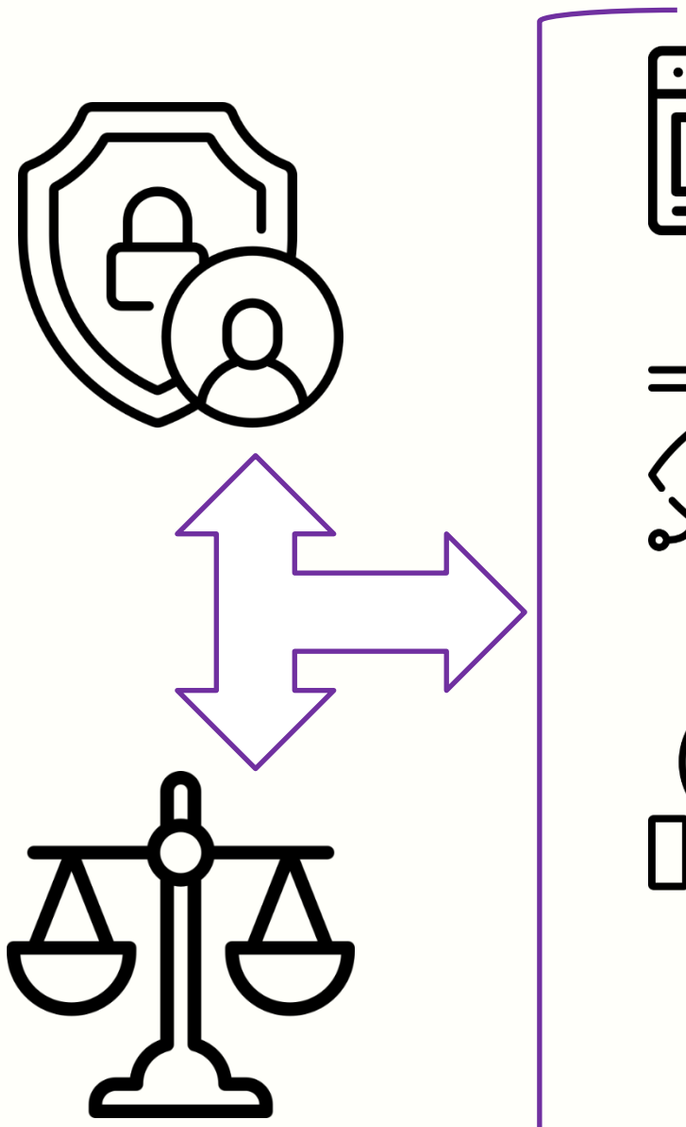
RQ$_2$. *In quali applicativi machine learning specific, le implicazioni e le dipendenze tra fairness e privacy sono particolarmente rilevanti?*

t.depalma@studenti.unisa.it
https://github.com/andesrule
https://www.linkedin.com/in/thomas-de-palma-4459a1266

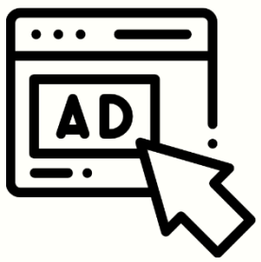
Fairness, Privacy, Ethics in sistemi di Machine Learning

Thomas De Palma
Università degli Studi di Salerno

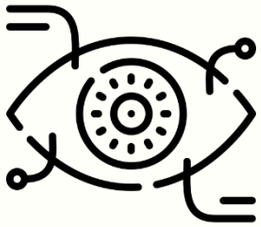
Risultati



> **RQ$_2$.** *In quali applicativi machine learning specific, le implicazioni e le dipendenze tra fairness e privacy sono particolarmente rilevanti?*



Sistemi pubblicitari personalizzati



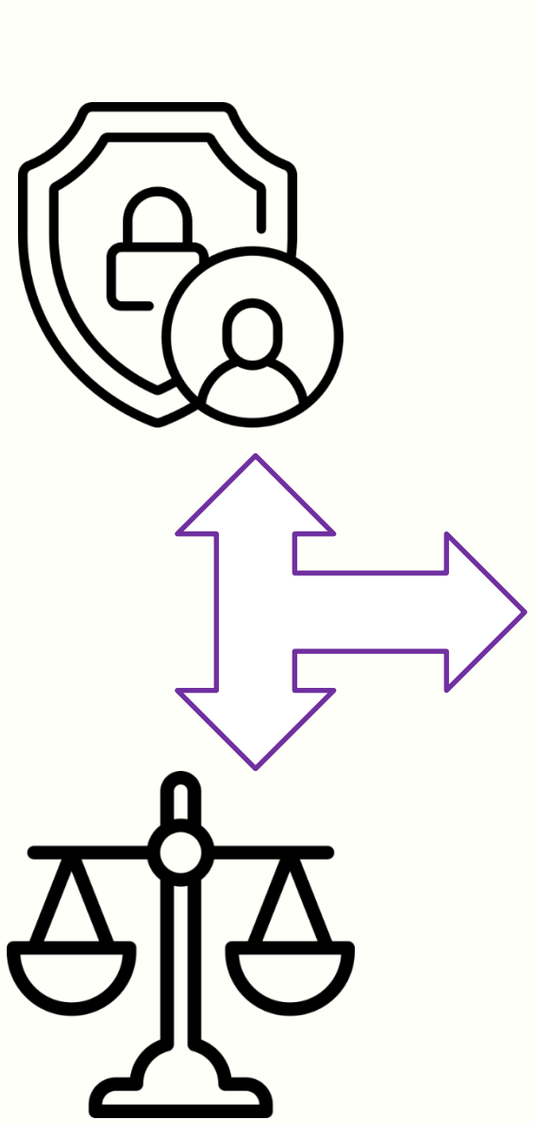
Computer Vision

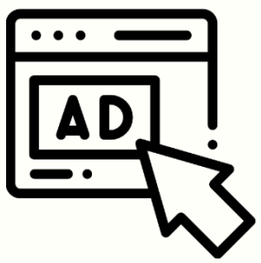
Sanitá

t.depalma@studenti.unisa.it
https://github.com/andesrule
https://www.linkedin.com/in/thomas-de-palma-4459a1266

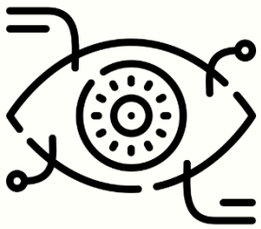
Risultati

RQ$_2$. *In quali applicativi machine learning specific, le implicazioni e le dipendenze tra fairness e privacy sono particolarmente rilevanti?*





Sistemi pubblicitari personalizzati



Computer Vision

Sanitá

Smart Cities

t.depalma@studenti.unisa.it
https://github.com/andesrule
https://www.linkedin.com/in/thomas-de-palma-4459a1266

Fairness, Privacy, Ethics in sistemi di Machine Learning

Thomas De Palma
Università degli Studi di Salerno

Risultati

ID	Modello / Algoritmo di ML	Dataset	Conclusioni
P9	PPA basati su GAN: • Mo-PAE • Traj-GAN	• MDC • GEOLIFE	La fairness individuale non viene garantita dagli algoritmi considerati. Le metriche di fairness di gruppo non subiscono violazioni.
P11	Classificatori che sfruttano tecniche di pixelizzazione e sfocatura: • K-Nearest Neighbour (KNN) • Naive Bayes (NB) • Support Vector Classfier (SVC) • Multi-Layer Perceptron (MLP)	PUBFIG	Vengono riportate discriminazioni nei confronti di attributi quali sesso e razza. L'iniquitá non dipende dall'utilizzo di classificatori che sfruttano tecniche di offuscamento delle immagini
P12	LGBM addestrato tramite privacy differenziale locale	• ADULT • ACS • LSAC	L'utilizzo del framework di privacy differenziale locale non inficia significativamente sulle performance e non genera iniquitá
P13	Algoritmi di logistic regression (PFLR e PFLR*) a cui viene applicata privacy differenziale	• ADULT • DUTCH	Negli algoritmi proposti vengono garantiti sia i requisiti di privacy che fairness preservando la precisione delle predizioni.
P15	PrivFairFL, un framework di federated learning che combina privacy differenziale e SMC	• ADS • MovieLens-1K	La soluzione proposta risulta efficace per garantire fairness di gruppo pur preservando la privacy.

RQ$_3$. *Esistono, ad oggi, strumenti automatici atti a misurare, trattare in maniera congiunta le implicazioni dirette tra privacy e fairness nello sviluppo ML?*

Per rispondere alla terza domanda di ricerca sono state individuate le tecniche adottate in ogni documento per misurare l'impatto di metodi per la tutela della privacy sulle nozioni di fairness.

È stata generata una tabella che contenesse in sintesi gli algoritmi di ML utilizzati, i dataset e le conclusioni raggiunte.

t.depalma@studenti.unisa.it
https://github.com/andesrule
https://www.linkedin.com/in/thomas-de-palma-4459a1266

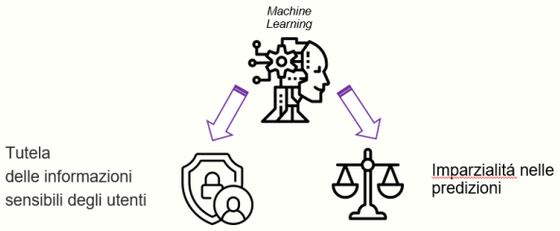
Sviluppi Futuri

Analizzare le implicazioni di metriche di fairness individuali su modelli implementati tramite paradigmi PUT (privacy-utility trade-off).

Questi ultimi infatti hanno dato risultati promettenti in quanto giá a priori garantiscono la utility del modello, mentre le definizioni individuali di fairness rappresentano ancora un ostacolo nella ricerca.

t.depalma@studenti.unisa.it
https://github.com/andesrule
https://www.linkedin.com/in/thomas-de-palma-4459a1266

Introduzione e Background



Privacy e Fairness sono due importanti vincoli qualitativi nella progettazione di un modulo di Machine Learning.

t.depalma@studenti.unisa.it
https://github.com/andesrule
https://www.linkedin.com/in/thomas-de-palma-4459a1266

Metodologia

È stata effettuata una revisione sistematica della letteratura. Sono state individuate le seguenti research questions:

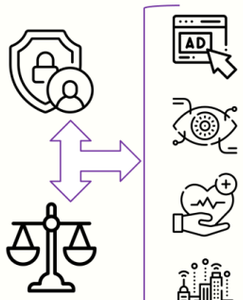
- **RQ$_1$.** *Esistono relazioni di dipendeza tra fairness e privacy nello sviluppo di soluzioni di machine learning?*
- **RQ$_2$.** *In quali applicativi machine learning specific, le implicazioni e le dipendenze tra fairness e privacy sono particolarmente rilevanti?*
- **RQ$_3$.** *Esistono, ad oggi, strumenti automatici atti a misurare, trattare in maniera congiunta le implicazioni dirette tra privacy e fairness nello sviluppo ML?*

t.depalma@studenti.unisa.it
https://github.com/andesrule
https://www.linkedin.com/in/thomas-de-palma-4459a1266

Fairness, Privacy, Ethics in sistemi di Machine Learning
Thomas De Palma
Università degli Studi di Salerno

Risultati

RQ$_2$. *In quali applicativi machine learning specific, le implicazioni e le dipendenze tra fairness e privacy sono particolarmente rilevanti?*



- Sistemi pubblicitari personalizzati
- Computer Vision
- Sanitá
- Smart Cities

t.depalma@studenti.unisa.it
https://github.com/andesrule
https://www.linkedin.com/in/thomas-de-palma-4459a1266

Fairness, Privacy, Ethics in sistemi di Machine Learning
Thomas De Palma
Università degli Studi di Salerno

Sviluppi Futuri

Analizzare le implicazioni di metriche di fairness individuali su modelli implementati tramite paradigmi PUT (privacy-utility trade-off).

Questi ultimi infatti hanno dato risultati promettenti in quanto giá a priori garantiscono la utility del modello, mentre le definizioni individuali di fairness rappresentano ancora un ostacolo nella ricerca.

t.depalma@studenti.unisa.it
https://github.com/andesrule
https://www.linkedin.com/in/thomas-de-palma-4459a1266

Fairness, Privacy, Ethics in sistemi di Machine Learning
Thomas De Palma
Università degli Studi di Salerno

treedom®
let's green the planet

Questa tesi ha contribuito a piantare un albero in Kenya

Fairness, Privacy, Ethics in sistemi di Machine Learning

Grazie!

Thomas De Palma

t.depalma@studenti.unisa.it
https://github.com/andesrule
https://www.linkedin.com/in/thomas-de-palma-4459a1266