



Corso di Laurea Magistrale in Informatica

# Machine Learning Fairness

## Analisi empirica dello stato della pratica

Prof. Fabio Palomba  
Dott.ssa. Giulia Sellitto

Carmine Ferrara  
Mat: 0522500990



✉ [c.ferrara49@studenti.unisa.it](mailto:c.ferrara49@studenti.unisa.it)

🌐 <https://cferrara98.github.io/>

🌐 <https://tinyurl.com/27zbux8z>

sesa<sup>lab</sup>  
SOFTWARE ENGINEERING  
SALERNO



# Outline

01 Introduzione e Background

02 Problematica e Obiettivi

03 Metodologia di Ricerca

04 Risultati di Ricerca

05 Implicazioni e Studi Futuri

Machine Learning Fairness  
Analisi empirica dello stato della pratica



c.ferrara49@studenti.unisa.it



<https://cferrara98.github.io/>



<https://tinyurl.com/27zbux8z>





Machine Learning Fairness  
Analisi empirica dello stato della pratica

# 01

## Introduzione e Background



✉ [c.ferrara49@studenti.unisa.it](mailto:c.ferrara49@studenti.unisa.it)

🌐 <https://cferrara98.github.io/>

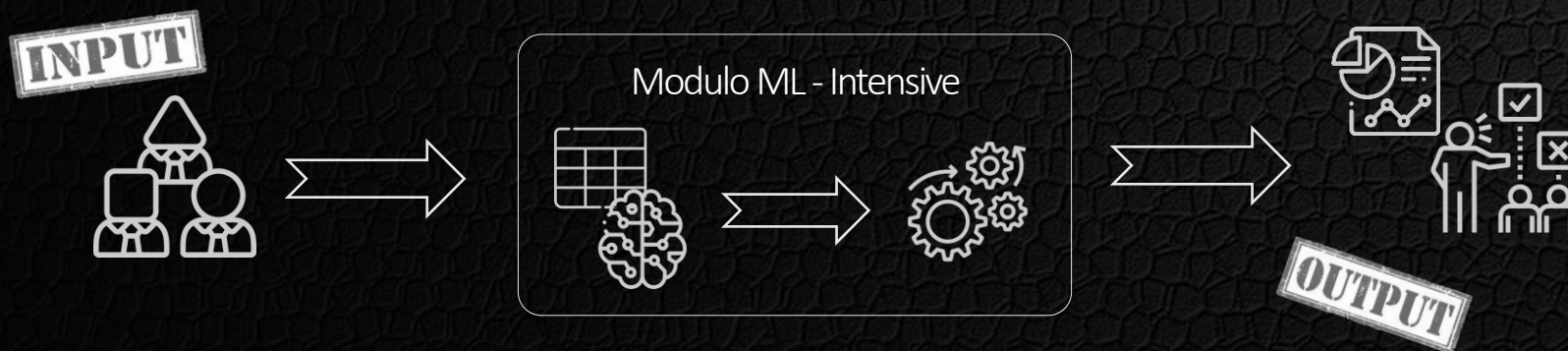
🌐 <https://tinyurl.com/27zbux8z>





# Introduzione e Background

*Nuove tipologie di vulnerabilità nello sviluppo ML-Intensive*



Data bias

Molto spesso le discriminazioni di **un** modulo ML hanno origine dall'introduzione di Bias

Learning bias



*Le soluzioni di Machine Learning sono progettate e sviluppate da umani,  
e gli umani naturalmente riproducono il loro giudizio all'interno delle Soluzioni di Machine Learning*



✉ [c.ferrara49@studenti.unisa.it](mailto:c.ferrara49@studenti.unisa.it)

🌐 <https://cferrara98.github.io/>

🌐 <https://tinyurl.com/27zbux8z>



# Introduzione e Background

## *Alcuni Esempi Famosi*



Nel 2015, un tool di Recruiting di Amazon, era solito giudicare effettuare discriminazioni sul Gender dei candidati.

Nel 2014 Compass, un tool americano AI - Intensive per l'analisi di colpevolezza, era solito giudicare facilmente, come colpevoli, le persone di colore.



Per alcune professioni quali “Dottore o Dottoressa”, la traduzione dal turco di Google Translate non tengono conto del Gender del soggetto.



✉ [c.ferrara49@studenti.unisa.it](mailto:c.ferrara49@studenti.unisa.it)

🌐 <https://cferrara98.github.io/>

🌐 <https://tinyurl.com/27zbux8z>



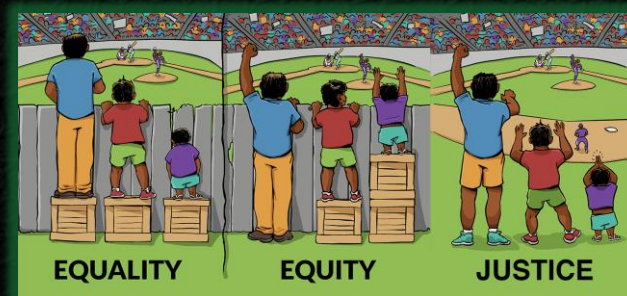
# Introduzione e Background

## Machine Learning Fairness



Una caratteristica complessa che coinvolge innumerevoli fattori e contesti sociali...

... che a seconda dei domini applicative può essere trattata in svariati modi



*Fairness: Set di requisiti, metodi e tecniche che rendono una soluzione di intelligenza artificiale eticamente corretta.*



✉ [c.ferrara49@studenti.unisa.it](mailto:c.ferrara49@studenti.unisa.it)

🌐 <https://cferrara98.github.io/>

🌐 <https://tinyurl.com/27zbux8z>



# Introduzione e Background

## Machine Learning Fairness

Un vero e proprio *Hot Topic* di ricerca

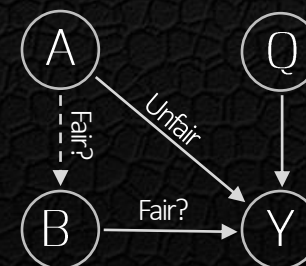
Con più di 20 differenti *definizioni e metriche*

Approcci Probabilistici



Approcci di Similarità

Approcci Causali



*Fairness: Set di requisiti, metodi e tecniche che rendono una soluzione di intelligenza artificiale eticamente corretta.*



✉ [c.ferrara49@studenti.unisa.it](mailto:c.ferrara49@studenti.unisa.it)

🌐 <https://cferrara98.github.io/>

🌐 <https://tinyurl.com/27zbux8z>



Machine Learning Fairness  
Analisi empirica dello stato della pratica

# 02

## Problema e Obiettivi

sesə<sup>lab</sup>  
SOFTWARE ENGINEERING  
SALERNO



c.ferrara49@studenti.unisa.it



<https://cferrara98.github.io/>



<https://tinyurl.com/27zbux8z>

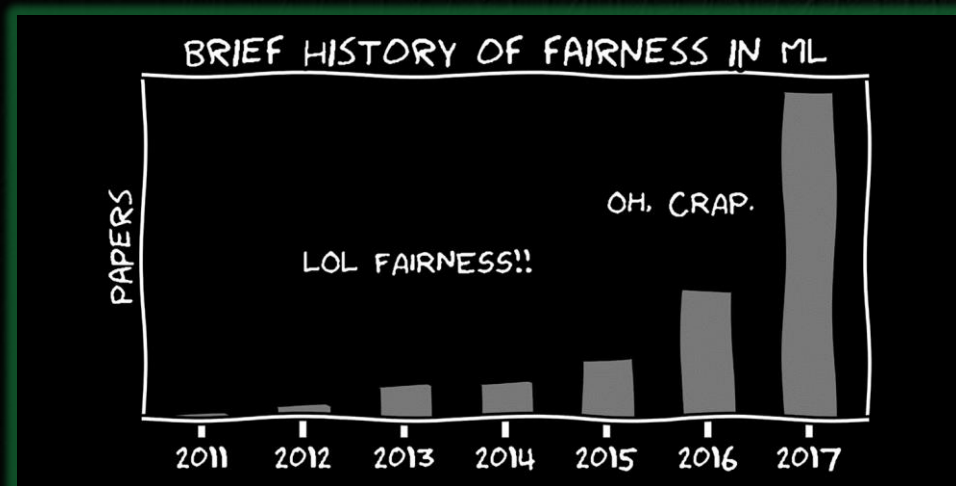




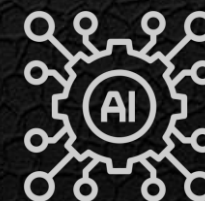
# Problema e Obiettivi

## *Osservare la pratica lavorativa nel trattamento di Fairness?*

È osservabile come negli ultimi anni Fairness sia cresciuta considerevolmente come argomento di ricerca...



Dai numerosi studi di intelligenza artificiale atti a migliorare i livelli di fairness di un sistema ML-Intensive...



... Fino a quelli ingegneristici che trattano fairness... come un vero aspetto di qualità nello sviluppo ML-Intensive



✉ c.ferrara49@studenti.unisa.it

🌐 <https://cferrara98.github.io/>

🌐 <https://tinyurl.com/27zbux8z>



# Problema e Obiettivi

*Osservare la pratica lavorativa nel trattamento di Fairness?*

Ma si riscontra facilmente come la comunità scientifica abbia poco riscontro con i professionisti

Come Fairness è approcciata realmente in un contest lavorativo AI-Intensive?



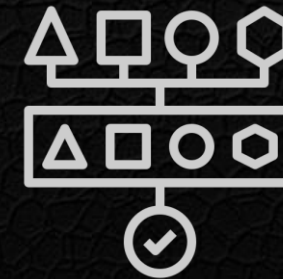
Quali sono le definizioni o gli approcci maggiormente utilizzati?

Quali sono le figure necessarie in un Team Fair-Oriented?



Quanto Fairness è rilevante  
Rispetto alter specifiche?

In che fasi di sviluppo ML viene trattata Fairness?



Quanto mature sono le compagnie Fair-Oriented?



✉ [c.ferrara49@studenti.unisa.it](mailto:c.ferrara49@studenti.unisa.it)

🌐 <https://cferrara98.github.io/>

🌐 <https://tinyurl.com/27zbux8z>



Machine Learning Fairness  
Analisi empirica dello stato della pratica

# 03

## Metodologia di ricerca

sesə<sup>lab</sup>  
SOFTWARE ENGINEERING  
SALERNO



c.ferrara49@studenti.unisa.it



<https://cferrara98.github.io/>



<https://tinyurl.com/27zbux8z>





# Metodologia di Ricerca

## *Investigazione Empirica*

Per rispondere ai quesiti formulati, si è scelto di coinvolgere direttamente i professionisti...

È stata scelta la metodologia del Survey dato che consentisse di:



Ottenere risposte omogenee e facilmente analizzabili;  
Coinvolgere figure professionali con differenti background;  
Collezionare dati in maniera rapida e mirata!



I contenuti del survey sono stati validati e modificati  
con l'ausilio di un **Test Pilota** effettuato da studenti



Il Survey è stato progettato secondo gli obiettivi formulati,  
ma la struttura non ne rispecchia l'ordine



✉ [c.ferrara49@studenti.unisa.it](mailto:c.ferrara49@studenti.unisa.it)

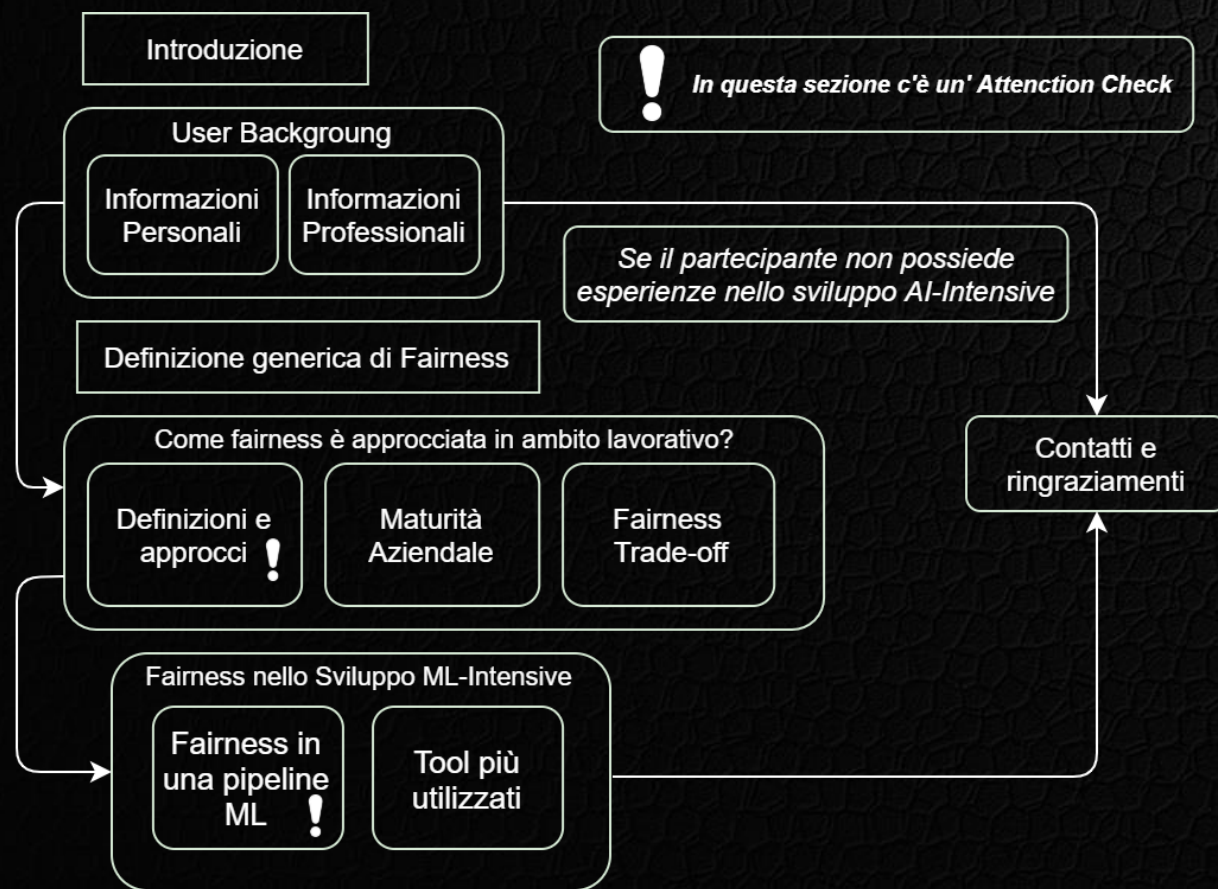
🌐 <https://cferrara98.github.io/>

🌐 <https://tinyurl.com/27zbux8z>



# Metodologia di Ricerca

## Dettagli sulla fase di studio Empirico



### Disseminazione

Il survey è stato disponibile dal 12 al 16 maggio 2022



Piattaforma nota nel mondo della ricerca per questa tipologia di studi

### Pulizia dei dati

In totale sono state raccolte **203 risposte**, di cui

19 inconsistenti e 68 compilate con poca esperienza

Al termine della fase di pulitura  
116 risposte totali sono state considerate per l'analisi dei risultati



✉ [c.ferrara49@studenti.unisa.it](mailto:c.ferrara49@studenti.unisa.it)

🌐 <https://cferrara98.github.io/>

🌐 <https://tinyurl.com/27zbux8z>

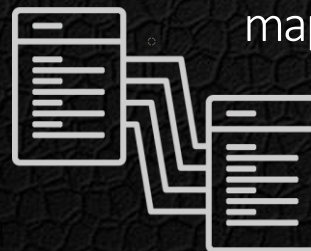


# Metodologia di Ricerca

## *Analisi dei Dati*



Statistica descrittiva e Grafici



In fase di analisi è stato ricostruita la mappatura tra quesiti del Survey e gli obiettivi di ricerca

Prima di passare alla vera e propria analisi dei dati si sono rese necessarie differenti operazioni di pre-processing tra cui trasformazioni di scala o l'utilizzo di acronimi



✉ [c.ferrara49@studenti.unisa.it](mailto:c.ferrara49@studenti.unisa.it)

🌐 <https://cferrara98.github.io/>

🌐 <https://tinyurl.com/27zbux8z>



Machine Learning Fairness  
Analisi empirica dello stato della pratica

# 04

## Risultati di Ricerca

sesə<sup>lab</sup>  
SOFTWARE ENGINEERING  
SALERNO

✉ [c.ferrara49@studenti.unisa.it](mailto:c.ferrara49@studenti.unisa.it)

🌐 <https://cferrara98.github.io/>

🌐 <https://tinyurl.com/27zbux8z>

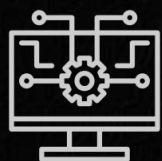




# Risultati di Ricerca

## *Composizione del Campione*

68% dei partecipanti dichiara di essere di provenienza europea



Più della metà dei partecipanti dichiara di avere anni di esperienza in Ingegneria del software, ma anche ruoli come Data Scientists o Manager sono ampiamente rappresentati



Circa il 90% dei partecipanti all'indagine possiede un titolo di studi triennale o superiori



Tutti i 116 partecipanti hanno esperienza lavorativa con Machine Learning e Fairness



✉ [c.ferrara49@studenti.unisa.it](mailto:c.ferrara49@studenti.unisa.it)

🌐 <https://cferrara98.github.io/>

🌐 <https://tinyurl.com/27zbux8z>



# Risultati di Ricerca – Obiettivo di ricerca 1

## *Definizioni ed approcci al trattamento di fairness*

Definizioni e metriche formali basate su similarità matematica o approcci pratici basati sull'analisi di dipendenze tra feature sensibili risultano essere lievemente più applicabili rispetto altri noti in letteratura

Definizioni e Metriche Formali

Approcci Pratici

***Similarità degli Individui***

Over

***Metriche di predizione o Casual Reasoning***

***Analisi di dipendenze dei Dati***

Over

***Configurazioni di Learning***

I professionisti suggeriscono altre tecniche specifiche per il trattamento di Fairness



***Metodologie empiriche o Correlazione statistica.***



✉ [c.ferrara49@studenti.unisa.it](mailto:c.ferrara49@studenti.unisa.it)

🌐 <https://cferrara98.github.io/>

🌐 <https://tinyurl.com/27zbux8z>



# Risultati di Ricerca – Obiettivo di ricerca 2

## Team Ottimale per il trattamento di Fairness



Essenziali **figure di management** che monitorino i livelli di fairness del sistema, oltre che il coinvolgimento diretto di **esperti specifici** durante lo sviluppo fair oriented

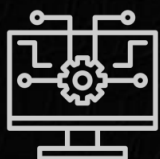


Figure **professionali trasversali quali Data Scientists e Ingegneri del Software** abbiano **responsabilità equiparabili nel trattamento delle specifiche etiche**



✉ [c.ferrara49@studenti.unisa.it](mailto:c.ferrara49@studenti.unisa.it)

🌐 <https://cferrara98.github.io/>

🌐 <https://tinyurl.com/27zbux8z>



# Risultati di Ricerca – Obiettivo di ricerca 3

## Fairness VS Altri Aspetti di Qualità

**Usability  
& Scalability**



A little Bit Over

**Fairness**



**Accuracy &  
Sicurezza**



Over

**Fairness**



L'etica di un modulo ML-Intensive è un aspetto poco maturo se paragonato ad altre specifiche non funzionali più standardizzate, tale considerazione diventa particolarmente vera con specifiche più tecniche quali Sicurezza o Accuracy!

Questi confronti hanno una variabilità elevata a seconda dello specifico dominio applicativo!

**Sicurezza** Over **Fairness** in ambito governativo



**Fairness** Over **Sicurezza** nell'Intrattenimento



✉ c.ferrara49@studenti.unisa.it

🌐 <https://cferrara98.github.io/>

🌐 <https://tinyurl.com/27zbux8z>



# Risultati di Ricerca – Obiettivo di ricerca 4

## *Fairness nello Sviluppo di Soluzioni ML-Intensive*



Fairness matura man mano che il modello evolve e matura

*Fasi di una pipeline di Machine Learning per il trattamento di Fairness*



✉ [c.ferrara49@studenti.unisa.it](mailto:c.ferrara49@studenti.unisa.it)

🌐 <https://cferrara98.github.io/>

🌐 <https://tinyurl.com/27zbux8z>



# Risultati di Ricerca – Obiettivo di ricerca 5

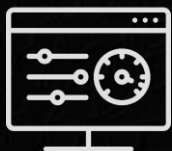
## *Fairness e Maturità Aziendale*

Il trattamento di fairness sia ancora poco maturo e standardizzato nelle pratiche aziendali ma questo aspetto è da considerarsi in costante miglioramento



Il 26% degli intervistati dichiara che la propria compagnia tratta Fairness in maniera sporadica e senza standard precisi

Un ulteriore 24% afferma che la propria compagnia adotti standard di sviluppo Fair Oriented



Della restante parte oltre il 30% dichiara di trattare le specifiche etiche con standard definiti e volti all'ottimizzazione



✉ [c.ferrara49@studenti.unisa.it](mailto:c.ferrara49@studenti.unisa.it)

🌐 <https://cferrara98.github.io/>

🌐 <https://tinyurl.com/27zbux8z>



Machine Learning Fairness  
Analisi empirica dello stato della pratica

# 05

## Implicazioni e Studi Futuri

sesə<sup>lab</sup>  
SOFTWARE ENGINEERING  
SALERNO



c.ferrara49@studenti.unisa.it



<https://cferrara98.github.io/>



<https://tinyurl.com/27zbux8z>





# Implicazioni e Studi Futuri

L'indagine empirica condotta lascia spazio a diversi spunti di riflessione...

Fairness è una specifica di rilevanza in ambito professionale, ma nuovi studi sono necessari affinché il suo trattamento venga standardizzato al pari di altre specifiche di qualità.



Esigenze di dominio e metodologie professionali alternative



Essenziali per la sistematizzazione di Definizioni e Approcci di Sviluppo standard per il Trattamento di Fairness

*Capire i requisiti Etici tramite l'ausilio di Scenari d'uso reali!*



✉ [c.ferrara49@studenti.unisa.it](mailto:c.ferrara49@studenti.unisa.it)

🌐 <https://cferrara98.github.io/>

🌐 <https://tinyurl.com/27zbux8z>



# Implicazioni e Studi Futuri

L'indagine empirica condotta lascia spazio a diversi spunti di riflessione...

La natura evolutiva dell'aspetto etico in un modulo ML è essenziale per eventuali futuri Standard.



✉ [c.ferrara49@studenti.unisa.it](mailto:c.ferrara49@studenti.unisa.it)

🌐 <https://cferrara98.github.io/>

🌐 <https://tinyurl.com/27zbux8z>



Grazie Per l'Attenzione!

```

graph TD
    Introduzione --> User_Background[User Background]
    User_Background --> Information_Personal[Information Personal]
    User_Background --> Information_Professional[Information Professional]
    Information_Professional --> Definizione_generale[Definizione generale di Fattori]
    Definizione_generale --> Come_fattori[Come fattori si applicano in ambito lavorativo?]
    Come_fattori --> Definizione_e_spiegazione[Definizione e spiegazione]
    Come_fattori --> Fattori_Trade_off[Fattori Trade off]
    Definizione_e_spiegazione --> Fattori_Relazione_M_M[Fattori con Relazione M-M. Insieme]
    Fattori_Relazione_M_M --> Fattori_Pipeline_ML[Fattori in una pipeline ML]
    Fattori_Relazione_M_M --> Test_Sofisticati[Test più sofisticati]
    Fattori_Pipeline_ML --> Confronto_Risultati[Confronto di risultati]
    Test_Sofisticati --> Confronto_Risultati
    Confronto_Risultati --> Pulizia_dati[Pulizia dei dati]
    Pulizia_dati --> Totale_Risposte[In totale sono state raccolte 208 risposte, di cui 19 inconsistenti e 162 complete con buona esperienza]
    Totale_Risposte --> Risultati_Finali[Al termine della fase di pulizia 116 risposte totali sono state considerate per l'analisi dei risultati]
    Attention_Check[In questa sezione c'è un Attention Check]
  
```

**Dettagli sulla fase di studio Empirico**

**Introduzione**

**User Background**

**Information Personal**

**Information Professional**

**Definizione generale di Fattori**

**Come fattori si applicano in ambito lavorativo?**

**Definizione e spiegazione**

**Fattori Trade off**

**Fattori con Relazione M-M. Insieme**

**Fattori in una pipeline ML**

**Test più sofisticati**

**Confronto di risultati**

**Pulizia dei dati**

**In totale sono state raccolte 208 risposte, di cui 19 inconsistenti e 162 complete con buona esperienza**

**Al termine della fase di pulizia 116 risposte totali sono state considerate per l'analisi dei risultati**

**Attenzione!** In questa sezione c'è un Attention Check

**Disposizione**

Il survey è stato disponibile dal 12 al 16 maggio 2022

**Profic**

Platforma online nel mondo della ricerca per questa tipologia di studi

# Implicazioni e Studi Futuri

L'indagine empirica condotta lascia spazio a diversi spunti di riflessione...

La natura evolutiva dell'aspetto etico in un modulo ML è essenziale per eventuali futuri Standard.

Analisi di Bad & Best Practices Specifiche

Standard di Responsabilità in un Team Fair Oriented

Protocolli Gestionali Evolutivi Fair Oriented


CLARKE&F@STUDENTIUNIVERSITÀ  
https://clarkeandf.github.io/  
https://www.linkedin.com/company/clarke-and-f/674120387/

sesa  
SISTEMI E SERVIZI  
ECONOMICI

**freedom**<sup>®</sup>  
let's green the planet

Questa tesi ha contribuito a piantare un albero in



 <https://tinyurl.com/27zbux8z>