

# UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Filosofia, Sociologia, Pedagogia e Psicologia Applicata

Corso di Laurea *Triennale* in  
SCIENZE PSICOLOGICHE SOCIALI  
E DEL LAVORO



TRASPARENZA E FIDUCIA NELL'INTELLIGENZA ARTIFICIALE:  
OPERAZIONALIZZAZIONE DEL CONCETTO DI TRASPARENZA

*Relatore:* Prof.ssa ANNA SPAGNOLLI

*Dipartimento di Psicologia Generale*

*Laureando:* GIOVANNI  
MONTELEONE

matricola n. 2021518

A.A. 2022/2023

*A mamma, papà, Francesco e Andrea*

<b>INDICE</b>	
<b>INTRODUZIONE</b> .....	1
<b>Intelligenza Artificiale (IA) ed <i>Explainable AI (XAI)</i></b> .....	<b>1</b>
<b>Domande di ricerca</b> .....	<b>2</b>
<b>METODI</b> .....	3
<b>RISULTATI</b> .....	7
<b>RQ1: Come viene operazionalizzata la trasparenza?</b> .....	<b>9</b>
A) Quali informazioni vengono rese trasparenti all'utente? .....	9
B) In che modalità venivano rese trasparenti? .....	10
C) Quando le informazioni vengono rese trasparenti all'utente?.....	10
D) Come veniva misurata la trasparenza?.....	13
E) Si misura se l'utente ha visto e compreso le informazioni?.....	14
<b>RQ2: Che rapporto ha la trasparenza con la fiducia?</b> .....	<b>15</b>
<b>CONCLUSIONI</b> .....	16
<b>BIBLIOGRAFIA</b> .....	17
<b>RINGRAZIAMENTI</b> .....	20

## INTRODUZIONE

### **Intelligenza Artificiale (IA) ed *Explainable AI (XAI)***

L'intelligenza artificiale (IA) ha gradualmente acquisito maggiore rilevanza nell'ambito decisionale, spaziando dalla personalizzazione delle raccomandazioni, all'individuazione di fonti false, all'assegnazione di un posto di lavoro, fino alla gestione delle transazioni in ambito assicurativo (Angerschmid et al., 2022; Araujo et al., 2020).

Questo progresso è stato reso possibile grazie all'affinamento degli algoritmi, all'aumento della potenza di calcolo dei computer e alla crescente disponibilità di dati analizzabili (Morana et al., 2018, citato da Vössing et al., 2022).

Tuttavia, insieme a tale avanzamento, è venuta meno la chiarezza nei processi che guidano le scelte dell'IA. Come riporta Rai, (2020), gli algoritmi di *Machine Learning* e *Deep Learning* operano in maniera non comprensibile.

In risposta alla carenza, si è sviluppato un crescente interesse nella comunità scientifica verso l'*Explainable AI (XAI)*, un ambito di ricerca volto a rendere i sistemi di intelligenza artificiale più comprensibili per gli esseri umani ([Adadi & Berreda, 2018](#)).

Nel contesto XAI, vengono esaminati vari costrutti, che si estendono dalla chiarezza alla comprensibilità, dalla correttezza alla percezione dell'utente, dall'efficacia alla trasparenza e fiducia (Angerschmid et al., 2022; Yu & Li, 2022). Questo progetto di ricerca esamina quest'ultime variabili citate, con l'obiettivo di operationalizzare la trasparenza in ambito (X)AI e comprendere il suo legame con la fiducia. Hepworth et al. (2020) spiegano come la trasparenza sia molto usata in

ambito XAI, ma molto spesso viene poco definita. Inoltre, il legame con la fiducia non è del tutto chiaro.

### **Domande di ricerca**

Le due domande di ricerca (DR) che guidano questo studio sono le seguenti:

**DR 1: Come viene operazionalizzata la trasparenza?**

**DR 2: Che rapporto c'è tra trasparenza e fiducia?**

Per rispondere alla prima domanda, sono state formulate ulteriori cinque domande:

- a) Quali informazioni vengono rese trasparenti all'utente?
- b) In che modalità venivano rese trasparenti?
- c) Quando le informazioni vengono rese trasparenti all'utente?
- d) Come veniva misurata la trasparenza?
- e) Si misura se l'utente ha visto e compreso le informazioni?

La metodologia seguita per questa ricerca bibliografica è stata una ricerca sistematica, condotta mediante l'impiego del modello PRISMA (Liberati et al., 2009). Si trovano dettagli del metodo di indagine nella sezione 2, mentre nella sezione 3 sono riportati i risultati.

Seguono le discussioni sui risultati ottenuti e le conclusioni riassuntive.

## **METODI**

La review sistematica è stata condotta utilizzando la metodologia del Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA; Liberati et al., 2009)., al fine di esaminare la relazione tra la trasparenza dell'intelligenza artificiale e la fiducia dell'utente.

Per realizzare lo studio, sono state eseguite cinque ricerche per acquisire una letteratura specifica e aggiornata. Sono stati inclusi nella ricerca tutti gli articoli pubblicati dal 2021 fino alla fine di giugno 2023, utilizzando Scopus come database bibliografico di riferimento.

Per la ricerca sono state condotte cinque diverse analisi basate su un set comune di parole chiave, ovvero “transparency”, “artificial intelligence” e “trust\*”, ricercate in modo mirato all'interno dei titoli, degli abstract e delle parole chiave degli articoli. Tuttavia, per differenziare le analisi, sono stati applicati ulteriori filtri di esclusione e di inclusione utilizzando gli operatori AND NOT e AND.

Nella prima ricerca, sono stati esclusi i termini “government”, “health\*”, “food” e “finance\*”, al fine di scartare articoli che trattano specificamente di governo, salute, cibo e finanza. Successivamente, è stata condotta una seconda ricerca in cui sono stati ulteriormente esclusi i termini “manage\*”, oltre a quelli precedentemente menzionati. Infine, una terza ricerca ha escluso i termini “education” e “student\*”, ipotizzando di ottenere articoli sempre più pertinenti rispetto la tematica di interesse.

La quarta ricerca nasce con l'obiettivo di ampliare l'ambito di ricerca e cercare di ottenere risultati rilevanti dal punto di vista dell'utente. A partire dai parametri

della terza ricerca, è stato incluso tramite AND il filtro “perception”, in modo da centrare la ricerca sull’esperienza dell’utente.

Infine, con lo stesso obiettivo della quarta analisi, nasce la quinta ricerca, che includeva tramite AND non più il filtro “perception” bensì il filtro “user”.

Sono stati coinvolti un totale di 1124 articoli, gestiti mediante l’apposito software per i riferimenti bibliografici Zotero. Dopo aver eliminato i duplicati, il numero finale di articoli ottenuti era pari a 283. Successivamente, sono state escluse ulteriori 127 pubblicazioni, che includevano conference paper e capitoli di libri.

Al seguito di tale esclusione, è stato effettuato un lavoro di doppio screening dei titoli e degli abstract degli articoli rimanenti, con il principale obiettivo di selezionare gli articoli che soddisfacevano i seguenti criteri di inclusione:

- i) riferimento ad intelligenza artificiale e fiducia dell’utente,
- ii) pubblico generalista come destinatario, cercando di limitare persone con competenze in informatica
- iii) presenza di una rilevazione empirica della fiducia.

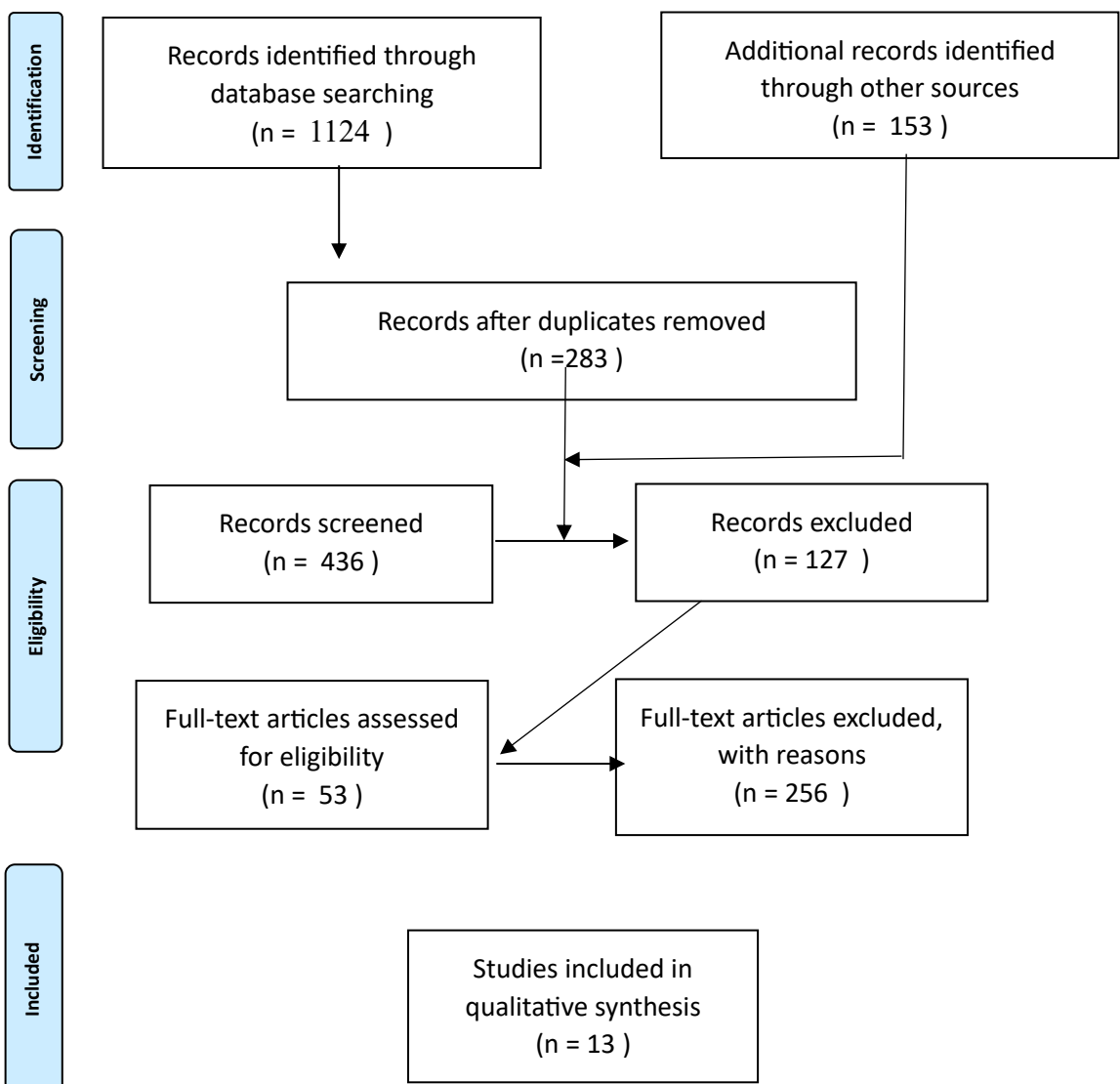
Inoltre, durante il processo di screening, sono stati selezionati gli articoli che esploravano in modo approfondito il concetto di trasparenza nell’ambito dell’intelligenza artificiale, al fine di integrare il corpus principale della ricerca e fornendo un quadro teorico più completo.

Il processo di selezione degli articoli pertinenti è stato poi completato attraverso una operazione di backward and forward con l’utilizzo di “Researchrabbit”, software di ricerca che usufruisce di intelligenza artificiale con algoritmi

proprietari per attingere alle citazioni e raccogliere articoli precedenti e successivi rispetto ai primi analizzati.

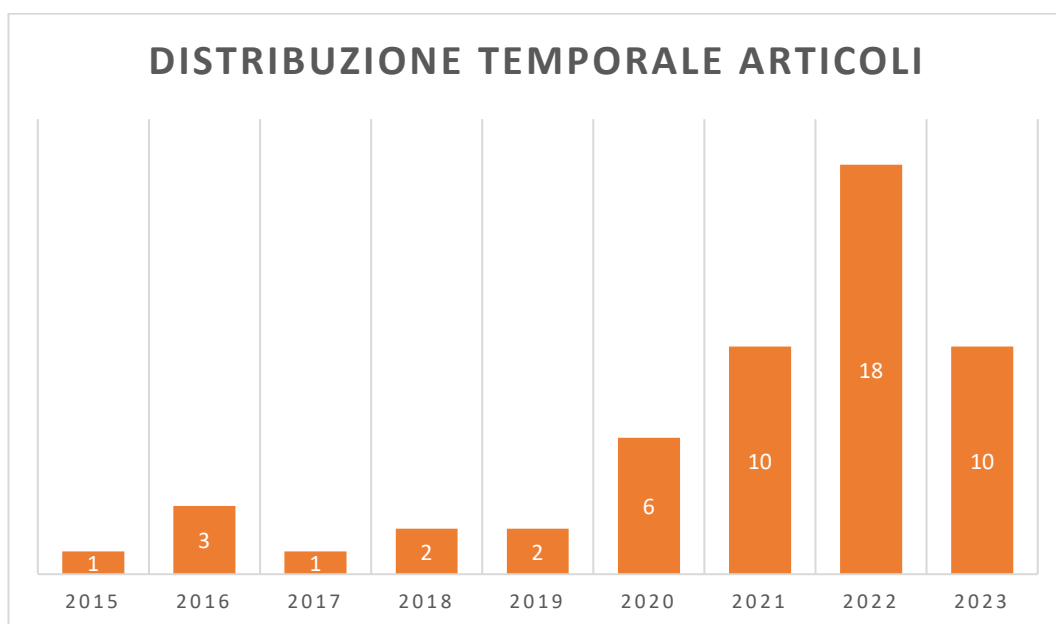
Lo screening è stato di ulteriori 150 articoli, di cui 30 soddisfano i criteri precedentemente citati.

Di seguito, un resoconto mediante un diagramma di flusso:





Date le premesse, la maggioranza degli articoli è compresa nell'intervallo temporale 2021-2023. Tuttavia, l'analisi copre un arco temporale che va dal 2015 all'anno attuale. La distribuzione temporale dei documenti è rappresentata chiaramente nel diagramma seguente:



Dei 53 articoli, sono stati analizzati nella sezione risultati 13 esperimenti.

## RISULTATI

Tutte le ricerche esaminate indagano l'utilizzo dell'intelligenza artificiale nel processo decisionale. I diversi ambiti di applicazione e le differenze tra i partecipanti nei vari studi sono riportati nella tabella 1

Tabella 1

Titolo articolo	Ambito di applicazione IA	Numero Partecipanti	Info partecipanti
<b>Artificial Intelligence Decision-Making Transparency and Employees' Trust: The Parallel Multiple Mediating Effect of Effectiveness and Discomfort</b> (Yu & Li, 2022)	Decisione aziendale	235	50% uomini, 50% donne 76% non lavora in IT
<b>Being Trustworthy is Not Enough: How Untrustworthy Artificial Intelligence (AI) Can Deceive the End-Users and Gain Their Trust</b> (Banovic et al., 2023)	Decisione in una partita di scacchi	120	51% senza competenze in IA,
<b>Building trust in automatic video interviews using various AI interfaces: Tangibility, immediacy, and transparency</b> (Suen & Hung, 2023)	Interviste automatizzate	282	68% donne, 32 uomini 69% non lavora in IT
<b>Designing for Confidence: The Impact of Visualizing Artificial Intelligence Decisions</b> (Karran et al., 2022)	Riconoscimento delle immagini	206	60% uomini, 40% donne 70% non esperti in IA
<b>Effects of Fairness and Explanation on Trust in Ethical AI</b> (Angerschmid et al., 2022)	Decisione su assicurazioni sanitarie	25	60% uomini, 40% donne Studenti
<b>Employees' Appraisals and Trust of Artificial Intelligences' Transparency and Opacity</b> (Yu, Li & Fan, 2023)	Decisione aziendale	375	49% uomini 51% donne 71% non lavora in IT

<b>Explaining Why the Computer Says No: Algorithmic Transparency Affects the Perceived Trustworthiness of Automated Decision-Making</b> (Grimmelikhuijsen, 2023)	Assegnazione visto; assegnazione sostegni economici	897	50% donne, 50% uomini
<b>In AI We Trust? Effects of Agency Locus and Transparency on Uncertainty Reduction in Human–AI Interaction</b> (Liu, 2021)	Detenzione fake news; valutazione gradevolezza sui social	491	51% donne, 49% uomini
<b>The automation of leadership functions: Would people trust decision algorithms?</b> (Hoddinghaus, Sondern & Hertel, 2021)	Decisione aziendale	333	62% donne, 38% uomini 89% non lavora in IT
<b>The Effects of Explanations in Automated Essay Scoring Systems on Student Trust and Motivation</b> (Conjin, Kahr & Snijders, 2023)	Correzione compiti scolastici	152	53% uomini, 47% donne studenti
<b>The effect of transparency and trust on intelligent system acceptance: Evidence from a user-based study</b> (Wanner et al., 2022)	Manutenzione industriale automatizzata	160	69% uomini 31% donne
<b>The perception of humanness in conversational journalism: An algorithmic information-processing perspective</b> (Donghee, 2021)	Erogazione notizie	760	51% donne, 49% uomini
<b>When AI moderates online content: effects of human collaboration and interactive transparency on user trust</b> (Molina & Sundar, 2022)	Individuazione contenuti non etici	676	52% uomini 46 % donne 2% altro

## **RQ1: Come viene operazionalizzata la trasparenza?**

### **A) Quali informazioni vengono rese trasparenti all'utente?**

Gli esperimenti presi in considerazione hanno esaminato diversi aspetti legati al processo decisionale effettuato dall'IA. (Angerschmid et al., 2022; Banovic et al., 2023; Conjin, Kahr & Snijders, 2023; Donghee, 2021; Grimmelikhuijsen, 2023; Hoddinghaus, Sondern & Hertel, 2021; Karran et al., 2022; Liu, 2021; Molina & Sundar, 2022; Suen & Hung, 2023; Wanner et al., 2022; Yu & Li, 2022; Yu, Lin & Fan, 2023).

Tuttavia, le diverse definizioni di trasparenza considerate nei vari articoli hanno portato a divergenze nella tipologia di informazioni esplicate.

Per esempio, Yu e Li (2022), Wanner et al., (2022) e Yu, Lin & Fan (2023), concordano nel definire la trasparenza nell'IA come l'abilità della IA stessa di rivelare la logica alla base della sua decisione;

Conjine, Kahr e Snijders (2023) Donghee (2021) e Grimmelikhuijsen (2023), pur parlando di trasparenza algoritmica, riportano tra loro definizioni divergenti:

Conjine, Kahr e Snijders (2023) e Grimmelikhuijsen (2023) condividono un punto di vista simile rispetto alla possibilità che esperti possano analizzare l'algoritmo e che gli utenti possano comprendere il processo decisionale; Donghee (2021) considera la trasparenza come la chiarezza delle raccomandazioni fornite dell'algoritmo, con una interpretazione più tautologica.

Nel lavoro di Angerschmid et al (2022), non viene riportata alcuna definizione di trasparenza, ma viene considerata come *explainability*.

La scelta di cosa rendere trasparente dipende, pertanto, dal quadro teorico adottato dagli autori nei vari studi. I dettagli sulle informazioni trattate nei singoli articoli, si rimanda alla Tabella 2.

**B) In che modalità venivano rese trasparenti?**

In tutti gli esperimenti presi in considerazione, sono state utilizzate diverse forme di testo al fine di garantire trasparenza dei processi decisionali dell'IA. Si differenziano i contributi di Karran et al. (2022) e Wanner et al. (2022), poiché le loro spiegazioni integrano al testo delle immagini esplicative. Il grado e il contenuto di trasparenza variavano in base alle variabili prese in esame da ciascun esperimento.

**C) Quando le informazioni vengono rese trasparenti all'utente?**

Un altro aspetto analizzato riguarda il momento della presentazione delle informazioni trasparenti: tra gli esperimenti, il 54% delle volte l'elemento trasparente veniva mostrato successivamente alla scelta compiuta dalla IA, mentre nel restante 46% dei casi l'IA illustrava il modo in cui avrebbe operato prima che svolgesse il proprio compito.

Tabella 2

Titolo articolo	Definizione di trasparenza	Info trasparente
<b>Artificial Intelligence Decision-Making Transparency and Employees' Trust: The Parallel Multiple Mediating Effect of Effectiveness and Discomfort</b> (Yu & Li, 2022)	viene reso disponibile il come e il perché viene presa una decisione	-Motivazione - Processo che ha portato alla motivazione (dati)
<b>Being Trustworthy is Not Enough: How Untrustworthy Artificial Intelligence (AI) Can Deceive the End-Users and Gain Their Trust</b> (Banovic et al., 2023)	Trasparenza come forma di spiegazione	Viene manipolato il livello di abilità dell'IA nel compiere una scelta
<b>Building trust in automatic video interviews using various AI interfaces: Tangibility, immediacy, and transparency</b> (Suen & Hung, 2023)	Grado per cui gli utenti comprendo come e perché l'AI decide qualcosa (Hoff & Bashir, 2015)	cosa analizza; come analizza
<b>Designing for Confidence: The Impact of Visualizing Artificial Intelligence Decisions</b> (Karran et al., 2022)	Fornire spiegazioni sul comportamento di sistema	Modello per cui l'IA ha compiuto la scelta (dati)
<b>Effects of Fairness and Explanation on Trust in Ethical AI</b> (Angerschmid et al., 2022)	No def	Esempi sul comportamento dell'IA; valori (dati) per cui ha preso la decisione
<b>Employees' Appraisals and Trust of Artificial Intelligences' Transparency and Opacity</b> (Yu, Li & Fan, 2023)	Le informazioni che la IA lascia rispetto al suo operato	Dati utilizzati per prendere la sua decisione

<b>Explaining Why the Computer Says No: Algorithmic Transparency Affects the Perceived Trustworthiness of Automated Decision-Making</b> (Grimmelikhuijsen, 2023)	accessibilità alla struttura algoritmica; spiegabilità sul come o perché ha preso una decisione	Informare sulla possibilità di accedere; spiegazione del come ha preso la decisione
<b>In AI We Trust? Effects of Agency Locus and Transparency on Uncertainty Reduction in Human–AI Interaction</b> (Liu, 2021)	Motivazione della propria decisione	Spiegazione del come ha preso la decisione
<b>The automation of leadership functions: Would people trust decision algorithms?</b> (Hoddinghaus, Sondern & Hertel, 2021)	Descrizione di attributi che favoriscono la comprensione tra fiduciante e fiduciario	Viene solo reso trasparente che la decisione è presa da IA, né come né perché
<b>The Effects of Explanations in Automated Essay Scoring Systems on Student Trust and Motivation</b> (Conjin, Kahr & Snijders, 2023)	Evidenza rispetto ai funzionamenti interni	Criteri di scelta dell'ai
<b>The effect of transparency and trust on intelligent system acceptance: Evidence from a user-based study</b> (Wanner et al., 2022)	abilità di spiegare la logica dietro le proprie decisioni	Valori usati per segnalare il comando
<b>The perception of humanness in conversational journalism: An algorithmic information-processing perspective</b> (Donghee, 2021)	No def	Come i dati venivano collezionati, analizzati e usati
<b>When AI moderates online content: effects of human collaboration and interactive transparency on user trust</b> (Molina & Sundar, 2022)	Logica dietro le sue decisioni; significatività per l'utente finale	Parole usate dal sistema per prendere decisione; possibilità di interagire

#### **D) Come veniva misurata la trasparenza?**

Nell'analisi della letteratura condotta da Wachter et al. (2017) ([Wachter et al., 2017](#)), è stata rilevata la carenza di studi che indagano la percezione della trasparenza dal punto di vista dell'utente. Negli esperimenti presi in considerazione, questa prospettiva non è sempre stata presa in considerazione:

Nei lavori Angerschmid et al. (2022), Banovic et al. (2023), Karran et al. (2022) e Molina e Sundar (2022) è stato indagato il legame tra la spiegabilità del sistema e la fiducia, ma senza un'approfondita esplorazione di come gli utenti elaborino le informazioni presentate.

Altri autori, come Liu (2021) e Suen e Hung (2023), hanno effettuato un controllo della manipolazione, ma non strettamente focalizzato sulla percezione di trasparenza.

D'altro canto, Yu e Li (2022) e Yu, Lin & Fan (2023) hanno adottato elementi già validati da uno studio precedenti per misurare la trasparenza dal punto di vista soggettivo dell'utente.<sup>2</sup>

Anche Wanner et al (2022) hanno fatto riferimento a lavori presenti in letteratura, citando i lavori Madsen and Gregor (2000) e di Cramer et al. (2008), mentre Grimmelikhuijsen (2023), Hoddinghaus, Sondern & Hertel (2021) e Conjin, Kahr & Snijders (2023) hanno creato elementi di misurazione specifici all'interno dei loro studi.

Infine, Donghee (20021) ha sviluppato elementi di misurazione che si concentrano sull' *explainability* (capacità di spiegazione), non direttamente sulla trasparenza. Nonostante ciò, il contenuto era molto simile agli elementi riportati nel lavoro di Yu e Li (2022).



Tutto ciò è in linea con quanto sottolineato da Hepworth et al. (2020), ovvero la presenza di confusione in letteratura tra i concetti trasparenza ed *explainability*. I due termini sono spesso utilizzati come sinonimi ed è assente una distinzione marcata tra di essi.

**E) Si misura se l'utente ha visto e compreso le informazioni?**

L'analisi della comprensione delle informazioni avviene principalmente mediante l'utilizzo di item, come nei lavori di Grimmelikhuijsen (2023), Liu (2021), Wanner et al. (2022) e Donghee (2021). Si differenziano il lavoro di Conjin, Kahr e Snijders (2023), che prova a fornire una indicazione sul tempo speso dall'utente ad analizzare le informazioni elicitate, e il lavoro di Molina e Sundar (2022), che utilizza un self report.

La scala di Zhao et al, presente nei lavori di Yu e Li (2022) e Yu, Lin & Fan (2023), sebbene valuti la percezione di trasparenza, non valuta la comprensione delle informazioni. Quest'ultima non viene valutata nemmeno nei lavori di Banovic et al. (2023), Suen e Hung (2023) e Angerschmid et al. (2022).

**RQ2: Che rapporto ha la trasparenza con la fiducia?**

È emerso un impatto positivo della trasparenza sulla fiducia in tutti i contesti esaminati, ad eccezione dello studio condotto da Conjin, Kahr e Snijders (2023).

Tuttavia, gli stessi autori suggeriscono la particolarità dell'ambito di applicazione.

La componente attiva del partecipante che interagiva con l'IA contribuiva ad aumentare il legame tra trasparenza e fiducia. (Donghee, (2021); Grimmelikhuijsen, 2023; ; Karran et al., 2022; Liu, 2021; Molina & Sundar , 2023; Suen & Hung, 2023; Yu & Li, 2022;)

Sono stati evidenziati vari fattori che aumentano in positivo il legame tra i due costrutti considerati, come la possibilità di interagire con la trasparenza (Molina & Sundar, 2023), la somiglianza del testo scritto dalla IA a quella di un testo scritto da un umano (Donghee, 2021).

Altri invece segnalano che un eccessivo carico di informazione può compromettere l'affidabilità e la fiducia provata nei confronti dell'IA (Grimmelikhuijsen, 2023; Karran et al., 2022; Suen & Hung, 2023).

Yu e Li (2022) confermano l'importanza della trasparenza percepita nel rapporto con la fiducia, mentre per i lavori di Molina e Sundar (2022) e Wanner et al. (2022) la trasparenza è una componente della affidabilità, quest'ultima correlata alla fiducia dell'utente.

L'avvertenza di Banovic et al. è sul come la trasparenza possa indurre a fiducia anche con informazioni false.

## **CONCLUSIONI**

L'analisi condotta evidenzia che il costrutto di trasparenza nell'ambito dell'Intelligenza Artificiale (IA) principalmente si riferisce alla capacità della IA di spiegare la logica delle sue decisioni (Yu & Li, 2022).

Le informazioni rese trasparenti, così come le modalità e il momento in cui vengono presentate, mostrano uniformità tra i vari esperimenti. Tuttavia, si sono riscontrate delle variazioni nella misura del costrutto: sebbene alcuni studi dimostrino l'importanza della misurare la percezione di trasparenza dal punto di vista dell'utente e la sua comprensione delle informazioni ricevute (Wanner et al., 2022), poche ricerche hanno adottato un approccio rigoroso nell'analisi dei due aspetti, compromettendo così la solidità della misurazione stessa.

L'indagine infine mostra come l'analisi del legame tra la trasparenza e la fiducia cambi in base ai riferimenti teorici utilizzati, andando ad attribuire più o meno rilevanza alla trasparenza. Al netto di eccezioni relative all'ambito di utilizzo della IA (Donghee, 2021), la correlazione dei due costrutti è positiva.

## **BIBLIOGRAFIA**

- Alessa Angerschmid, Kevin Theuermann, Andreas Holzinger, Fang Chen, & Jianlong Zhou. (2022). Effects of Fairness and Explanation on Trust in Ethical AI. *Lecture Notes in Computer Science*. [https://doi.org/10.1007/978-3-031-14463-9\\_4](https://doi.org/10.1007/978-3-031-14463-9_4)
- Araujo, T., Helberger, N., Kruikemeier, S., & de Vreese, C. H. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & SOCIETY*, 35(3), 611–623. <https://doi.org/10.1007/s00146-019-00931-w>
- Banovic, N., Yang, Z., Ramesh, A., & Liu, A. (2023). Being Trustworthy is Not Enough: How Untrustworthy Artificial Intelligence (AI) Can Deceive the End-Users and Gain Their Trust. *Proceedings of the ACM on Human-Computer Interaction*, 7(1 CSCW). Scopus. <https://doi.org/10.1145/3579460>
- Conijn, R., Kahr, P., & Snijders, C. (2023). The Effects of Explanations in Automated Essay Scoring Systems on Student Trust and Motivation. *Journal of Learning Analytics*, 10(1), 37–53. <https://doi.org/10.18608/jla.2023.7801>
- Dong-Hee Shin & Donghee Shin. (2021). The perception of humanness in conversational journalism: An algorithmic information-processing perspective: *New Media & Society*. <https://doi.org/10.1177/1461444821993801>
- Grimmelikhuijsen, S. (2023). Explaining Why the Computer Says No: Algorithmic Transparency Affects the Perceived Trustworthiness of Automated Decision-Making. *Public Administration Review*, 83(2), 241–262. Scopus. <https://doi.org/10.1111/puar.13483>
- Hepworth, A. J., Baxter, D. P., Hussein, A., Yaxley, K. J., Debie, E., & Abbass, H. A. (2021). Human-Swarm-Teaming Transparency and Trust Architecture. *IEEE/CAA Journal of Automatica Sinica*, 8(7), 1281–1295. Scopus. <https://doi.org/10.1109/JAS.2020.1003545>

Höddinghaus, M., Sondern, D., & Hertel, G. (2021). The automation of leadership functions: Would people trust decision algorithms? *Computers in Human Behavior*, *116*, 106635. <https://doi.org/10.1016/j.chb.2020.106635>

Karran, A. J., Demazure, T., Hudon, A., Senecal, S., & Léger, P.-M. (2022). Designing for Confidence: The Impact of Visualizing Artificial Intelligence Decisions. *Frontiers in Neuroscience*, *16*. Scopus. <https://doi.org/10.3389/fnins.2022.883385>

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration. *PLOS Medicine*, *6*(7), e1000100. <https://doi.org/10.1371/journal.pmed.1000100>

Liu, B. (2021). In AI We Trust? Effects of Agency Locus and Transparency on Uncertainty Reduction in Human-AI Interaction. *Journal of Computer-Mediated Communication*, *26*(6), 384–402. Scopus. <https://doi.org/10.1093/jcmc/zmab013>

Molina, M. D., & Sundar, S. S. (2022). When AI moderates online content: Effects of human collaboration and interactive transparency on user trust. *Journal of Computer-Mediated Communication*, *27*(4). Scopus. <https://doi.org/10.1093/jcmc/zmac010>

Rai, A. (2020). Explainable AI: from black box to glass box. *Journal of the Academy of Marketing Science*, *48*(1), 137–141. <https://doi.org/10.1007/s11747-019-00710-5>

Suen, H.-Y., & Hung, K.-E. (2023). Building trust in automatic video interviews using various AI interfaces: Tangibility, immediacy, and transparency. *Computers in Human Behavior*, 143. Scopus. <https://doi.org/10.1016/j.chb.2023.107713>

Vössing, M., Kühl, N., Lind, M., & Satzger, G. (2022). Designing Transparency for Effective Human-AI Collaboration. *Information Systems Frontiers*, 24(3), 877–895. Scopus. <https://doi.org/10.1007/s10796-022-10284-3>

Wanner, J., Herm, L.-V., Heinrich, K., & Janiesch, C. (2022). The effect of transparency and trust on intelligent system acceptance: Evidence from a user-based study. *Electronic Markets*, 32(4), 2079–2102. Scopus. <https://doi.org/10.1007/s12525-022-00593-5>

Yu, L., & Li, Y. (2022). Artificial Intelligence Decision-Making Transparency and Employees' Trust: The Parallel Multiple Mediating Effect of Effectiveness and Discomfort. *Behavioral Sciences*, 12(5). Scopus. <https://doi.org/10.3390/bs12050127>

Yu, L., Li, Y., & Fan, F. (2023). Employees' Appraisals and Trust of Artificial Intelligences' Transparency and Opacity. *Behavioral Sciences*, 13(4). Scopus. <https://doi.org/10.3390/bs13040344>

## **RINGRAZIAMENTI**

Ringrazio la mia relatrice Spagnolli Anna, che mi ha seguito scrupolosamente durante il percorso con grandissima professionalità e umanità.

Ringrazio mamma e papà che, con amore, hanno fatto di tutto pur di fornirmi gli strumenti necessari per coltivare i miei interessi.

Ringrazio Francesco e Andrea, i miei fratelli, i miei punti di riferimento.

Ringrazio Ciccio e Arianna, Mary e Andrea, per il loro preziosissimo contributo nel mio inserimento a Padova.

Ringrazio la mia meravigliosa e grande famiglia.

Infine, ringrazio Simona, Veronica, Tommaso, Hamza e Walid, miei coinquilini e amici che, con il loro supporto, hanno aggiunto alla mia esperienza un inestimabile valore.