

# POLITECNICO DI TORINO

Laurea Magistrale in Ingegneria Gestionale



Tesi di Laurea Magistrale

## AI e decision making in contesti burocratici complessi: approcci sperimentali per la valutazione dell'impatto organizzativo nel settore assicurativo

Relatore

Prof. Emilio PAOLUCCI

Ing. Simone TRINCHERO

Candidato

Chiara SCIARAFFA

26 Marzo 2026

*“Il mio aereo si staccava dal suolo  
ed era come un muro che si ergeva  
tra il nostro passato e il mio oggi”*



# Tabella dei Contenuti

<b>Elenco delle figure</b>	VII
<b>1 Introduzione</b>	1
1.1 Crescente diffusione dell'intelligenza artificiale nei processi organizzativi	3
1.2 Intelligenza artificiale nei contesti burocratici . . . . .	4
1.3 Trasformazione del lavoro peritale nell'integrazione uomo-AI . . . . .	6
1.4 Contesto di ricerca e inquadramento del lavoro di tesi . . . . .	9
1.5 Obiettivi e domande di ricerca . . . . .	10
<b>2 Analisi della letteratura e gap di ricerca</b>	14
2.1 Complessità e incertezza nelle attività supportate da AI . . . . .	17
2.2 Limiti dell'AI al crescere della complessità e dell'eterogeneità dei task	19
2.3 Black box ed explainable AI nei contesti decisionali complessi . . .	20
2.3.1 Discrezionalità decisionale e limiti dell'affidamento algoritmico	22
2.4 Impatto dell'AI su produttività e prestazioni decisionali . . . . .	24
2.5 Costruzione del dataset di addestramento e ruolo del ground truth .	26
2.6 Limiti della letteratura esistente e posizionamento della ricerca . . .	28
<b>3 Metodologia</b>	30
3.1 Approccio metodologico . . . . .	30
3.1.1 Disegno sperimentale . . . . .	31
3.2 Architettura dell'esperimento, partecipanti e contesto sperimentale .	35

3.2.1	Partecipanti e contesto sperimentale . . . . .	35
3.2.2	Condizioni sperimentali, controbilanciamento e procedura . .	36
3.3	Costruzione dei materiali decisionali . . . . .	39
3.3.1	Manuale operativo . . . . .	40
3.3.2	Corso introduttivo e video-pillole formative . . . . .	43
3.3.3	Scelta delle perizie . . . . .	45
3.4	Progettazione delle domande e implementazione delle condizioni sperimentali . . . . .	46
3.5	Strumento di supporto basato su intelligenza artificiale . . . . .	47
3.5.1	Implementazione sperimentale del supporto AI . . . . .	49
3.6	Variabili, metriche e strumenti di analisi . . . . .	50
3.6.1	Discrezionalità decisionale . . . . .	51
3.6.2	Accuratezza delle valutazioni rispetto al <i>ground truth</i> . . . .	52
3.6.3	Affidamento al sistema di supporto AI . . . . .	54
3.6.4	Tempo di esecuzione del compito . . . . .	55
<b>4</b>	<b>Esperimento</b>	<b>57</b>
4.1	Obiettivo dell'esperimento . . . . .	58
4.2	Procedura sperimentale . . . . .	59
4.3	Svolgimento delle sessioni sperimentali . . . . .	63
4.3.1	Esempio di struttura del test sperimentale . . . . .	64
4.4	Raccolta dei dati . . . . .	66
<b>5</b>	<b>Analisi dei risultati</b>	<b>69</b>
5.1	Descrizione del campione e dei dati raccolti . . . . .	69
5.2	Confronto tra le due condizioni . . . . .	70
5.3	Impatto del supporto AI sulla discrezionalità decisionale . . . . .	71
5.3.1	Costruzione della misura . . . . .	71
5.3.2	Risultati . . . . .	72
5.3.3	Interpretazione . . . . .	72

5.3.4	Discrezionalità nelle risposte binarie . . . . .	76
5.3.5	Risultati . . . . .	77
5.3.6	Interpretazione . . . . .	78
5.3.7	Discrezionalità decisionale a livello di partecipante . . . . .	79
5.3.8	Sintesi comparativa dei risultati . . . . .	81
5.4	Accuratezza delle decisioni . . . . .	82
5.4.1	Costruzione della variabile di accuratezza . . . . .	83
5.4.2	Risultati descrittivi . . . . .	84
5.4.3	Confronti statistici . . . . .	84
5.4.4	Interpretazione dei risultati . . . . .	86
5.4.5	Confronto dell'accuratezza tra domande di causa ed effetto . . . . .	87
5.4.6	Accuratezza della stima economica . . . . .	88
5.4.7	Coincidenza delle risposte con il ground truth nella condizione AI . . . . .	91
5.4.8	Comportamento rispetto alla correttezza del suggerimento algoritmico . . . . .	92
5.5	Affidamento degli utenti inesperti al sistema di supporto . . . . .	93
5.5.1	Risultati aggregati . . . . .	94
5.5.2	Analisi a livello individuale . . . . .	96
5.5.3	Risultati aggregati . . . . .	97
5.6	Impatto del supporto AI sui tempi di esecuzione . . . . .	100
5.6.1	Preparazione e normalizzazione dei dati . . . . .	100
5.6.2	Analisi esplorativa e identificazione dei valori anomali . . . . .	101
5.6.3	Analisi statistica . . . . .	102
5.6.4	Interpretazione . . . . .	103
5.7	Percezione soggettiva del compito sperimentale . . . . .	104
5.8	Risultati e key findings . . . . .	105
5.9	Analisi degli scostamenti . . . . .	109
5.10	Sintesi dei risultati empirici . . . . .	110

<b>6</b>	<b>Conclusioni e sviluppi futuri</b>	<b>112</b>
6.1	Risultati principali . . . . .	112
6.2	Contributo della ricerca . . . . .	113
6.3	Sviluppi futuri . . . . .	114
<b>A</b>	<b>Manuale per la valutazione di perizie fotografiche in contesto sperimentale</b>	<b>117</b>
	Sommario . . . . .	117
A.1	Introduzione e finalità del manuale . . . . .	118
A.2	Che cos'è una perizia: definizione generale . . . . .	119
A.3	La causa del danno: come riconoscerla . . . . .	120
A.4	Il danno (effetto): cosa osservare nelle immagini . . . . .	123
A.5	Lettura del suggerimento AI . . . . .	126
A.6	La stima economica: cosa rappresenta e come leggerla . . . . .	127
A.7	Criteri generali per la valutazione dell'indennizzabilità della perizia	129
<b>B</b>	<b>Tabelle risultati</b>	<b>132</b>
B.1	TEST A SENZA AI . . . . .	134
B.2	TEST A CON AI . . . . .	135
B.3	TEST B SENZA AI . . . . .	135
B.4	TEST B CON AI . . . . .	135
B.5	TEST C SENZA AI . . . . .	136
B.6	TEST C CON AI . . . . .	136
	<b>Bibliografia</b>	<b>137</b>

# Elenco delle figure

3.1	Esempi possibili cause . . . . .	43
3.2	Esempi possibili danni . . . . .	43
3.3	Legenda cromatica del suggerimento visivo fornito di supporto AI nel tool reale . . . . .	48
3.4	Esempi di suggerimenti visivi forniti dal supporto AI nel tool reale .	49
4.1	Diagramma di flusso della procedura sperimentale . . . . .	62
4.2	Esempio schermate test . . . . .	65
4.3	Esempio schermate test . . . . .	66
5.1	Discrezionalità media per condizione . . . . .	74
5.2	Differenza di discrezionalità per perizia ( $\Delta_i$ ) . . . . .	75
5.3	Discrezionalità media nelle risposte binarie per condizione . . . . .	78
5.4	Differenza di discrezionalità per singola perizia ( $\Delta_i$ ) . . . . .	79
5.5	Discrezionalità decisionale media dei partecipanti nelle condizioni con e senza supporto algoritmico. . . . .	81
5.6	Accuratezza decisionale nelle tre condizioni sperimentali . . . . .	85
5.7	Accuratezza media nelle tre condizioni sperimentali con intervalli di confidenza al 95%. . . . .	86
5.8	Accuratezza media delle risposte nelle domande relative alla causa (Q01) e all'effetto (Q02). . . . .	88
5.9	Accuratezza della stima economica nelle due condizioni sperimentali	90

5.10	Follow rate per condizione di suggerimento AI . . . . .	95
5.11	Follow rate medio con intervalli di confidenza al 95% per condizione AI . . . . .	95
5.12	Differenza individuale nel follow rate ( $\Delta follow = Follow\_correct -$ $Follow\_wrong$ ) . . . . .	97
5.13	Distribuzione dei tempi di esecuzione prima e dopo l'esclusione degli outlayer . . . . .	101
A.1	Esempi rottura tubo carico e rottura tubo scarico . . . . .	122
A.2	Esempi rottura tubo scarico . . . . .	122
A.3	Esempi di rottura guarnizioni e canotto . . . . .	123
A.4	Esempio macchia e muffa chiara . . . . .	124
A.5	Esempio muffa scura e gocciolamento . . . . .	125
A.6	Esempio perdita d'acqua e rialzamento . . . . .	125
A.7	Esempi di scrostamenti . . . . .	125
A.8	Danneggiamento finiture . . . . .	126



# Capitolo 1

## Introduzione

L'intelligenza artificiale (Artificial Intelligence, AI) può essere definita come l'insieme di tecniche e metodi computazionali finalizzati a consentire a sistemi informatici di svolgere compiti che, se eseguiti da un essere umano, richiederebbero capacità cognitive quali l'apprendimento, il ragionamento, la percezione e la risoluzione di problemi. Nel corso degli ultimi decenni, l'AI ha conosciuto un'evoluzione significativa, favorita dallo sviluppo di algoritmi di apprendimento automatico, dalla crescente disponibilità di dati digitali e dall'aumento della capacità computazionale.

Nelle applicazioni contemporanee, una parte rilevante dei sistemi di intelligenza artificiale è basata su approcci data-driven, in particolare su tecniche di machine learning, che prevedono l'addestramento del modello a partire da un dataset di esempi storici. Attraverso tale processo, l'algoritmo apprende relazioni statistiche e pattern ricorrenti nei dati, che vengono successivamente utilizzati per effettuare classificazioni, previsioni o suggerimenti su nuovi casi. In questo paradigma, le prestazioni del sistema risultano strettamente dipendenti dalla qualità, dalla completezza e dalla rappresentatività del dataset di addestramento, nonché dalle modalità con cui il ground truth viene definito.

Negli ultimi anni, l'intelligenza artificiale è divenuta oggetto di un crescente

interesse applicativo da parte di imprese, organizzazioni pubbliche e operatori professionali, che ne stanno progressivamente esplorando il potenziale come strumento operativo a supporto delle proprie attività. Sempre più frequentemente, enti e società scelgono di sviluppare o adottare soluzioni basate su AI per affrontare problematiche legate alla gestione di grandi volumi di informazioni, alla standardizzazione dei processi e all'integrazione di fonti informative eterogenee, in risposta a un contesto caratterizzato da una complessità informativa e procedurale in costante aumento.

Questa tendenza risulta particolarmente evidente nei contesti professionali e istituzionali, nei quali l'intelligenza artificiale viene progressivamente integrata nei flussi di lavoro esistenti attraverso lo sviluppo di strumenti dedicati e soluzioni applicative specifiche, spesso progettate per rispondere a esigenze operative ben definite e a vincoli procedurali stringenti. In tali contesti, l'AI non viene introdotta come tecnologia astratta, ma come componente di sistemi socio-tecnici più ampi, all'interno dei quali interagisce con operatori umani, procedure formalizzate e sistemi informativi preesistenti.

Alla luce di queste considerazioni, l'intelligenza artificiale deve essere interpretata come una tecnologia abilitante, il cui impatto dipende in larga misura dal contesto applicativo e dalle modalità di integrazione nei processi decisionali. Comprendere in che modo l'AI venga utilizzata nella pratica, quali attività venga chiamata a supportare e come si inserisca in contesti caratterizzati da elevata complessità informativa e procedurale rappresenta un passaggio fondamentale per analizzarne gli effetti sui processi decisionali reali. In molti contesti professionali, tuttavia, la difficoltà dei processi decisionali non deriva tanto da una elevata complessità procedurale, quanto piuttosto dalla complessità informativa dei casi analizzati. Le attività decisionali seguono spesso procedure relativamente standardizzate, ma richiedono l'interpretazione, l'integrazione e la classificazione di informazioni eterogenee e talvolta incomplete. In tali situazioni, il supporto di sistemi basati su intelligenza artificiale può risultare particolarmente rilevante nel facilitare l'analisi

delle informazioni disponibili e nell'orientare il processo decisionale. Su queste basi si inserisce il presente lavoro, che si concentra sull'impiego dell'intelligenza artificiale come strumento di supporto in contesti decisionali ad alta complessità informativa, con particolare attenzione alle implicazioni per il ruolo dell'operatore umano.

## 1.1 Crescente diffusione dell'intelligenza artificiale nei processi organizzativi

L'intelligenza artificiale ha assunto un ruolo sempre più centrale nei processi organizzativi e istituzionali, configurandosi come una delle principali componenti della più ampia trasformazione digitale che sta interessando imprese e amministrazioni pubbliche. Tale diffusione è stata favorita dalla convergenza di diversi fattori, tra cui l'aumento della capacità computazionale, la disponibilità di grandi volumi di dati digitali e il progresso delle tecniche di apprendimento automatico, che hanno reso possibile l'impiego dell'AI su scala operativa e non più esclusivamente sperimentale.

In questo contesto, l'intelligenza artificiale viene progressivamente integrata all'interno dei sistemi organizzativi come tecnologia in grado di automatizzare, standardizzare e velocizzare una pluralità di attività, incidendo sulle modalità di funzionamento dei processi interni. Come osservato da Newman et al. (2022) [1], l'AI non rappresenta una rottura radicale rispetto alla logica burocratica tradizionale, bensì un suo potenziamento tecnologico. Gli autori evidenziano come le tecnologie digitali avanzate, e in particolare l'intelligenza artificiale, tendano a rafforzare alcuni tratti tipici della burocrazia, quali la formalizzazione delle procedure, la codificazione delle regole e l'orientamento alla coerenza decisionale, piuttosto che eliminarli.

L'adozione crescente dell'AI si inserisce dunque in una traiettoria di continuità

con i processi di razionalizzazione amministrativa, all'interno della quale le tecnologie algoritmiche vengono utilizzate per rendere più efficiente la gestione delle informazioni, migliorare il coordinamento organizzativo e supportare l'applicazione sistematica di criteri decisionali predefiniti. In tale prospettiva, l'intelligenza artificiale viene spesso incorporata in strumenti applicativi specifici e soluzioni verticali, progettate per rispondere a esigenze operative ben definite e per integrarsi nei flussi di lavoro esistenti, piuttosto che come sistemi generalisti o autonomi.

Solo in una fase successiva, e in modo progressivo, l'impiego dell'AI si è esteso anche ad attività che coinvolgono direttamente i processi decisionali, affiancando gli operatori umani nella valutazione delle informazioni e nella formulazione delle scelte. Questo passaggio segna un'evoluzione rilevante del ruolo dell'intelligenza artificiale all'interno delle organizzazioni, poiché introduce sistemi algoritmici in ambiti tradizionalmente caratterizzati da responsabilità individuali, discrezionalità e vincoli istituzionali. Tali dinamiche risultano particolarmente evidenti nei contesti burocratici, nei quali l'introduzione dell'AI solleva interrogativi specifici in termini di controllo, legittimità e distribuzione del potere decisionale.

## 1.2 Intelligenza artificiale nei contesti burocratici

Nel presente lavoro, per contesti burocratici si intendono quegli ambiti organizzativi, prevalentemente istituzionali e professionali, nei quali i processi decisionali sono regolati da procedure formalizzate, vincoli normativi stringenti e da un uso sistematico della documentazione. Tali contesti sono caratterizzati da una struttura gerarchica, da una chiara attribuzione di responsabilità e dall'applicazione impersonale delle regole, con l'obiettivo di garantire coerenza, prevedibilità e legittimità delle decisioni. Come evidenziato nella formulazione classica di Weber (1998) [2], *la burocrazia rappresenta una forma di organizzazione razionale fondata sulla standardizzazione delle procedure e sulla riduzione dell'arbitrarietà individuale attraverso criteri formalizzati.*

Alla luce di queste caratteristiche, l'introduzione dell'intelligenza artificiale nei contesti burocratici non costituisce una semplice innovazione tecnologica, ma incide direttamente sulla struttura e sul funzionamento dei processi decisionali. Come osservato da Newman et al. (2022) [1], le tecnologie digitali avanzate, e in particolare l'AI, tendono a rafforzare alcuni tratti fondamentali della burocrazia weberiana, quali la formalizzazione delle valutazioni, la codificazione delle regole e l'orientamento alla coerenza decisionale, piuttosto che sostituirli. In questo senso, l'intelligenza artificiale può essere interpretata come un'estensione tecnologica della razionalità burocratica, inserendosi in continuità con i processi di razionalizzazione amministrativa già in atto.

Tuttavia, la letteratura più recente evidenzia come l'integrazione dell'intelligenza artificiale nei contesti burocratici introduca anche elementi di discontinuità rilevanti, in particolare con riferimento al ruolo della discrezionalità e al controllo umano sul processo decisionale. Bullock et al. (2019) [3] sottolineano come i sistemi di AI, quando impiegati per supportare o orientare decisioni amministrative, possano assumere il ruolo di veri e propri machine agents, contribuendo attivamente alla definizione dell'output decisionale. In tali circostanze, la discrezionalità non viene eliminata, ma si trasforma, spostandosi parzialmente dall'operatore umano al sistema algoritmico e, indirettamente, ai dati di addestramento e alle scelte progettuali incorporate nel modello.

Nei contesti burocratici, la discrezionalità rappresenta infatti una componente inevitabile del processo decisionale, poiché le regole formali risultano spesso incomplete o incapaci di coprire l'intera varietà delle situazioni reali. L'adozione di sistemi di intelligenza artificiale, addestrati su decisioni storiche considerate valide all'interno di uno specifico contesto istituzionale, può quindi portare alla riproduzione automatica di pratiche consolidate, senza una valutazione critica delle loro implicazioni normative o sostanziali. Come osservato da Bullock, Huang e Kim (2022) [4], ciò solleva problematiche rilevanti in termini di trasparenza, responsabilità e mantenimento di un controllo umano effettivo sulle decisioni prodotte dal

sistema.

Nel complesso, la letteratura suggerisce che l'impiego dell'intelligenza artificiale nei contesti burocratici non possa essere valutato esclusivamente in termini di efficienza o accuratezza, ma debba essere analizzato alla luce della complessità organizzativa, del ruolo della discrezionalità e delle modalità di interazione tra sistemi algoritmici e decisori umani. Tali aspetti risultano particolarmente rilevanti nei processi tecnico-amministrativi complessi, nei quali la decisione finale richiede l'interpretazione congiunta di informazioni eterogenee e l'applicazione di criteri formali a casistiche non completamente standardizzabili.

### **1.3 Trasformazione del lavoro peritale nell'integrazione uomo-AI**

All'interno dei contesti burocratici assicurativi, il lavoro del perito rappresenta una componente centrale del processo decisionale, in quanto contribuisce in modo determinante alla definizione degli elementi tecnici e informativi su cui si fonda la decisione preliminare relativa alla validità della claim e alla quantificazione del danno. L'attività peritale si colloca tradizionalmente in una fase intermedia del processo liquidativo e si sviluppa attraverso un insieme di attività tra loro interconnesse, che comprendono:

- l'acquisizione e l'analisi delle informazioni disponibili sul sinistro;
- l'interpretazione delle evidenze tecniche raccolte;
- la valutazione delle cause e degli effetti del danno;
- la valutazione economica dell'evento sulla base di criteri tecnici e prezziari di riferimento;
- la formalizzazione delle valutazioni effettuate in un documento strutturato a supporto della decisione amministrativa.

Sebbene tali attività siano svolte all'interno di un quadro procedurale formalizzato, esse lasciano ampio spazio alla discrezionalità professionale del perito, che si manifesta sia nella selezione delle informazioni ritenute rilevanti sia nelle modalità di interpretazione delle evidenze disponibili. Una parte significativa del valore aggiunto dell'attività peritale risiede infatti nella conoscenza tacita maturata attraverso l'esperienza sul campo, che si sviluppa secondo logiche di *learning by doing* e che difficilmente può essere completamente formalizzata o codificata. Questa conoscenza consente al perito esperto di riconoscere schemi ricorrenti, valutare eccezioni e interpretare correttamente situazioni che non risultano pienamente standardizzabili.

In molti casi, la complessità dell'attività peritale non deriva tanto dalla sequenza procedurale delle operazioni da svolgere, che risulta relativamente strutturata, quanto piuttosto dalla necessità di interpretare e classificare informazioni eterogenee provenienti da fonti diverse. Il perito è chiamato a integrare documentazione testuale, evidenze fotografiche e informazioni tecniche al fine di ricondurre il caso specifico a categorie interpretative rilevanti per la valutazione del sinistro. In questo senso, **la difficoltà del compito risiede principalmente nella gestione della complessità informativa dei casi analizzati** piuttosto che nella definizione delle procedure operative da seguire.

L'introduzione di sistemi di intelligenza artificiale in questa fase del processo comporta una progressiva riorganizzazione delle modalità operative del lavoro peritale. In particolare, l'AI viene impiegata per supportare alcune attività tradizionalmente svolte dall'esperto, quali l'analisi preliminare delle immagini e dei dati disponibili, il confronto con casistiche pregresse presenti nel database storico e la strutturazione delle informazioni rilevanti ai fini della valutazione. Attraverso l'addestramento su dataset peritali precedenti, il sistema è in grado di fornire un primo inquadramento tecnico del sinistro, orientando il processo valutativo nelle sue fasi iniziali.

Un ulteriore ambito di intervento dell'AI riguarda la redazione del documento

di valutazione assicurativa. Il sistema può contribuire alla produzione di un testo strutturato, nel quale vengono organizzati in modo coerente gli elementi essenziali della perizia, favorendo una maggiore uniformità nella forma e nella completezza informativa del documento. In questo contesto, l'apporto umano rimane tuttavia essenziale per l'integrazione, l'interpretazione e la validazione dei contenuti prodotti, soprattutto nei casi caratterizzati da elementi di ambiguità o complessità.

All'interno di questo quadro si colloca anche la stima economica del danno, che rappresenta una delle componenti più delicate dell'attività peritale. I sistemi di intelligenza artificiale tendono a basare tale valutazione su criteri standardizzati e prezzari uniformi, separando la componente tecnica della stima da ulteriori considerazioni di natura commerciale o negoziale. Questo approccio può contribuire a rendere il processo più coerente e tracciabile, pur inserendosi in un ambito nel quale l'esperienza del perito e la conoscenza del contesto specifico continuano a svolgere un ruolo rilevante.

È infine importante sottolineare che, anche in presenza di strumenti di supporto basati su intelligenza artificiale, la decisione finale rimane formalmente in capo al perito, che conserva la responsabilità della valutazione complessiva del caso e dell'interpretazione delle informazioni disponibili. Nel complesso, l'integrazione dell'AI nel lavoro peritale non implica una sostituzione dell'esperto umano, ma introduce una trasformazione delle modalità attraverso cui le attività vengono svolte e le decisioni preliminari vengono supportate. Questa riconfigurazione del ruolo del perito, del rapporto tra discrezionalità professionale e standardizzazione algoritmica, costituisce il presupposto per la definizione degli obiettivi di ricerca del presente lavoro.

## 1.4 Contesto di ricerca e inquadramento del lavoro di tesi

Il presente lavoro di tesi si colloca all'interno delle attività di ricerca svolte nell'ambito del progetto **DARE (Decision-making and Artificial intelligence for Risk Evaluation)**, un progetto di ricerca sviluppato presso il **Politecnico di Torino** e finalizzato allo studio e allo sviluppo di soluzioni basate su intelligenza artificiale a supporto dei processi di valutazione e gestione dei sinistri assicurativi non-motor. Il progetto nasce dall'esigenza di affrontare un ambito applicativo caratterizzato da elevata complessità decisionale e forte eterogeneità delle casistiche, elementi che rendono particolarmente critico l'impiego di strumenti di automazione tradizionale.

All'interno del progetto DARE, l'attenzione è stata rivolta in particolare ai sinistri relativi all'ambito abitativo, quali danni da acqua, rotture di elementi strutturali, danneggiamenti a vetri e serramenti ed eventi analoghi. In questi contesti, il processo di valutazione del danno richiede l'analisi congiunta di documentazione eterogenea, evidenze fotografiche e informazioni tecniche, e si fonda in larga misura su competenze specialistiche e conoscenza esperienziale maturata attraverso la pratica professionale. Tali caratteristiche rendono il processo peritale difficilmente formalizzabile e pongono sfide rilevanti allo sviluppo di sistemi di supporto automatizzati.

Nel quadro delle attività del progetto, l'autrice della presente tesi ha partecipato in qualità di ricercatrice allo sviluppo di un sistema di supporto basato su intelligenza artificiale e su un'ontologia di dominio, progettato per assistere alcune fasi del processo peritale, in particolare l'analisi delle immagini peritali, l'organizzazione delle informazioni rilevanti e il supporto alla stima del danno. Lo sviluppo del sistema si inserisce in una linea di ricerca progressiva, che ha previsto una prima fase di progettazione e implementazione dello strumento, seguita da una sperimentazione preliminare volta a esplorarne il funzionamento e le potenzialità.

In una fase precedente del progetto, è stato infatti condotto un primo esperimento con l'obiettivo di verificare la capacità del sistema di supportare il processo di valutazione peritale in condizioni controllate. Tale sperimentazione iniziale ha permesso di testare l'architettura del sistema, la coerenza delle informazioni fornite e la sua integrazione nel flusso decisionale, evidenziando al contempo la necessità di un'analisi più approfondita dell'interazione tra utente e strumento.

Il presente lavoro di tesi si inserisce in continuità con questa sperimentazione preliminare e ne rappresenta un'estensione mirata. In particolare, la tesi si propone di approfondire in modo sistematico l'impatto del sistema di supporto sul processo decisionale umano, concentrandosi su utenti inesperti e analizzando come l'interazione con lo strumento influenzi la discrezionalità decisionale, l'accuratezza delle valutazioni e l'efficienza operativa. L'obiettivo è quello di fornire evidenze empiriche sull'efficacia del supporto AI in un contesto che riproduce la complessità informativa tipica di un ambiente burocratico reale.

## 1.5 Obiettivi e domande di ricerca

Nonostante la crescente diffusione di sistemi di supporto basati su intelligenza artificiale nei processi decisionali, la letteratura esistente ha prevalentemente analizzato l'interazione tra giudizio umano e raccomandazioni algoritmiche in contesti sperimentali o su task decisionali semplificati. Numerosi studi si concentrano infatti su ambienti di laboratorio o su compiti artificiali, progettati per isolare specifiche dinamiche cognitive, ma che solo in parte riflettono la complessità dei contesti decisionali reali.

I contributi empirici che analizzano l'utilizzo di sistemi di supporto basati su AI in contesti operativi reali risultano invece più limitati. Inoltre, molti di questi studi si concentrano su compiti caratterizzati da una struttura decisionale relativamente semplice e da una limitata complessità informativa, nei quali le informazioni rilevanti sono già esplicitamente disponibili o facilmente interpretabili. **Rimane**

**pertanto ancora poco esplorato il comportamento degli utenti quando tali strumenti vengono introdotti in contesti decisionali caratterizzati da processi relativamente strutturati ma da elevata complessità informativa** dei casi analizzati, nei quali l'interpretazione delle informazioni disponibili richiede attività di classificazione, valutazione e integrazione di evidenze eterogenee.

Nel capitolo seguente verranno presentati alcuni dei contributi presenti in letteratura che analizzano l'utilizzo di sistemi di supporto algoritmico in compiti caratterizzati da maggiore complessità informativa, evidenziando come tali studi siano ancora relativamente limitati. Alla luce di queste considerazioni, **il presente lavoro di tesi si propone di analizzare l'impatto dell'introduzione di un sistema di supporto basato su intelligenza artificiale nel processo decisionale peritale, con particolare riferimento a utenti inesperti.** Lo studio si concentra sull'utilizzo di un tool di supporto sviluppato internamente presso il Politecnico di Torino nell'ambito del progetto DARE, progettato per assistere l'attività di valutazione tecnica e di stima economica del danno in contesti assicurativi caratterizzati da elevata complessità informativa.

L'obiettivo generale della ricerca è valutare in che modo l'introduzione di tale sistema di supporto influenzi il processo decisionale degli utenti inesperti, considerando sia la qualità delle decisioni assunte sia le modalità attraverso cui tali decisioni vengono formulate. In particolare, l'attenzione è rivolta alle dinamiche di discrezionalità decisionale, all'accuratezza delle valutazioni rispetto a un riferimento definito come *ground truth*. inteso come insieme di valutazioni precedentemente considerate corrette all'interno del contesto di riferimento, alle modalità di affidamento al sistema di supporto e all'efficienza del processo decisionale in termini di tempi di esecuzione.

A tal fine, l'esperimento è strutturato come un confronto sistematico tra due condizioni sperimentali: una condizione in assenza di supporto basato su intelligenza artificiale e una condizione in cui gli utenti possono avvalersi del tool di supporto. Il confronto tra le due condizioni consente di isolare l'effetto del sistema sulle diverse

dimensioni analizzate, mantenendo costanti le caratteristiche dei partecipanti e dei task decisionali.

Un primo obiettivo specifico consiste nell'analizzare l'impatto del supporto AI sulla discrezionalità decisionale degli utenti inesperti. In particolare, la ricerca intende osservare se e in che misura la presenza del tool modifichi il modo in cui gli utenti interpretano le informazioni disponibili, selezionano le evidenze rilevanti e formulano le proprie valutazioni, rispetto a una condizione in cui tali scelte sono effettuate senza il supporto del sistema.

Un ulteriore obiettivo specifico concerne la valutazione dell'accuratezza delle decisioni assunte dagli utenti inesperti. Le valutazioni fornite dai partecipanti vengono confrontate con un riferimento definito come *ground truth*, inteso come insieme di valutazioni precedentemente considerate corrette all'interno del contesto di riferimento. L'analisi mira a verificare se l'utilizzo del tool consenta agli utenti di avvicinare le proprie decisioni a tali valutazioni di riferimento e se la presenza del supporto AI contribuisca a ridurre la variabilità introdotta dalla soggettività individuale, rispetto alla condizione priva di supporto.

Un successivo obiettivo della ricerca riguarda l'analisi delle modalità di affidamento degli utenti inesperti al sistema di supporto basato su intelligenza artificiale. In particolare, si intende osservare in che misura gli utenti accettino, integrino o mettano in discussione le indicazioni fornite dal tool, e come tali modalità di interazione si riflettano sul processo decisionale complessivo.

Infine, la ricerca si propone di analizzare l'impatto del sistema di supporto sui tempi di esecuzione del compito decisionale, confrontando il tempo impiegato dagli utenti inesperti per completare le valutazioni nelle condizioni con e senza l'ausilio dell'intelligenza artificiale. Tale analisi consente di valutare se il tool influisca anche sull'efficienza del processo decisionale, oltre che sulla sua qualità.

Alla luce di questi obiettivi, il lavoro di tesi è guidato dalle seguenti domande di ricerca:

- **RQ1:** In che modo l'introduzione di un sistema di supporto basato su intelligenza artificiale influisce sulla discrezionalità decisionale degli utenti inesperti nel processo peritale?
- **RQ2:** L'utilizzo del supporto AI consente agli utenti inesperti di migliorare l'accuratezza delle proprie valutazioni rispetto al *ground truth*?
- **RQ3:** In che misura gli utenti inesperti si affidano alle indicazioni fornite dal sistema di supporto e come tale affidamento orienta le loro decisioni?
- **RQ4:** L'introduzione del sistema di supporto influisce sui tempi di esecuzione del compito decisionale?

Nel complesso, attraverso il confronto tra le due condizioni sperimentali, la ricerca intende fornire una valutazione empirica del ruolo dell'intelligenza artificiale come strumento di supporto nei processi decisionali peritali, chiarendo in che modo un sistema sviluppato ad hoc possa influenzare le decisioni assunte da utenti con limitata esperienza in un contesto burocratico complesso.

## Capitolo 2

# Analisi della letteratura e gap di ricerca

Prima di analizzare l'introduzione di sistemi di intelligenza artificiale all'interno di uno specifico contesto applicativo, risulta fondamentale distinguere le diverse tipologie di attività e comprendere in quali casi l'intervento dell'AI possa effettivamente contribuire a migliorare lo svolgimento del compito e in quali, invece, non rappresenti uno strumento di supporto adeguato. A tal fine, la letteratura propone diverse classificazioni delle attività in funzione del grado di automazione consentito e del ruolo attribuito all'intelligenza artificiale nel processo decisionale.

Una tassonomia ricorrente consente di individuare almeno tre macro-categorie di task, differenziate sulla base delle competenze richieste, del livello di variabilità delle casistiche e della complessità interpretativa del compito.

Una prima categoria comprende le attività che richiedono un elevato grado di creatività, intelligenza emotiva e capacità di comprensione del contesto, nonché la valutazione di elementi qualitativi difficilmente formalizzabili. In tali task, il processo decisionale dipende in larga misura dall'esperienza, dal giudizio umano e dalla capacità di integrare informazioni implicite, contestuali o non strutturate. Per queste ragioni, la letteratura concorda nel ritenere che tali attività non siano

adeguatamente automatizzabili e debbano rimanere prevalentemente in capo all'essere umano, con un ruolo limitato o marginale dell'intelligenza artificiale. In questi casi, uno dei principali rischi legati all'introduzione dell'AI è rappresentato dalla ripetitività delle soluzioni, in quanto compiti di questo tipo richiedono un significativo sforzo innovativo, elemento non assimilabile a un tool di intelligenza artificiale che opera replicando pattern appresi in fase di addestramento.

Una seconda categoria riguarda le attività che richiedono la visione, l'analisi e la classificazione di grandi quantità di dati, spesso eterogenei e complessi, ma caratterizzati da schemi riconoscibili e da regole operative relativamente stabili. In tali contesti, l'intelligenza artificiale può fornire un supporto significativo all'operatore umano, ad esempio attraverso la pre-elaborazione delle informazioni, il riconoscimento di pattern ricorrenti o la riduzione del carico cognitivo. La letteratura evidenzia come, in questi casi, la soluzione più efficace non consista in una completa automazione del processo decisionale, bensì in una sinergia uomo-AI, nella quale il sistema automatizzato affianca l'operatore umano, lasciando a quest'ultimo la responsabilità della decisione finale.

Una terza categoria comprende infine le attività ripetitive e routinarie, caratterizzate da grandi volumi di dati, bassa variabilità delle casistiche e da procedure altamente standardizzate. In presenza di tali condizioni, i task risultano particolarmente adatti all'esecuzione automatica tramite sistemi di intelligenza artificiale, che possono operare con elevati livelli di efficienza, coerenza e scalabilità. In questi casi, l'intervento umano tende a limitarsi alle fasi di supervisione, controllo e gestione delle eccezioni, mentre l'AI assume un ruolo centrale nell'esecuzione operativa del compito.

Questa distinzione risulta fondamentale per comprendere il ruolo appropriato dell'intelligenza artificiale nei processi decisionali complessi e per evitare un'eccessiva delega automatica in contesti nei quali il giudizio umano rimane imprescindibile. In particolare, numerosi studi sottolineano come i benefici dell'AI emergano soprattutto quando il sistema viene impiegato come strumento di supporto all'interno di attività

collocate tra la seconda e la terza categoria, mentre l'automazione completa di task ad alta complessità interpretativa comporta rischi significativi in termini di affidabilità e qualità decisionale.

Questa distinzione risulta fondamentale per comprendere il ruolo appropriato dell'intelligenza artificiale nei processi decisionali complessi e per evitare un'eccessiva delega automatica in contesti nei quali il giudizio umano rimane imprescindibile. In particolare, numerosi studi sottolineano come i benefici dell'AI emergano soprattutto quando il sistema viene impiegato come strumento di supporto all'interno di attività collocate tra la seconda e la terza categoria, in cui l'analisi di grandi quantità di informazioni e il riconoscimento di pattern ricorrenti possono contribuire a ridurre il carico cognitivo dell'operatore umano e a facilitare l'elaborazione delle informazioni rilevanti.

**Tuttavia, risultano ancora limitate le evidenze empiriche che esaminano in modo sistematico l'impatto dell'AI sul processo decisionale umano in attività caratterizzate da elevata complessità informativa e da margini di discrezionalità interpretativa, soprattutto quando tali decisioni vengono assunte a livello individuale.** Gran parte degli studi esistenti si concentra infatti su contesti sperimentali o su task semplificati, mentre risultano meno numerose le analisi condotte su compiti decisionali che richiedono l'integrazione di informazioni eterogenee in contesti applicativi realistici.

Alla luce di queste considerazioni, il presente lavoro si inserisce in questo filone di ricerca analizzando l'interazione tra utente e sistema di supporto basato su intelligenza artificiale in un contesto decisionale caratterizzato da elevata complessità informativa. In particolare, **l'esperimento esamina il comportamento decisionale individuale degli utenti inesperti chiamati a svolgere valutazioni peritali**, al fine di comprendere in che modo l'introduzione di un sistema di supporto possa influenzare le modalità con cui le informazioni vengono interpretate e utilizzate nel processo decisionale.

## **2.1 Complessità e incertezza nelle attività supportate da AI**

Negli ultimi anni, la letteratura scientifica ha dedicato crescente attenzione all'adozione di sistemi di intelligenza artificiale a supporto dei processi decisionali umani, in particolare in contesti caratterizzati da bassa incertezza strutturale del compito e da un'elevata standardizzazione delle procedure operative. In tali ambiti, l'AI viene impiegata per supportare decisioni ripetute, documentate e soggette a vincoli formali, con l'obiettivo di migliorare l'efficienza, la coerenza e la qualità complessiva del processo decisionale. Le attività analizzate presentano livelli contenuti di variabilità e una struttura decisionale relativamente stabile, caratteristiche che rendono tali contesti particolarmente adatti a una formalizzazione algoritmica del processo decisionale.

In questi contesti, la letteratura mostra in modo consistente come l'utilizzo di strumenti di intelligenza artificiale sia in grado di produrre significativi incrementi di produttività, in particolare a beneficio degli utenti con minore esperienza, contribuendo a ridurre le differenze di performance rispetto agli operatori più esperti. Evidenze empiriche in tal senso sono fornite dallo studio di Noy et al. (2023) [5], che analizza l'impatto dell'utilizzo di ChatGPT in attività di scrittura professionale di livello medio. I risultati mostrano una riduzione significativa dei tempi di completamento dei compiti per il gruppo con accesso al sistema AI e, al contempo, una diminuzione delle disuguaglianze di performance tra i partecipanti, come evidenziato dalla riduzione della correlazione tra i risultati dei compiti successivi rispetto al gruppo di controllo.

Risultati analoghi emergono anche in contesti operativi reali. Kanazawa et al. (2025) [6], ad esempio, analizzano l'impatto di un sistema di supporto basato su AI (AI Navi) sulla produttività dei tassisti di Yokohama, mostrando come l'utilizzo dell'algoritmo consenta una riduzione media del tempo di ricerca dei clienti pari a circa il 5%. In questo caso, l'AI opera su un task fortemente standardizzato,

caratterizzato da grandi volumi di dati e da regole decisionali relativamente semplici, confermando l'efficacia del supporto automatizzato in attività a bassa complessità decisionale.

Ulteriori evidenze provengono dall'ambito giuridico. Choi e Schwarcz (2024) [7] hanno valutato l'impatto dell'assistenza di sistemi di intelligenza artificiale generativa sul ragionamento giuridico, confrontando le prestazioni di studenti in prove d'esame con e senza accesso a GPT-4. I risultati mostrano un miglioramento significativo nelle prestazioni relative a domande semplici e standardizzate, mentre l'effetto risulta nullo o non significativo per quesiti caratterizzati da maggiore complessità interpretativa.

In modo coerente, Dell'Acqua et al. (2023) [8] distinguono tra compiti che rientrano nelle capacità operative dell'AI, definiti come compiti "dentro la frontiera", e compiti che si collocano al di fuori di tali capacità. I risultati mostrano che, per i compiti "dentro la frontiera", l'utilizzo dell'AI produce incrementi significativi sia in termini di produttività sia di qualità dell'output. Al contrario, per i compiti caratterizzati da maggiore complessità interpretativa, l'utilizzo del supporto AI può condurre a performance inferiori rispetto alla condizione di controllo, evidenziando il rischio di un affidamento inappropriato allo strumento algoritmico in contesti decisionali complessi.

Nel complesso, tali evidenze suggeriscono che l'impiego di sistemi di intelligenza artificiale risulti particolarmente efficace in attività standardizzate, ripetitive e a bassa incertezza, nelle quali l'AI può contribuire ad aumentare l'efficienza operativa e la qualità dell'output, soprattutto per gli utenti meno esperti. Tuttavia, come emerge chiaramente dagli stessi studi, tali benefici tendono a ridursi o a scomparire all'aumentare della complessità interpretativa del compito, rendendo meno prevedibili gli effetti del supporto algoritmico e ponendo le basi per una riflessione più ampia sul ruolo dell'AI nei contesti decisionali complessi.

## **2.2 Limiti dell'AI al crescere della complessità e dell'eterogeneità dei task**

Accanto alle evidenze positive relative all'impiego dell'intelligenza artificiale in compiti standardizzati e a bassa incertezza, la letteratura ha progressivamente messo in luce come l'efficacia dei sistemi di supporto basati su AI tenda a ridursi al crescere della complessità decisionale e dell'eterogeneità delle casistiche. In tali contesti, pur in presenza di procedure formalizzate e di vincoli istituzionali, il processo decisionale non può essere ricondotto a schemi ripetitivi, ma richiede un'elevata capacità di interpretazione, contestualizzazione e integrazione di informazioni eterogenee e spesso non strutturate.

A differenza dei task caratterizzati da bassa variabilità, nei processi decisionali complessi le singole casistiche presentano differenze sostanziali che incidono in modo determinante sull'esito della valutazione. La complessità non è quindi riconducibile unicamente all'incertezza dell'esito, bensì alla necessità di considerare molteplici fattori qualitativi, relazioni tra elementi diversi e specificità del contesto, che difficilmente possono essere formalizzati in regole generali. In tali condizioni, i sistemi di intelligenza artificiale incontrano limiti strutturali legati alla loro capacità di generalizzazione, poiché i modelli tendono a basarsi su pattern ricorrenti appresi dai dati storici, risultando meno efficaci quando le nuove casistiche si discostano in modo significativo da quelle osservate in fase di addestramento.

La letteratura evidenzia come, in presenza di elevata eterogeneità dei task, l'utilizzo di sistemi di supporto basati su AI produca risultati meno stabili e meno prevedibili rispetto ai contesti a bassa complessità decisionale. In questi casi, l'output dell'algoritmo può risultare formalmente coerente con i dati disponibili, ma non adeguato a cogliere le peculiarità del singolo caso, rendendo necessario un intervento umano sostanziale per la corretta interpretazione e validazione della decisione proposta. Tuttavia, tali osservazioni derivano prevalentemente da analisi concettuali o da studi condotti in contesti sperimentali semplificati, mentre risultano

ancora limitate le evidenze empiriche che analizzano sistematicamente questi effetti in contesti decisionali realistici.

Tali limiti assumono particolare rilevanza nell'interazione tra essere umano e sistema di intelligenza artificiale. Nei contesti decisionali complessi, infatti, la difficoltà dell'AI nel gestire l'eterogeneità delle casistiche incide direttamente sia sull'affidabilità oggettiva del supporto fornito sia sulla fiducia soggettiva che l'utente ripone nel sistema. Proprio in risposta a queste criticità, una parte significativa della letteratura si è concentrata sull'analisi delle caratteristiche del sistema di intelligenza artificiale che possono mitigare tali limiti, in particolare attraverso il confronto tra approcci di tipo black box e soluzioni di explainable artificial intelligence, volte a rendere più trasparente e comprensibile il processo decisionale dell'algoritmo.

Parallelamente, numerosi studi hanno indagato gli effetti dell'introduzione dell'AI sui comportamenti decisionali degli utenti, analizzando come il supporto automatizzato influenzi la produttività, la discrezionalità e la qualità delle scelte, soprattutto in funzione del livello di esperienza dell'operatore umano. In questo senso, l'aumento della complessità decisionale non rappresenta soltanto un limite tecnico per i sistemi di intelligenza artificiale, ma diventa un elemento chiave per comprendere le dinamiche di affidamento e di uso critico del supporto algoritmico nei processi decisionali reali.

## **2.3 Black box ed explainable AI nei contesti decisionali complessi**

Nei contesti decisionali ad alta complessità, una parte rilevante della letteratura si è concentrata sulla valutazione dell'accuratezza dei sistemi di intelligenza artificiale e sul grado di affidamento che i professionisti tendono a riporre nei suggerimenti forniti dal sistema. In questo ambito risulta particolarmente diffusa la comparazione tra modelli di tipo black box, che forniscono output decisionali privi di spiegazioni sul

funzionamento interno dell'algoritmo, e approcci di explainable artificial intelligence (XAI), progettati per rendere esplicite le motivazioni alla base delle raccomandazioni prodotte.

Numerosi studi, prevalentemente in ambito medico, evidenziano come la presenza di spiegazioni influenzi in modo significativo la fiducia dell'utente, la percezione di controllo e l'accettazione del sistema, pur non traducendosi necessariamente in un miglioramento dell'accuratezza decisionale o in una modifica sostanziale dell'esito finale delle decisioni. In tali contesti, la spiegabilità non viene quindi interpretata come una garanzia di correttezza dell'output, bensì come uno strumento volto a supportare l'interazione tra professionista e sistema automatizzato.

In questa direzione, Senoner et al. (2024) [9] analizzano l'effetto dell'utilizzo di strumenti di explainable AI, sotto forma di heatmap visive, sulla capacità di esperti di valutare criticamente le previsioni dell'algoritmo in contesti reali di collaborazione uomo-AI. I risultati mostrano come la disponibilità di spiegazioni visive consenta agli utenti di confrontare le previsioni dell'AI con la propria conoscenza del dominio, favorendo una maggiore adesione alle previsioni corrette e una più efficace individuazione di quelle errate, rispetto a sistemi di tipo black box.

Risultati complementari emergono dallo studio etnografico di Lebovitz et al. (2022) [10], che esplora le modalità di interazione tra professionisti medici e sistemi di intelligenza artificiale opachi in diversi reparti ospedalieri. Gli autori evidenziano come, in assenza di spiegazioni sul funzionamento dell'algoritmo, l'AI possa generare incertezza soprattutto quando i risultati divergono dal giudizio clinico iniziale. In tali situazioni, alcuni professionisti sviluppano pratiche di "interrogazione dell'AI", integrando in modo critico l'output del sistema con la propria esperienza (engaged augmentation), mentre in altri casi l'AI viene utilizzata in modo superficiale o ignorata (unengaged augmentation). Lo studio sottolinea come il valore del supporto basato su AI dipenda non solo dalle prestazioni tecniche del sistema, ma anche dalle pratiche organizzative e dal livello di coinvolgimento degli utenti.

Nel complesso, questi contributi mostrano come la spiegabilità dell'AI assuma

un ruolo centrale nei contesti decisionali complessi, non tanto in termini di miglioramento diretto dell'accuratezza, quanto per le sue implicazioni sull'affidabilità percepita del sistema e sulle modalità di interazione tra professionista e tecnologia. Tuttavia, la letteratura risulta ancora limitata nell'analizzare in modo sistematico come tali caratteristiche influenzi

### **2.3.1 Discrezionalità decisionale e limiti dell'affidamento algoritmico**

Va innanzitutto considerato che l'intelligenza artificiale, allo stato attuale dello sviluppo tecnologico, non può essere considerata uno strumento infallibile. I sistemi basati su apprendimento automatico tendono a generalizzare a partire dai pattern osservati nei dati di addestramento e possono incontrare difficoltà nel distinguere correttamente tra casi apparentemente simili ma caratterizzati da differenze sostanziali rilevanti ai fini decisionali. Tale limite risulta particolarmente critico in contesti decisionali complessi, caratterizzati da un'elevata eterogeneità delle casistiche e da una forte dipendenza dal contesto specifico, nei quali il processo decisionale richiede l'esercizio di un significativo grado di discrezionalità umana.

In particolare, Kolkman et al. (2024) [11] sottolineano come gli algoritmi di previsione decisionale siano tipicamente addestrati su un ground truth che non coincide necessariamente con la "decisione corretta" in senso normativo o sostanziale, ma piuttosto con decisioni storicamente adottate e considerate valide all'interno di uno specifico contesto istituzionale. Di conseguenza, un algoritmo può produrre un output formalmente coerente con il dataset di riferimento pur replicando decisioni peritali o giudiziarie errate. Come affermato dagli autori, "*legal prediction algorithms do not and cannot accurately perform the complex legal reasoning that is required in the decision-making process*", evidenziando i limiti intrinseci dell'automazione nei processi decisionali ad alta complessità interpretativa.

Per tale ragione, nello studio di Kolkman et al. (2024) [11] l'algoritmo non viene impiegato per predire direttamente le decisioni delle autorità giudiziarie,

ma per svolgere compiti più semplici e ripetitivi, quali la ricerca di casi simili all'interno di archivi strutturati o la redazione preliminare di documenti. Questa scelta metodologica mira a ridurre il rischio di un'eccessiva delega decisionale al sistema, preservando il ruolo centrale dell'operatore umano nell'esercizio della discrezionalità decisionale.

Considerazioni analoghe emergono anche nello studio di Mishra et al. (2024) [12], che analizza i limiti di generalizzazione dei modelli di intelligenza artificiale applicati al riconoscimento automatico delle crepe strutturali. Gli autori evidenziano come il dataset di addestramento presenti una gamma ristretta di caratteristiche delle crepe, quali larghezza, pattern e colore di sfondo, limitando la capacità del modello di adattarsi a scenari reali caratterizzati da maggiore variabilità. In tali condizioni, la presenza di rumore nei dati, ad esempio variazioni di illuminazione, texture del materiale o sfondi complessi, può influire negativamente sulla precisione delle previsioni dell'AI, riducendone l'affidabilità operativa.

Nel complesso, questi contributi mostrano come l'affidamento all'intelligenza artificiale nei contesti decisionali complessi debba essere valutato con particolare cautela. Sebbene tecniche di validazione umana e approcci di apprendimento debolmente supervisionato possano contribuire a garantire livelli accettabili di affidabilità entro ambiti circoscritti, i limiti del ground truth e della capacità di generalizzazione dei modelli rendono problematica un'eccessiva delega decisionale al sistema. In questo quadro, l'efficacia dei sistemi di supporto basati su intelligenza artificiale dipende dalla capacità di integrare in modo equilibrato il suggerimento algoritmico con il giudizio umano, mantenendo la responsabilità finale in capo all'operatore. Tali considerazioni evidenziano la necessità di analizzare gli effetti dell'interazione uomo-AI non solo in termini di prestazioni tecniche, ma anche rispetto alla discrezionalità, alla produttività e alla qualità complessiva delle decisioni assunte, temi affrontati nel paragrafo successivo.

## **2.4 Impatto dell'AI su produttività e prestazioni decisionali**

Un ulteriore filone ampiamente esplorato nella letteratura riguarda l'analisi dell'impatto degli strumenti di supporto basati su intelligenza artificiale sulla produttività e sulle prestazioni decisionali, attraverso il confronto tra condizioni sperimentali con e senza l'ausilio del sistema. Le evidenze empiriche mostrano in modo consistente come l'effetto dell'AI vari sensibilmente in funzione del livello di esperienza dell'operatore umano.

In particolare, numerosi studi indicano che gli utenti meno esperti tendono a beneficiare maggiormente del supporto fornito dall'AI, registrando incrementi significativi sia in termini di accuratezza sia di riduzione dei tempi di esecuzione del compito. In questo senso, Caplin et al. (2024) mostrano come l'assistenza dell'AI migliori le prestazioni soprattutto per individui con basse abilità di base, evidenziando inoltre che, a parità di competenze, i benefici risultano maggiori per gli utenti meglio calibrati, ossia consapevoli dei propri limiti. Gli autori sottolineano come la riduzione della miscalibrazione individuale possa contribuire in modo significativo alla diminuzione delle disuguaglianze di performance, in misura persino superiore rispetto all'effetto diretto dell'introduzione dell'AI.

Risultati analoghi emergono anche in ambiti operativi e professionali. Campero et al. (2022) [13] mostrano come l'utilizzo di GPT-3 consenta a soggetti privi di competenze di programmazione di svolgere compiti di sviluppo software con prestazioni comparabili a quelle di programmatori umani, a costi inferiori. Allo stesso modo, Dell'Acqua et al. (2023) [8], in uno studio condotto in collaborazione con il Boston Consulting Group, evidenziano come l'introduzione di GPT-4 migliori significativamente la produttività e la qualità del lavoro nei compiti che rientrano nelle capacità del modello, con effetti particolarmente marcati per i consulenti con prestazioni inferiori alla media. Benefici analoghi per gli utenti meno esperti sono riportati anche da Peng et al. (2023) [14] nell'uso di strumenti di supporto

alla programmazione e da Brynjolfsson et al. (2025) [15], che documentano un incremento medio della produttività pari al 14%, con un impatto maggiore sui lavoratori meno qualificati e un effetto limitato sugli esperti.

Nel complesso, tali risultati suggeriscono che l'intelligenza artificiale possa svolgere un ruolo rilevante nel ridurre le disuguaglianze di performance e nell'accelerare la curva di apprendimento degli utenti meno esperti, favorendo la diffusione di conoscenza precedentemente tacita. Tuttavia, tali evidenze si concentrano prevalentemente su indicatori di output e di efficienza, lasciando in secondo piano l'analisi del processo decisionale sottostante e delle modalità attraverso cui tali risultati vengono conseguiti.

Al contrario, per gli utenti più esperti l'effetto del supporto AI risulta più eterogeneo. In diversi casi, l'introduzione dell'AI non produce miglioramenti significativi e può generare frustrazione, soprattutto quando il sistema fornisce suggerimenti in contrasto con il giudizio umano o quando il processo decisionale dell' algoritmo risulta opaco. In questo senso, Jussupow et al. (2021) [16] evidenziano come, in presenza di disconferma, gli esperti tendano a mantenere una maggiore fiducia nel proprio giudizio rispetto agli utenti inesperti, mostrando una minore propensione ad accettare il suggerimento automatizzato.

Parallelamente, la letteratura segnala come gli utenti meno esperti siano maggiormente esposti a fenomeni di automation bias e overreliance, accettando il suggerimento del sistema anche quando esso risulta errato. In tali circostanze, l'interazione con l'AI può condurre a una riduzione della discrezionalità umana, con l'utente che tende a ricercare giustificazioni a posteriori per confermare la decisione proposta dal sistema, piuttosto che valutare criticamente l'informazione disponibile. Questi risultati evidenziano la necessità di analizzare in modo più approfondito come l'introduzione dell'AI influisca non solo sulle prestazioni finali, ma anche sulle dinamiche decisionali che conducono all'esito osservato.

## 2.5 Costruzione del dataset di addestramento e ruolo del ground truth

Un aspetto centrale nello sviluppo di sistemi di supporto basati su intelligenza artificiale riguarda il processo attraverso cui i dati utilizzati per l'addestramento del modello vengono selezionati, strutturati e annotati. Nei contesti professionali, infatti, molte attività decisionali si fondano sulla classificazione di casi concreti all'interno di categorie interpretative che guidano l'applicazione delle regole operative. Nella letteratura organizzativa, tale meccanismo è descritto da Mintzberg (1979) [17] come *pigeonholing*, ossia il processo attraverso cui il professionista analizza una situazione specifica, la riconduce a una categoria diagnostica e applica il programma decisionale associato.

Nel caso dei sistemi di intelligenza artificiale, questo processo di classificazione deve essere esplicitamente formalizzato attraverso la definizione di ontologie di dominio e la costruzione di dataset annotati. Attraverso tali strumenti, i domain expert e i data scientist definiscono le categorie interpretative che consentono al sistema di riconoscere e classificare i diversi tipi di evento osservabili nella realtà operativa. Il dataset di addestramento e il *ground truth* rappresentano pertanto il risultato di un processo di strutturazione concettuale dei dati che rende esplicite le categorie utilizzate nel processo decisionale professionale.

La costruzione del dataset di addestramento rappresenta quindi una fase particolarmente critica nello sviluppo di sistemi basati su AI. Il *ground truth*, infatti, si basa spesso su dati storici eterogenei non solo nella sostanza, ma anche nella forma, nella completezza e nella qualità delle informazioni disponibili. Prima di definire un dataset utilizzabile, è pertanto necessario un articolato processo di selezione, pulizia e standardizzazione dei dati, volto a eliminare i casi incompleti, incoerenti o non strutturati e a garantire la coerenza delle categorie utilizzate per l'addestramento del modello.

Le difficoltà associate alla costruzione di dataset affidabili sono ampiamente

documentate anche nella letteratura. Kolkman et al. (2024) [11] evidenziano, ad esempio, come lo sviluppo del sistema analizzato abbia richiesto oltre un anno di lavoro preliminare dedicato alla scrematura e alla normalizzazione del dataset, al fine di ottenere un insieme di dati coerente e utilizzabile per l'addestramento del modello. Analogamente, Mishra et al. (2024) [12] mostrano come dataset caratterizzati da una gamma limitata di caratteristiche osservabili possano compromettere la capacità del sistema di generalizzare a scenari reali più complessi.

Considerazioni analoghe emergono anche nel contesto applicativo del progetto di ricerca DARE, nel quale lo sviluppo del sistema di supporto decisionale ha richiesto un esteso lavoro preliminare di strutturazione delle categorie del dominio assicurativo e di costruzione del dataset di addestramento. Questo processo ha comportato la definizione di un'ontologia del dominio finalizzata alla classificazione degli eventi peritali, nonché la successiva annotazione dei casi storici utilizzati per l'addestramento del modello. In questo modo, il processo di classificazione tipico dell'attività professionale viene formalizzato e reso utilizzabile da un sistema algoritmico in grado di supportare l'identificazione delle categorie rilevanti nel processo decisionale.

In contesti professionali caratterizzati da processi decisionali altamente strutturati ma da elevata eterogeneità informativa dei casi analizzati, la qualità della rappresentazione dei dati e delle categorie utilizzate per l'addestramento del modello assume quindi un ruolo particolarmente critico. Tuttavia, come discusso nella sezione successiva, gran parte della letteratura empirica sull'interazione uomo-AI analizza il supporto algoritmico in contesti sperimentali semplificati, che difficilmente riflettono la complessità informativa tipica dei contesti professionali reali.

## **2.6 Limiti della letteratura esistente e posizionamento della ricerca**

Nonostante la solidità e l'ampiezza dei contributi analizzati, emerge un limite rilevante nella letteratura attuale. La maggior parte degli studi empirici sull'impiego di sistemi di intelligenza artificiale a supporto dei processi decisionali è infatti condotta in contesti sperimentali controllati, su task semplificati o artificialmente costruiti, che difficilmente riproducono l'incertezza informativa, procedurale e documentale tipica dei contesti decisionali reali. In tali studi, l'interazione uomo-AI viene spesso analizzata attraverso attività isolate, standardizzate e decontestualizzate, limitando la possibilità di generalizzare i risultati a scenari operativi complessi.

In particolare, una parte significativa della letteratura si concentra sul supporto fornito dall'AI a utenti umani non esperti, ma in relazione a compiti a bassa complessità cognitiva o scarsamente rappresentativi di un'attività professionale strutturata. In molti contesti professionali, tuttavia, il processo decisionale non si riduce all'esecuzione di un compito isolato, ma implica la classificazione di situazioni reali all'interno di categorie interpretative che guidano l'applicazione delle procedure operative. Come evidenziato nella letteratura organizzativa sul funzionamento delle professioni, tale attività di diagnosi e categorizzazione rappresenta una componente centrale del lavoro esperto. Nonostante ciò, la maggior parte degli studi sull'interazione uomo-AI tende a trascurare questa dimensione, concentrandosi su task semplificati che non riflettono la complessità dei processi di classificazione tipici dei contesti professionali reali. In questi casi, l'obiettivo degli studi è prevalentemente quello di analizzare l'effetto del supporto algoritmico sulle prestazioni individuali, sui bias cognitivi o sulle dinamiche di affidamento, senza interrogarsi sulla possibilità che tale supporto consenta all'utente di svolgere, in modo affidabile, compiti che nella pratica reale sono riservati a figure esperte.

Risultano pertanto ancora scarse le evidenze empiriche che affrontano esplicitamente il tema della sostituzione dell'esperto attraverso sistemi di supporto basati

su intelligenza artificiale, soprattutto in contesti professionali caratterizzati da elevata complessità informativa. Tale lacuna appare particolarmente rilevante nei domini tecnico-amministrativi e peritali, nei quali il processo decisionale richiede la valutazione congiunta di informazioni eterogenee, quali documentazione testuale non strutturata, immagini tecniche e stime di natura economica, ed è fortemente vincolato da procedure formali e da un elevato grado di discrezionalità umana.

Il presente lavoro di tesi si colloca all'interno di questo spazio di ricerca ancora poco esplorato e si distingue dalla letteratura esistente per l'obiettivo di valutare se e in che misura un utente inesperto, supportato da un sistema di intelligenza artificiale progettato per l'uso operativo reale, sia in grado di svolgere un compito decisionale tipicamente affidato a un esperto umano. A differenza di studi che analizzano l'interazione uomo-AI su task astratti o semplificati, **il presente lavoro adotta un contesto sperimentale che riproduce in modo realistico un'attività professionale concreta, quale la valutazione peritale in ambito assicurativo.**

In tal senso, il contributo della tesi non si limita a esaminare l'effetto del supporto AI sulle prestazioni decisionali o sulle dinamiche di affidamento, ma mira a fornire evidenze empiriche su un quesito di rilevanza pratica e applicativa: la possibilità che sistemi di supporto basati su intelligenza artificiale possano abilitare, almeno in parte, l'esecuzione di compiti complessi da parte di utenti non esperti, avvicinandosi alle condizioni operative di un mestiere reale. Il presente lavoro intende pertanto colmare tale lacuna, contribuendo alla letteratura sull'interazione uomo-AI attraverso un'analisi empirica condotta in un contesto decisionale realistico, caratterizzato da elevata complessità informativa, vincoli procedurali e centralità del giudizio umano.

# Capitolo 3

## Metodologia

### 3.1 Approccio metodologico

Il presente studio adotta un approccio sperimentale controllato finalizzato ad analizzare l'impatto di un sistema di supporto basato su intelligenza artificiale sul processo decisionale peritale di utenti inesperti. In particolare, l'esperimento è progettato per valutare come l'introduzione del tool influenzi l'accuratezza delle decisioni, la discrezionalità decisionale, le dinamiche di affidamento al sistema e i tempi di esecuzione del compito, rispetto a una condizione priva di supporto tecnologico.

L'attenzione è rivolta esclusivamente a **utenti inesperti**, al fine di osservare in modo diretto se e in che misura il supporto algoritmico consenta loro di avvicinare le proprie valutazioni a un riferimento ritenuto corretto, il cosiddetto *ground truth*, riducendo al contempo la variabilità e la soggettività delle decisioni. Il *ground truth* è definito sulla base di valutazioni peritali validate nell'ambito del progetto DARE ed è utilizzato come riferimento operativo condiviso, pur non rappresentando una verità oggettiva in senso assoluto (vedi paragrafo 2.5).

La scelta di coinvolgere utenti inesperti si inserisce nel solco di numerosi esperimenti presenti nella letteratura sull'interazione uomo-AI, che utilizzano disegni

sperimentali controllati per analizzare l'effetto dei suggerimenti algoritmici sulle decisioni individuali. Alcuni studi sono stati condotti in contesti professionali reali coinvolgendo esperti del dominio, come nel caso della diagnostica radiologica analizzata da Lebovitz et al. (2022) [10] o delle attività di consulenza strategica studiate da Dell'Acqua et al. (2023) [8]. Altri contributi adottano invece contesti sperimentali controllati con partecipanti non necessariamente esperti, come negli studi di Caplin et al. (2024) [18] sul supporto algoritmico in compiti di classificazione o di Campero et al. (2022) [13] e Peng et al. (2023) [14] sull'utilizzo dell'AI nello sviluppo software.

In linea con questo secondo filone, il presente esperimento coinvolge partecipanti privi di esperienza professionale nel dominio peritale. Tuttavia, al fine di ridurre il divario tra la conoscenza dei partecipanti e quella degli esperti che operano nel contesto reale, è stata introdotta una fase preliminare di formazione. Prima dell'esperimento i partecipanti hanno avuto accesso a un manuale introduttivo, a una serie di videopillole formative e a un breve corso esplicativo, progettati per fornire le conoscenze di base necessarie alla comprensione delle perizie e del processo valutativo. Questa fase di training consente di avvicinare il livello di familiarità dei partecipanti con il dominio decisionale a quello degli esperti, pur mantenendo i vantaggi metodologici di un esperimento controllato.

### 3.1.1 Disegno sperimentale

Il disegno sperimentale adottato combina elementi di un disegno *within-subject* e *between-subject*. Ogni partecipante è esposto a tutte le condizioni sperimentali previste, ma ciascuna perizia viene presentata una sola volta per evitare effetti di apprendimento e familiarità con i singoli casi. Questa scelta consente di controllare la variabilità inter-individuale legata alle capacità cognitive e alle differenze personali, permettendo un confronto diretto delle prestazioni dello stesso soggetto in presenza e in assenza del supporto AI. In particolare, l'esperimento include **tre condizioni**: (i) assenza di supporto AI, (ii) supporto AI con suggerimento corretto,

(iii) supporto AI con suggerimento errato. L'inclusione di suggerimenti errati consente di analizzare non solo i benefici del supporto algoritmico, ma anche le dinamiche di affidamento e condizionamento decisionale in presenza di informazioni fuorvianti.

Per evitare effetti di apprendimento e di familiarità con i singoli casi, ciascuna perizia viene presentata una sola volta a ogni partecipante. Al contempo, l'adozione di uno schema di controbilanciamento tra partecipanti consente di garantire che, a livello aggregato, ogni perizia venga valutata in tutte le condizioni sperimentali. Questa impostazione permette di combinare i vantaggi di un disegno *within-subject*, utilizzato per analizzare l'accuratezza delle valutazioni e i tempi di esecuzione, con la possibilità di stimare la discrezionalità decisionale su gruppi indipendenti di partecipanti.

In particolare, mentre il confronto intra-soggetto consente di isolare l'effetto del supporto AI sulle prestazioni individuali, la distribuzione delle stesse perizie su condizioni diverse e su partecipanti distinti rende possibile misurare la variabilità delle risposte associate a ciascuna combinazione perizia-condizione. La discrezionalità decisionale viene così operazionalizzata attraverso la dispersione delle risposte (misurata tramite deviazione standard), evitando che essa sia contaminata da effetti di apprendimento o da dipendenze statistiche tra osservazioni ripetute.

Questa scelta metodologica rappresenta un'estensione coerente del disegno sperimentale adottato da Jussupow et al. (2021) [16], nel quale un disegno *within-subject* viene utilizzato per analizzare l'interazione tra giudizio umano e supporto algoritmico, garantendo un elevato controllo della variabilità individuale. Rispetto a tale approccio, il presente studio introduce un ulteriore livello di controllo, separando analiticamente le misure di performance individuale da quelle di variabilità intersoggettiva. Questa impostazione consente inoltre di analizzare esplicitamente il grado di discrezionalità decisionale tra i partecipanti, dimensione che risulta ancora relativamente poco esplorata nella letteratura sperimentale sull'interazione tra giudizio umano e supporto algoritmico.

Un'impostazione analoga, seppur applicata a un diverso dominio, è riscontrabile anche nello studio di Kolkman et al. (2024) [11], in cui l'impatto di un sistema di intelligenza artificiale sulle decisioni legali viene analizzato attraverso un esperimento controllato con e senza assistenza AI. In quel contesto, i partecipanti valutano gli stessi casi giuridici in entrambe le condizioni, consentendo un confronto diretto tra decisioni supportate e non supportate dall'algoritmo. I risultati mostrano come l'introduzione dell'AI influenzi in modo significativo il processo decisionale umano, riducendo la percentuale di valutazioni conformi al giudizio del giudice dal 67% al 48%. Tali evidenze rafforzano la rilevanza di un disegno sperimentale in grado di cogliere non solo le variazioni di accuratezza, ma anche le trasformazioni più sottili nei meccanismi decisionali indotte dall'informazione algoritmica.

Tutti i partecipanti sono posti in una condizione **blind**: non hanno visionato in precedenza le immagini peritali né i materiali oggetto di valutazione. Questo accorgimento metodologico è volto a evitare bias di familiarità e a garantire che le decisioni siano prese esclusivamente sulla base delle informazioni fornite durante la sessione sperimentale.

Le valutazioni espresse dai partecipanti sono raccolte tramite **scala binaria** e **scala Likert a cinque punti**. Ai fini dell'analisi comparativa con quanto espresso dal *ground truth*, la scala Likert viene convertita in formato binario seguendo il criterio semantico strutturale di Suarez et al. (2024) [19]: le categorie 1–2 sono ricondotte alla classe negativa, le categorie 4–5 alla classe positiva, mentre la categoria centrale (3), che esprime esplicitamente una posizione di incertezza o neutralità, viene esclusa dall'analisi. Tale criterio di dicotomizzazione non deriva da un'ottimizzazione data-driven, ma dalla natura semantica della scala ed è coerente con quanto proposto in letteratura da Suarez et al. (2024) [19], garantendo la preservazione dell'informazione e la comparabilità dei risultati. Le metriche di accuratezza, discrezionalità decisionale, affidamento al sistema e tempi di esecuzione derivate da tale impostazione sono definite e discusse in dettaglio nel paragrafo 3.4.

Il numero di perizie somministrate a ciascun partecipante è fissato a 12. Tale

scelta è motivata sia da considerazioni operative emerse nel progetto DARE, che indicano un tempo medio di 4–5 minuti per perizia, sia da evidenze empiriche presenti in letteratura, che mostrano come un numero di item compreso tra 10 e 20 sia necessario per ottenere stime stabili di accuratezza e discrezionalità in compiti decisionali complessi. La numerosità degli item è inoltre calibrata per contenere il carico cognitivo del partecipante ed evitare effetti di affaticamento nelle fasi finali dell’esperimento.

Un ulteriore elemento di convergenza deriva dalla tesi di Wiles (2024) [20], che nel capitolo sperimentale dedicato all’uso dell’AI come supporto decisionale impiega sistematicamente set di 12 valutazioni per partecipante. Anche in questo caso, gli effetti dell’assistenza algoritmica sulla coerenza interna e sulla convergenza dei giudizi emergono solo con un numero di item sufficiente a compensare la variabilità intrinseca del comportamento umano in compiti interpretativi. Gli altri articoli analizzati, quali Brynjolfsson et al. (2025) [15], Noy et al. (2023) [5] e Peng et al. (2023) [14], pur appartenendo a domini differenti (scrittura, programmazione, produttività), confermano implicitamente lo stesso principio statistico: gli studi sperimentali che misurano performance, accuratezza o variazione della qualità richiedono almeno una dozzina di task per persona.

In generale, compiti che perseguono la stima dell’accuratezza e della discrezionalità (specialmente in contesti in cui l’AI influenza il giudizio umano) necessitano di un numero minimo di osservazioni per partecipante compreso tra 10 e 20. Alla luce di tale convergenza metodologica, e considerata la natura semi-strutturata delle perizie dell’esperimento (analoghe per complessità ai casi legali e diagnostici analizzati negli articoli), è stato fissato un numero di 12 perizie per partecipante, ritenuto il valore empiricamente più solido per ottenere stime stabili di accuratezza, discrezionalità e coerenza decisionale nelle due condizioni sperimentali previste (con e senza AI).

Nel complesso, il disegno metodologico adottato consente di isolare in modo rigoroso l’effetto del supporto AI sul processo decisionale degli utenti inesperti,

mantenendo il controllo sulla variabilità individuale, sugli effetti di apprendimento e sulle differenze intrinseche tra le perizie.

## 3.2 Architettura dell'esperimento, partecipanti e contesto sperimentale

L'architettura dell'esperimento è stata progettata in modo coerente con il disegno metodologico descritto nel paragrafo precedente, con l'obiettivo di garantire un confronto controllato tra le diverse condizioni sperimentali e di osservare in modo sistematico l'impatto del supporto basato su intelligenza artificiale sul processo decisionale degli utenti. In particolare, la struttura dell'esperimento è finalizzata a isolare l'effetto del supporto AI mantenendo sotto controllo la variabilità individuale, l'eterogeneità delle perizie e le condizioni di fruizione dei materiali decisionali.

### 3.2.1 Partecipanti e contesto sperimentale

I partecipanti all'esperimento sono studenti di ingegneria, selezionati in quanto utenti inesperti rispetto all'attività peritale oggetto di studio. Tale scelta consente di analizzare in modo diretto l'effetto del supporto AI su soggetti privi di esperienza pregressa nel dominio applicativo e, al contempo, di ridurre al minimo eventuali differenze cognitive legate a background formativi eterogenei. Tutti i partecipanti prendono parte all'esperimento su base volontaria e sono informati in anticipo sulle modalità di svolgimento, sulla durata complessiva della prova e sul trattamento dei dati raccolti.

Ai partecipanti viene chiaramente specificato che **non è richiesto di svolgere analisi tecniche** specialistiche analoghe a quelle effettuate da periti professionisti, ma di immedesimarsi nel ruolo decisionale peritale ed esprimere il proprio giudizio sulla base delle informazioni fornite e della formazione ricevuta nella fase preliminare all'esperimento.

I partecipanti possono completare il test da remoto, in un arco temporale predefinito, utilizzando un unico tentativo di svolgimento; il tempo di esecuzione viene registrato automaticamente dalla piattaforma. I partecipanti sono posti in una condizione *blind*: non hanno avuto accesso preventivo alle immagini peritali né ai materiali di valutazione utilizzati durante la prova. Oltre alle risposte alle domande sperimentali, viene raccolta anche l'informazione relativa alla media accademica dei partecipanti, al fine di consentire analisi esplorative sull'eventuale relazione tra rendimento universitario e prestazioni decisionali.

### 3.2.2 Condizioni sperimentali, controbilanciamento e procedura

In questo contesto, l'esperimento adotta un disegno *within-subject* articolato in tre condizioni sperimentali:

- assenza di supporto AI;
- supporto AI con suggerimento corretto;
- supporto AI con suggerimento errato.

Ciascun partecipante è esposto a tutte le condizioni sperimentali, mentre ogni singola perizia viene visualizzata una sola volta da ciascun soggetto, così da evitare effetti di apprendimento diretto sullo stesso item.

Le dodici perizie selezionate sono assegnate alle condizioni sperimentali secondo uno schema di **controbilanciamento tra partecipanti**. In questo modo ciascun partecipante valuta perizie differenti in ciascuna condizione, mentre ogni singola perizia viene complessivamente osservata in tutte le condizioni sperimentali da gruppi distinti di partecipanti. Cosicché nessun soggetto è esposto più volte alla stessa perizia, ma l'insieme dei partecipanti garantisce una copertura completa di tutte le combinazioni perizia–condizione.

Questa strategia di assegnazione consente di evitare effetti di apprendimento legati alla ripetizione degli stessi casi e, al contempo, di controllare l'eterogeneità

intrinseca tra le perizie, assicurando la comparabilità delle condizioni sperimentali e la validità interna del disegno di ricerca.

Dal punto di vista operativo, l'esperimento si articola in:

**Fase preliminare di formazione.**

I partecipanti svolgono una fase preliminare di formazione, finalizzata a fornire tutte le informazioni necessarie per la corretta comprensione del contesto sperimentale e delle attività richieste. La formazione si articola in una componente asincrona e in un breve momento sincrono online della durata di circa 30 minuti, svolto qualche giorno prima dell'avvio del test.

La componente asincrona è costituita da un **manuale operativo** e da una serie di **video pillole introduttive**, attraverso i quali vengono illustrati il contesto dell'esperimento, il ruolo richiesto ai partecipanti, le modalità di svolgimento della prova e la durata complessiva dell'attività.

Nel materiale formativo vengono inoltre presentati i criteri generali per l'analisi delle perizie, con particolare riferimento all'identificazione delle cause e degli effetti del danno, alla valutazione dell'indennizzabilità delle perizie e alla formulazione della stima economica.

La componente sincrona, della durata di circa 30 minuti, si svolge in modalità online ed è finalizzata a uniformare ulteriormente il livello di comprensione dei partecipanti prima dell'inizio della prova. Durante tale sessione vengono ripresi sinteticamente i contenuti principali delle video-pillole e viene chiarita la struttura operativa del test.

Al termine della spiegazione è previsto uno spazio dedicato a eventuali domande di chiarimento. Le richieste dei partecipanti possono riguardare esclusivamente aspetti organizzativi e procedurali, senza fornire indicazioni circa la correttezza delle risposte o anticipazioni sui contenuti delle perizie oggetto del test. Tale scelta è volta a garantire omogeneità informativa e a preservare la validità interna dell'esperimento.

**Materiali di supporto.** Ai partecipanti vengono forniti materiali di supporto

sotto forma di un manuale operativo sintetico e di brevi video-pillole tematiche. Il manuale, della lunghezza di circa dieci pagine, è strutturato nei seguenti capitoli:

- introduzione all'esperimento e ruolo dei partecipanti;
- definizione generale di perizia e sua struttura;
- cause del danno e criteri di riconoscimento, corredati da immagini esplicative;
- tipologie di danno e modalità di identificazione, supportate da esempi visivi;
- Suggerimento AI;
- stima economica del danno e interpretazione dei range di prezzo associati a cause e danni;
- criteri generali per la valutazione dell'indennizzabilità della perizia

Le video-pillole riprendono la struttura e i contenuti del manuale, utilizzando le stesse immagini ed esempi, con l'obiettivo di uniformare il livello di preparazione iniziale dei partecipanti. Il manuale può essere consultato anche durante lo svolgimento dell'esperimento.

**Svolgimento dell'esperimento.** Il test completato da da remoto, in un arco temporale predefinito, utilizzando un unico tentativo di svolgimento; il tempo di esecuzione viene registrato automaticamente dalla piattaforma. I partecipanti valutano le dodici perizie assegnate attraverso un'interfaccia strutturata che presenta il materiale visivo e testuale necessario alla valutazione. Nelle condizioni con supporto AI, l'interfaccia mostra anche il suggerimento prodotto dal sistema, sotto forma testuale. Nella condizione senza AI tali informazioni non sono presenti.

Per ciascuna perizia, i partecipanti rispondono a **quattro domande a risposta chiusa** su scala Likert a cinque punti o su scala binaria, progettate per raccogliere informazioni sulla valutazione tecnica del caso, sulla stima economica e sull'indennizzabilità della perizia.

La scelta di utilizzare una scala Likert a cinque punti è coerente con approcci adottati in studi precedenti in letteratura (si veda, ad esempio, Mishra et al., 2024 [12]) ed è motivata dall'esigenza di offrire ai partecipanti una gamma più articolata di giudizi. Rispetto a una risposta binaria, la scala Likert consente infatti di esprimere gradazioni di accordo o incertezza, riducendo il rischio che i partecipanti forniscano risposte casuali o forzate in presenza di ambiguità interpretative. Tale scelta risulta particolarmente appropriata nel contesto di valutazioni peritali simulate, caratterizzate da livelli variabili di complessità e da margini di incertezza intrinseci al compito decisionale. La risposta binaria sarà utilizzata solo per la domanda relativa all'indennizzabilità, in quanto tale scelta nell'ambito assicurativo non dipende dalle percezioni del perito e dalla sua discrezionalità decisionali, ma è legata ad una serie di criteri ben precisi.

### **3.3 Costruzione dei materiali decisionali**

Al fine di garantire la coerenza metodologica dell'esperimento e ridurre la variabilità non controllata nelle risposte dei partecipanti, è stata progettata una fase strutturata di costruzione dei materiali decisionali. Considerato che i soggetti coinvolti sono utenti inesperti rispetto all'attività peritale, tali materiali non sono stati concepiti con l'obiettivo di fornire una formazione tecnica approfondita, bensì di mettere i partecipanti nelle condizioni di comprendere il contesto valutativo, interpretare correttamente le informazioni disponibili ed esprimere giudizi coerenti e comparabili tra loro.

I materiali decisionali sono stati interamente sviluppati dall'autrice nell'ambito del progetto di ricerca e comprendono: (i) un manuale operativo, (ii) una serie di video-pillole formative, (iii) tabelle semplificate per la stima economica del danno a e (iv) un corso online introduttivo. L'insieme di tali strumenti è stato progettato per uniformare il livello di preparazione iniziale dei partecipanti e ridurre l'influenza di interpretazioni arbitrarie sul processo decisionale.

### 3.3.1 Manuale operativo

Il manuale operativo, della lunghezza di circa dieci pagine, è stato redatto dall'autrice ed è strutturato in *sette capitoli* principali. Il primo capitolo introduce il contesto dell'esperimento e chiarisce il ruolo richiesto ai partecipanti, specificando che non è loro richiesto di svolgere analisi tecniche di livello professionale, ma di immedesimarsi nel ruolo del perito sulla base delle informazioni fornite. Il secondo capitolo fornisce una definizione generale di perizia, descrivendone la struttura e le principali componenti informative.

Il terzo e il quarto capitolo sono dedicati rispettivamente alla causa del danno e al danno stesso. In queste sezioni vengono descritte le principali tipologie riscontrabili e i criteri per il loro riconoscimento, supportati da immagini esemplificative e suggerimenti operativi basati sull'esperienza maturata dall'autrice come ricercatrice nel progetto. Rispettivamente è possibile vedere alcuni esempi delle cause del danno in Figura 3.1 ed esempi di alcuni tra i possibili danni in figura 3.2.

Il quinto capitolo introduce la lettura del suggerimento AI sottoforma testuale, rimandando per gli aspetti tecnici di funzionamento al paragrafo 3.5 dedicato allo strumento di supporto.

Il sesto capitolo è dedicato alla stima economica e descrive le modalità di interpretazione dei range di prezzo associati alle diverse cause e tipologie di danno. In particolare, la stima economica complessiva associata a ciascuna perizia è ottenuta come somma del costo necessario per la ricerca e il ripristino della causa del danno e del costo relativo alla riparazione del danno stesso. A tal fine, sono state sviluppate due tabelle di supporto, costruite in forma semplificata per agevolare il processo valutativo. La Tabella 3.1 associa a ciascuna possibile tipologia di causa il corrispondente costo di ricerca e ripristino, mentre la Tabella 3.2 riporta i costi di riparazione del danno in funzione dell'elemento del fabbricato interessato e delle caratteristiche della finitura. Entrambe le tabelle sono incluse nel manuale fornito ai partecipanti e costituiscono un riferimento operativo per supportare la formulazione della stima economica durante la valutazione delle perizie.

<b>Dettaglio causa</b>	<b>Costo totale Ricerca e Ripristino [€]</b>
Rottura braga	600–750
Rottura tubo carico	650–700
Rottura tubo scarico acque grigie bidet/lavelli (immurate)	350–400
Rottura tubo scarico acque grigie doccia	750 (+200 se vasca da sostituire)
Rottura tubo scarico acque grigie vasca	600 (+200 manodopera +200 piatto doccia base)
Rottura tubo scarico acque nere (indoor, rottura tratto privato)	500–600
Rottura tubo scarico acque nere (indoor, tra più piani)	1000–1200 (quota parte se condominio)
Rottura tubo riscaldamento	650–750
Rottura tubo riscaldamento a pavimento	1100–1200
Rottura cassetta WC incasso	600–650
Rottura cassetta WC esterna	300–400
Rottura guarnizione	60
Rottura saracinesca bagno (valvole singole sanitari)	150–200

Tabella 3.1: Costi indicativi di ricerca e ripristino per le principali cause di danno idrico

Le tabelle utilizzate nel presente studio derivano da versioni più articolate sviluppate nell’ambito del progetto di ricerca DARE e validate rispetto al *ground truth*, con l’obiettivo di fornire una base di riferimento coerente per il calcolo automatico delle stime economiche da parte dello strumento di supporto basato su intelligenza artificiale. Al fine di renderle adeguate al contesto sperimentale e utilizzabili da utenti non esperti, tali tabelle sono state successivamente semplificate, eliminando i dettagli altamente tecnici e mantenendo esclusivamente le informazioni essenziali. In questo modo, un partecipante dotato di una formazione di base, fornita tramite il manuale, le video-pillole e il breve corso online, è in grado di effettuare una stima economica coerente entro i tempi limitati previsti per la fase di formazione. Le medesime tabelle, nella loro versione semplificata, sono inoltre impiegate dal tool di supporto basato su intelligenza artificiale descritto nel paragrafo successivo.

<b>Elemento fabbricato</b>	<b>Caratteristiche finitura</b>	<b>Etichetta segmentazione primaria</b>	<b>Costo totale ad elemento</b>
Muro	Intonaco	Macchia, gocciolamento	200–300 (ogni 20 mq)
Muro	Intonaco	Muffa	50 (ogni 10 mq)
Muro	Intonaco	Scrostamento	50 (ogni 10 mq)
Muro	Piastrelle	Gocciolamento, fughe bagnate, rilazamento	250–300 (ogni 10 mq)
Muro	Carta da parati	Macchia, gocciolamento, scrostamento	600 (ogni 10 mq)
Soffitto	Intonaco	Macchia, gocciolamento	200–250 (ogni 20 mq)
Pavimento	Parquet	Spargimento di liquidi	200 + 100 (ogni 10 mq) per ogni stanza
Pavimento	Parquet	Rilazamento	300–600 (ogni 5 mq)
Pavimento	Ceramica – piastrelle	Fughe bagnate, rilazamento, spargimento di liquidi	250 (ogni 5 mq)
Zoccolo	Legno	Rilazamento, spargimento di liquidi	100 (ogni 10 m lineari)

Tabella 3.2: Costi indicativi di ripristino in funzione dell'elemento fabbricato, della finitura e del danno osservato

Il settimo ed ultimo capitolo affronta l'indennizzabilità della perizia (unico tema in cui la risposta richiesta sarà di tipo binario). In particolare, tale scelta si baserà su dei criteri ben precisi quali: presenza di un danno visibile, causa riconducibile a fuoriuscita d'acqua o malfunzionamento di un impianto idrico, livello di manutenzione dell'immobile, coerenza tra causa dichiarata e danno osservabile, spese conseguenti al danno. La perizia può essere considerata indennizzabile se sono contemporaneamente soddisfatte tutte le seguenti condizioni, in mancanza anche di una sola di queste, la perizia non è indennizzabile.



(a) Rottura tubo scarico wc



(b) Rottura tubo carico acqua

Figura 3.1: Esempi possibili cause



(a) Muffa



(b) Scrostamento

Figura 3.2: Esempi possibili danni

Il manuale completo è reso disponibile ai partecipanti e può essere consultato anche durante lo svolgimento dell'esperimento. Una versione integrale del manuale è riportata in Appendice A.

### 3.3.2 Corso introduttivo e video-pillole formative

Prima dello svolgimento dell'esperimento è stata predisposta una breve fase di formazione preliminare, finalizzata a fornire ai partecipanti le conoscenze di base necessarie per comprendere il contesto decisionale in cui si sarebbero trovati ad operare. Tale fase si è articolata in due componenti complementari: un corso online

introduttivo svolto alcuni giorni prima dell'esperimento e una serie di video-pillole formative rese disponibili insieme al manuale.

Il **corso online**, della durata di circa 30 minuti, ha avuto l'obiettivo di introdurre in modo generale il contesto assicurativo dal punto di vista dell'attività peritale e delle principali logiche che guidano il lavoro dei periti nella valutazione dei sinistri. Durante l'incontro sono stati illustrati i concetti fondamentali relativi alla struttura di una perizia e ai criteri generalmente utilizzati nella valutazione dei danni. Questa fase preliminare ha permesso di fornire ai partecipanti un primo inquadramento del contesto decisionale in cui si sarebbe svolto l'esperimento, contribuendo a creare una base di conoscenze condivisa tra i partecipanti e a ridurre possibili differenze legate al diverso livello di familiarità con il tema trattato.

A supporto del manuale sono state inoltre realizzate brevi **video-pillole formative**, ideate e registrate dall'autrice sulla base di slide appositamente predisposte. Ogni video ha una durata media di circa due minuti e corrisponde a uno dei capitoli del manuale, riprendendone i contenuti essenziali e utilizzando le stesse immagini esplicative. La scelta di mantenere una durata contenuta per ciascun video è stata adottata per favorire una fruizione più immediata dei contenuti e per ridurre il rischio di sovraccarico cognitivo dei partecipanti, suddividendo le informazioni in unità brevi e facilmente assimilabili.

Le video-pillole affrontano in modo sintetico i principali passaggi della valutazione di una perizia assicurativa. In particolare, i contenuti trattati riguardano: (1) l'introduzione e le finalità dell'esperimento; (2) la definizione generale di perizia; (3) l'identificazione della causa del danno; (4) il riconoscimento dell'effetto del danno attraverso l'osservazione delle immagini; (5) la comprensione della stima economica della causa e del danno accertato; e (6) i criteri generali utilizzati per valutare l'indennizzabilità della perizia.

Le video-pillole sono state rese disponibili ai partecipanti insieme al manuale e potevano essere fruite autonomamente prima dell'esperimento e durante. L'utilizzo combinato di manuale, corso introduttivo e contenuti audiovisivi è stato progettato

con l'obiettivo di rendere la fase di preparazione quanto più possibile accessibile e omogenea, riducendo eventuali differenze individuali nel livello di familiarità con il tema assicurativo prima dello svolgimento del test.

### 3.3.3 Scelta delle perizie

La selezione delle dodici perizie utilizzate nell'esperimento è stata effettuata attraverso un processo strutturato, basato sull'intersezione di due insiemi di dati sviluppati nell'ambito del progetto di ricerca. Il primo insieme è costituito da un documento di perizie selezionate per test di stima economica, redatto durante il progetto e successivamente rivisto e validato rispetto al *ground truth*, rappresentato in questo caso dal giudizio del perito esperto. Tale documento contiene perizie per le quali sono disponibili stime economiche considerate affidabili e coerenti.

Il secondo insieme deriva da un'analisi quali-quantitativa condotta da altri ricercatori del progetto, nella quale a ciascuna perizia sono stati assegnati pesi relativi alle diverse caratteristiche informative presenti (ad esempio qualità delle immagini, presenza di fotografie del danno e della causa, completezza della documentazione). Sulla base di tali pesi, le perizie sono state classificate in categorie di qualità ("golden", "accettabili" e "da scartare").

Le perizie utilizzate nell'esperimento sono state selezionate come intersezione tra le perizie classificate come "golden" e quelle presenti nel file di stima economica validato. Questa scelta consente di disporre di casi che presentano tutte le caratteristiche informative principali e, al contempo, stime economiche coerenti con il *ground truth*, riducendo il rischio di ambiguità legate alla qualità del materiale di partenza.

### 3.4 Progettazione delle domande e implementazione delle condizioni sperimentali

Al fine di garantire il controllo delle condizioni sperimentali e la coerenza delle valutazioni proposte ai partecipanti, la progettazione delle domande e dei suggerimenti associati a ciascuna perizia è stata effettuata ex ante attraverso una struttura formale di definizione delle condizioni sperimentali.

Per ciascuna perizia selezionata, sono state definite in modo univoco:

- le affermazioni sottoposte ai partecipanti (relative a causa del danno, tipologia di danno e stima economica);
- la corrispondenza tra ciascuna affermazione e il riferimento ritenuto corretto sulla base del ground truth;
- la configurazione del suggerimento AI, distinguendo tra suggerimenti corretti e suggerimenti errati.

Le perizie sono state suddivise in due gruppi bilanciati, in modo tale che una parte fosse associata ad affermazioni corrette e l'altra ad affermazioni intenzionalmente scorrette. Tale associazione è rimasta invariata in tutte le condizioni sperimentali, così da evitare che differenze nelle prestazioni dei partecipanti potessero dipendere dalla formulazione delle domande anziché dalla presenza o assenza del supporto AI.

Le affermazioni sottoposte ai partecipanti (relative a causa del danno, tipologia di danno e stima economica) sono state definite in modo univoco per ciascuna perizia e mantenute **costanti in tutte le condizioni sperimentali**. In altri termini, per ogni perizia è stata stabilita ex ante la formulazione delle affermazioni e la relativa direzionalità (affermazione coerente con il riferimento oppure deliberatamente non coerente), così da garantire che eventuali differenze nelle risposte dipendano dalla condizione sperimentale e non da variazioni nella costruzione del questionario.

Per quanto riguarda il **supporto basato su intelligenza artificiale**, la configurazione del suggerimento non è fissata in modo univoco sulla singola perizia,

ma dipende dalla **condizione sperimentale**: nella condizione senza AI non viene mostrata alcuna informazione algoritmica; nella condizione con AI corretta viene presentato un suggerimento coerente con il riferimento adottato; nella condizione con AI errata viene invece presentato un suggerimento non coerente. Tale impostazione consente di isolare l'effetto dell'introduzione del supporto algoritmico e della sua affidabilità, mantenendo invariato il contenuto informativo delle perizie e delle domande tra condizioni.

### 3.5 Strumento di supporto basato su intelligenza artificiale

Lo strumento di supporto basato su intelligenza artificiale utilizzato nel presente studio è stato sviluppato nell'ambito del progetto di ricerca DARE presso il Politecnico di Torino, con l'obiettivo di assistere il processo decisionale peritale in contesti assicurativi caratterizzati da elevata eterogeneità informativa. Il sistema non è progettato per sostituire il giudizio umano, ma per fornire un supporto informativo strutturato a utenti non esperti durante l'analisi delle perizie.

Nel **contesto applicativo reale**, il tool AI è concepito per operare direttamente sulle immagini peritali, fornendo un **suggerimento di natura visiva** in affiancamento alle fotografie originali. Tale suggerimento assume la forma di una **segmentazione grafica** sovrapposta alle immagini, accompagnata da una **legenda cromatica** che consente di interpretare il significato delle aree evidenziate (vedi Figura 3.3).

I colori presenti nella legenda rappresentano differenti tipologie di informazione visiva rilevanti ai fini della valutazione peritale, tra cui:

- aree in cui sono visibili macchie, muffe o segni di bagnamento;
- elementi riconducibili a possibili rotture o danneggiamenti dell'impianto;

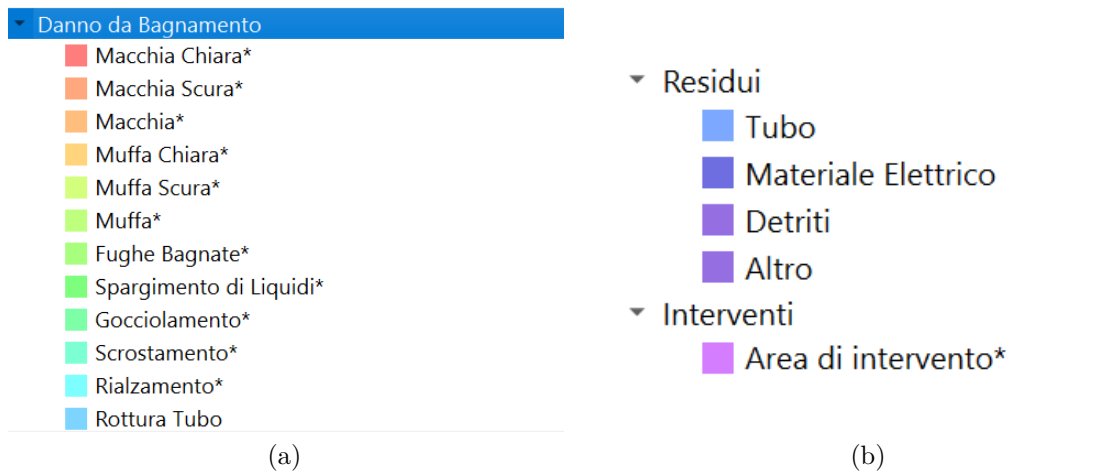


Figura 3.3: Legenda cromatica del suggerimento visivo fornito di supporto AI nel tool reale

- aree di intervento evidenziate (in viola), che indicano esclusivamente la collocazione spaziale del possibile danno o intervento, senza identificarne la tipologia specifica.

È importante sottolineare che le aree di intervento evidenziate non costituiscono una classificazione del danno, ma forniscono un'indicazione puramente spaziale. L'interpretazione della natura del danno rimane demandata all'utente, sulla base delle informazioni disponibili. In tal senso, il suggerimento visivo è concepito come un supporto all'osservazione e non come una risposta prescrittiva.

A titolo esemplificativo, in Figura 3.4 è riportato un esempio del suggerimento visivo prodotto dal sistema di supporto AI nel contesto applicativo reale. L'immagine mostra una fotografia peritale originale affiancata all'output del modello di analisi automatica, in cui sono evidenziate le aree di interesse mediante segmentazione grafica e codifica cromatica, coerentemente con la legenda riportata in Figura 3.3.

Le aree evidenziate rappresentano possibili zone di danno o di intervento individuate dal sistema sulla base delle caratteristiche visive rilevate. Come già sottolineato, tali evidenziazioni forniscono un'indicazione puramente spaziale e non costituiscono una classificazione automatica del danno, lasciando all'utente il compito di interpretarne la natura e la rilevanza peritale.

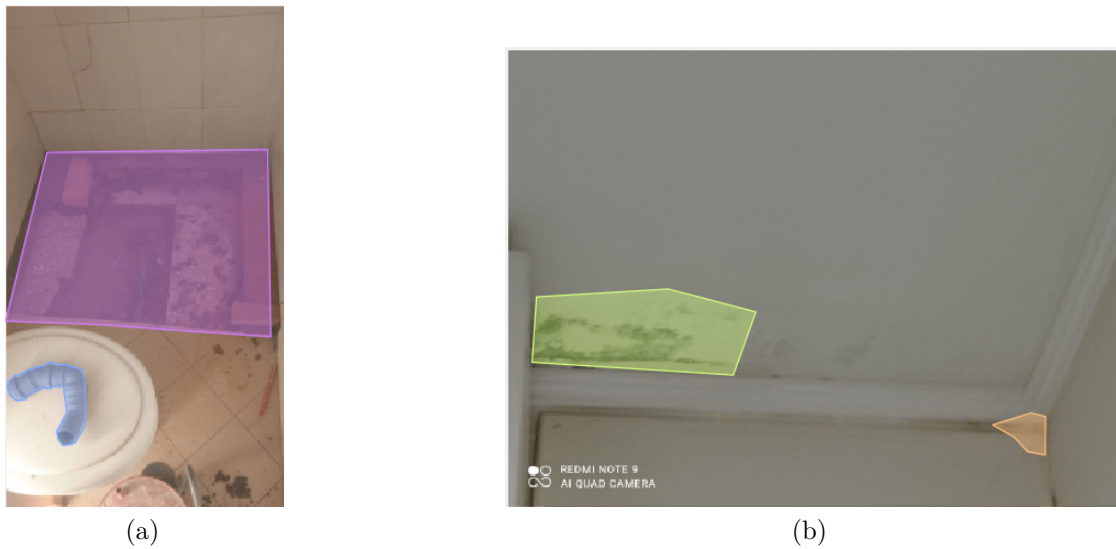


Figura 3.4: Esempi di suggerimenti visivi forniti dal supporto AI nel tool reale

### 3.5.1 Implementazione sperimentale del supporto AI

Per esigenze di controllo sperimentale e di standardizzazione delle condizioni di presentazione dello stimolo, nel presente studio il supporto fornito dal sistema di intelligenza artificiale **non è veicolato attraverso elementi visivi**, bensì sotto forma di **suggerimento testuale**. Tale suggerimento sintetizza l'output del sistema AI relativo alla valutazione del caso peritale e viene mostrato al partecipante esclusivamente in alcune condizioni sperimentali.

La scelta di utilizzare un suggerimento testuale consente di isolare l'effetto informativo del supporto AI, riducendo possibili confondimenti legati alla complessità percettiva delle immagini e garantendo una maggiore comparabilità tra le diverse condizioni sperimentali. Il contenuto del suggerimento testuale è coerente con le valutazioni che il sistema produrrebbe nel contesto reale, pur rinunciando alla componente di evidenziazione grafica.

Coerentemente con il disegno sperimentale adottato, i suggerimenti forniti dal sistema di intelligenza artificiale **non sono sempre corretti**. In alcune perizie il supporto risulta accurato, mentre in altre è parzialmente corretto o deliberatamente errato. Questa scelta metodologica consente di analizzare non solo l'effetto del

supporto AI sull'accuratezza delle decisioni, ma anche le dinamiche di affidamento, di condizionamento e di uso critico del suggerimento algoritmico da parte di utenti inesperti.

Ai partecipanti viene esplicitamente indicato che il suggerimento AI deve essere considerato come un'informazione aggiuntiva e non come una risposta da accettare automaticamente. Il giudizio finale rimane interamente in capo all'utente, che è chiamato a formulare la propria valutazione sulla base delle immagini peritali, delle indicazioni fornite nel manuale operativo e della propria interpretazione complessiva del caso.

Infine, il sistema fornisce indicazioni di supporto alla stima economica del danno basate su tabelle di valorizzazione semplificate sviluppate nell'ambito del progetto di ricerca. Tali tabelle, coerenti con quelle utilizzate dal tool AI nel contesto applicativo reale, consentono agli utenti inesperti di orientarsi nella formulazione di una stima economica plausibile, mantenendo la coerenza con il *ground truth* definito dal perito esperto.

Il supporto AI viene presentato ai partecipanti sotto forma di un breve suggerimento testuale, formulato in linguaggio naturale e mostrato insieme alle immagini peritali. A titolo esemplificativo, un possibile suggerimento fornito dal sistema è il seguente:

*“Suggerimento AI: La causa del danno è la rottura di del condotto di riscaldamento a pavimento.”*

Il suggerimento non fornisce indicazioni prescrittive sulla risposta corretta né sulla stima economica finale, ma rappresenta un'informazione aggiuntiva che il partecipante può decidere se integrare o meno nel proprio processo decisionale.

### **3.6 Variabili, metriche e strumenti di analisi**

Al fine di valutare in modo sistematico l'impatto del sistema di supporto basato su intelligenza artificiale sul processo decisionale degli utenti inesperti, l'esperimento

prevede la misurazione di quattro dimensioni principali: (i) la discrezionalità decisionale, (ii) l'accuratezza delle valutazioni rispetto a un riferimento ritenuto valido (*ground truth*), (iii) le modalità di affidamento al suggerimento AI e (iv) il tempo di esecuzione del compito. Ciascuna dimensione è associata a specifiche metriche di misura e a strumenti di analisi coerenti con il disegno sperimentale adottato. In questa sezione vengono descritte le principali variabili considerate nell'analisi empirica e le modalità con cui sono state operazionalizzate. La Tabella 3.3 fornisce una sintesi delle dimensioni analizzate, delle metriche utilizzate e dell'unità di analisi adottata per ciascuna variabile.

Tabella 3.3: Sintesi delle variabili analizzate

Variabile	Cosa misura	Metodo di analisi
Discrezionalità decisionale	Variabilità delle valutazioni tra partecipanti	Deviazione standard Test t di Student
Accuratezza delle valutazioni	Correttezza delle decisioni rispetto al <i>ground truth</i>	Confusion matrix Test chi-quadro
Affidamento al sistema AI	Tendenza a seguire il suggerimento dell'AI	Follow rate
Tempo di esecuzione	Efficienza nello svolgimento del compito	Test t di Student

### 3.6.1 Discrezionalità decisionale

La discrezionalità decisionale viene interpretata come la variabilità delle valutazioni fornite dai partecipanti in risposta alle stesse tipologie di perizie. In linea con la letteratura sui processi decisionali supportati da sistemi automatizzati (vedi Jussupow et al. (2021)[16]), una maggiore dispersione delle risposte è indicativa di un elevato grado di discrezionalità individuale, mentre una riduzione della variabilità suggerisce una maggiore convergenza valutativa indotta dal supporto tecnologico.

Nel presente studio, la discrezionalità decisionale viene misurata attraverso **l'analisi della dispersione** delle risposte fornite dai partecipanti su scala Likert

e binaria, calcolata separatamente per la condizione senza AI e per le condizioni con supporto AI. L'unità di analisi è rappresentata dalla singola perizia all'interno di ciascuna condizione sperimentale. Per ogni perizia e per ciascun item valutativo prima su scala Likert e poi su scala binaria, è stata calcolata la deviazione standard delle risposte fornite dai partecipanti appartenenti alla medesima condizione sperimentale. Formalmente, per una data perizia  $j$  e per una determinata condizione  $c$ , la deviazione standard è calcolata come:

$$SD_{jc} = \sqrt{\frac{1}{n_c - 1} \sum_{i=1}^{n_c} (x_{ijc} - \bar{x}_{jc})^2} \quad (3.1)$$

dove:

- $x_{ijc}$  rappresenta la risposta del partecipante  $i$  alla perizia  $j$  nella condizione  $c$ ;
- $\bar{x}_{jc}$  è la media delle risposte relative alla perizia  $j$  nella condizione  $c$ ;
- $n_c$  indica il numero di partecipanti assegnati alla condizione  $c$ .

Il confronto tra le due condizioni consente di valutare se e in che misura l'introduzione del suggerimento algoritmico contribuisca a ridurre la variabilità delle valutazioni, influenzando il processo di selezione e interpretazione delle informazioni rilevanti. Valori più elevati della deviazione standard indicano una maggiore eterogeneità nei giudizi e dunque un più alto livello di discrezionalità individuale; al contrario, una riduzione sistematica della dispersione nelle condizioni con AI suggerirebbe un effetto di convergenza valutativa potenzialmente indotto dal supporto algoritmico. Il confronto è stato effettuato sia a livello descrittivo sia mediante test statistici sulle differenze tra condizioni.

### 3.6.2 Accuratezza delle valutazioni rispetto al *ground truth*

L'accuratezza delle decisioni viene valutata confrontando le risposte fornite dai partecipanti con un riferimento ritenuto valido (*ground truth*), definito sulla base delle valutazioni effettuate da periti esperti nell'ambito del progetto DARE. Tale

riferimento non rappresenta una verità oggettiva in senso assoluto, ma costituisce il miglior proxy disponibile di una valutazione corretta e condivisa all'interno del dominio applicativo.

Per la raccolta delle risposte, è stata adottata sia una scala Likert a cinque punti sia una scala binaria. La scelta di una scala ordinale è coerente con quanto proposto da Mishra et al. (2024) [12], che evidenziano come una maggiore granularità della scala consenta ai partecipanti di esprimere in modo più accurato l'incertezza e riduca il rischio di risposte casuali o forzate. L'unica domanda con risposte su scala binaria è relativa all'indennizzabilità della perizia, fattore non sottoponibile al giudizio del singolo, in quanto scelta effettuata sulla base di criteri ben specifici (descritti nel paragrafo 3.3.1).

Ai fini dell'analisi dell'accuratezza, le risposte Likert vengono successivamente convertite in formato binario secondo un criterio semantico: i valori 1–2 sono associati alla classe negativa, i valori 4–5 alla classe positiva, mentre il valore centrale (3), che rappresenta una posizione di incertezza o neutralità, viene escluso dall'analisi. Tale scelta segue l'impostazione adottata in letteratura da Suarez et al. (2024) [19], e non deriva da un'ottimizzazione data-driven del cut-off, preservando il significato originario della scala [12].

L'accuratezza viene quindi valutata mediante metriche derivate dalla **confusion matrix**, consentendo di confrontare le prestazioni degli utenti inesperti nelle condizioni con e senza supporto AI. A partire dalla matrice di confusione è stata inoltre costruita una tabella di contingenza tra correttezza della risposta (corretta vs errata) e tipologia di domanda. Per verificare se le differenze osservate tra le categorie considerate fossero statisticamente significative è stato applicato il **test del chi-quadro di Pearson**. Tale test consente di valutare se la distribuzione delle risposte corrette ed errate differisca in modo significativo tra le diverse tipologie di domanda, verificando l'ipotesi nulla di indipendenza tra le due variabili.

In particolare, per ciascuna condizione sperimentale sono stati identificati i seguenti elementi: *True Positive* (TP), *True Negative* (TN), *False Positive* (FP) e

*False Negative* (FN). L'accuratezza complessiva (*Accuracy*) è stata calcolata come:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.2)$$

dove:

- *TP* rappresenta il numero di risposte correttamente classificate come positive;
- *TN* rappresenta il numero di risposte correttamente classificate come negative;
- *FP* indica le risposte erroneamente classificate come positive;
- *FN* indica le risposte erroneamente classificate come negative.

Tale formalizzazione consente di valutare in modo oggettivo la capacità dei partecipanti di fornire risposte corrette e di confrontare le performance tra le diverse condizioni sperimentali.

### 3.6.3 Affidamento al sistema di supporto AI

L'affidamento al sistema di supporto basato su intelligenza artificiale viene analizzato valutando il grado di allineamento tra le risposte fornite dai partecipanti e il suggerimento restituito dal tool nelle condizioni con AI. In linea con la letteratura sull'*automation bias* (Jussupow et al., 2021 [16]), un'elevata coincidenza tra decisione umana e output algoritmico può riflettere sia un utilizzo efficace del supporto decisionale sia forme di accettazione acritica del suggerimento.

Nel presente studio, l'affidamento viene misurato attraverso il *follow rate*. Il *follow rate* misura la frequenza con cui il partecipante fornisce una risposta coerente con il suggerimento dell'AI. A tal fine, le risposte originariamente raccolte su scala Likert vengono ricodificate in forma binaria, escludendo il valore centrale (3), che non esprime una scelta direzionale. Il *follow rate* è definito come:

$$FollowRate = \frac{N_{risposte\ uguali\ all'AI}}{N_{perizie\ valutate}}$$

Poiché il sistema di supporto fornisce suggerimenti che possono essere corretti o intenzionalmente errati, il *follow rate* viene ulteriormente distinto in:

$$FollowRate_{correct} = \frac{N_{risposte\ uguali\ all'AI\ corretto}}{N_{perizie\ con\ AI\ corretto}}$$

$$FollowRate_{wrong} = \frac{N_{risposte\ uguali\ all'AI\ errato}}{N_{perizie\ con\ AI\ errato}}$$

La differenza tra i due indicatori viene sintetizzata tramite:

$$\Delta = FollowRate_{correct} - FollowRate_{wrong}$$

Un valore elevato di  $\Delta$  indica un affidamento selettivo e ragionato al sistema di supporto, mentre valori prossimi allo zero suggeriscono una tendenza all'adesione indiscriminata al suggerimento dell'AI, riconducibile a fenomeni di affidamento cieco.

Questo approccio è coerente con il framework proposto da Ben-Michael et al. (2025) [21], che utilizza la distinzione tra *follow* e *override* per discriminare tra affidamento razionale e *automation bias* nelle decisioni umano-AI.

### 3.6.4 Tempo di esecuzione del compito

Un'ulteriore dimensione analizzata riguarda il tempo di esecuzione del compito, misurato come il tempo impiegato dai partecipanti per completare la valutazione delle perizie nelle condizioni con e senza supporto AI. L'analisi dei tempi consente di valutare se l'introduzione del tool influisca sull'efficienza del processo decisionale, oltre che sulla sua qualità.

Poiché le condizioni sperimentali prevedevano un diverso numero di perizie da esaminare (4 nella condizione senza supporto AI e 8 nella condizione con supporto AI), il tempo totale non risultava direttamente confrontabile. Al fine di garantire la comparabilità tra le condizioni, il tempo è stato pertanto normalizzato calcolando il *tempo medio per perizia* per ciascun partecipante:

$$T_{\text{senza AI},i}^* = \frac{T_{\text{senza AI},i}}{4} \quad T_{\text{con AI},i}^* = \frac{T_{\text{con AI},i}}{8}$$

Il tempo così normalizzato rappresenta una misura dell'efficienza decisionale individuale, espressa in secondi medi necessari per la valutazione di una singola perizia.

Per ciascun partecipante è quindi calcolata la differenza tra le due condizioni:

$$\Delta_i = T_{\text{senza AI},i}^* - T_{\text{con AI},i}^*$$

Valori positivi di  $\Delta_i$  indicano una riduzione del tempo medio per perizia in presenza del supporto AI (maggiore efficienza), mentre valori negativi indicano un aumento del tempo medio (maggiore carico cognitivo o necessità di verifica dei suggerimenti).

In linea con le raccomandazioni metodologiche per l'analisi di dati appaiati di Tokunaga (2018) [22], la distribuzione dei valori  $\Delta_i$  viene preliminarmente verificata mediante analisi grafica e test di normalità (Shapiro–Wilk). In funzione dell'esito di tale verifica, il confronto statistico viene effettuato mediante **paired t-test** oppure, in caso di violazione dell'ipotesi di normalità, tramite test non parametrico di Wilcoxon signed-rank.

Sono stati esclusi dall'analisi i tentativi incompleti e i tempi anomali riconducibili a interruzioni o inattività prolungata, identificati mediante criteri statistici oggettivi.

## Capitolo 4

# Esperimento

Il presente capitolo descrive lo svolgimento dell'esperimento condotto per valutare l'impatto di un sistema di supporto basato su intelligenza artificiale sul processo decisionale peritale di utenti inesperti. L'esperimento è stato progettato in coerenza con l'approccio metodologico illustrato nel Capitolo 3 e ha l'obiettivo di osservare, in un contesto controllato ma realistico, le modalità con cui l'introduzione del supporto AI influisce sulla discrezionalità decisionale, sull'accuratezza delle valutazioni, sull'affidamento al sistema e sui tempi di esecuzione del compito.

Lo studio si articola in una sequenza strutturata di fasi che comprendono una preparazione preliminare dei partecipanti, la somministrazione delle perizie secondo il disegno sperimentale definito e la raccolta sistematica delle risposte. L'esperimento prevede l'esposizione dei partecipanti a più condizioni sperimentali, con e senza supporto AI, al fine di consentire un confronto diretto tra le diverse modalità di decisione.

Nei paragrafi che seguono vengono descritti nel dettaglio il contesto operativo dell'esperimento, la procedura di somministrazione delle perizie, le modalità di raccolta dei dati e gli strumenti utilizzati per garantire la coerenza e la riproducibilità dello studio.

## 4.1 Obiettivo dell'esperimento

L'obiettivo dell'elaborato si colloca prevalentemente nella fase di progettazione e validazione del framework sperimentale volto a valutare l'impatto dell'introduzione di un sistema di supporto basato su intelligenza artificiale sul processo decisionale peritale di utenti inesperti. La presente tesi non si concentra su una sperimentazione estesa su larga scala, bensì sulla strutturazione metodologica dell'esperimento, sulla definizione delle variabili, delle misure e delle condizioni sperimentali, nonché sulla loro coerenza teorica e operativa.

In tale prospettiva, il lavoro culmina in un primo test pilota condotto su un campione ristretto, finalizzato a verificare la solidità dell'impianto sperimentale, la chiarezza delle istruzioni, la funzionalità tecnica degli strumenti utilizzati e l'adeguatezza delle metriche di misurazione. Questa fase preliminare ha lo scopo di raccogliere feedback sia interni (relativi alla struttura metodologica e all'operatività del disegno sperimentale) sia esterni (provenienti dai partecipanti), così da individuare eventuali criticità e margini di miglioramento.

L'obiettivo finale del lavoro è dunque la validazione della progettazione dell'esperimento, ponendo le basi per una successiva fase di implementazione su un campione più ampio, che sarà oggetto di sviluppi futuri e di ulteriori attività di ricerca. In questo senso, la tesi si configura come un contributo metodologico volto a garantire la robustezza e la replicabilità del disegno sperimentale prima della sua applicazione su larga scala.

L'esperimento si concentra su quattro dimensioni chiave del processo decisionale: la discrezionalità delle scelte, l'accuratezza delle valutazioni rispetto a un riferimento ritenuto valido (*ground truth*), il grado di affidamento degli utenti al suggerimento fornito dal sistema e il tempo di esecuzione del compito. Attraverso il confronto tra condizioni sperimentali con e senza supporto AI, l'obiettivo è comprendere se e in che misura il sistema contribuisca a rendere il processo decisionale più coerente, accurato ed efficiente, nonché a identificare eventuali effetti di condizionamento o di accettazione acritica del suggerimento algoritmico.

Un ulteriore obiettivo dello studio consiste nel verificare se il supporto AI consenta a utenti privi di esperienza nel dominio peritale di avvicinare le proprie prestazioni a valutazioni considerate corrette e consolidate, riducendo al contempo la variabilità soggettiva tipica dei contesti decisionali complessi. In questo senso, l'esperimento non intende valutare le prestazioni del sistema di intelligenza artificiale, ma osservare le dinamiche di interazione tra l'utente umano e il supporto algoritmico, ponendo attenzione al ruolo della discrezionalità e del giudizio critico nel processo decisionale.

Nel complesso, il lavoro di tesi è orientato alla progettazione, strutturazione e prima verifica del disegno sperimentale, con l'obiettivo di costruire un impianto metodologico solido e replicabile per lo studio del ruolo dell'intelligenza artificiale come strumento di supporto nei processi decisionali peritali. Attraverso il test pilota su campione ristretto, l'elaborato fornisce una prima raccolta di evidenze quantitative e qualitative funzionali a valutare la coerenza interna del framework, la chiarezza delle procedure e l'adeguatezza delle misure adottate.

Più che produrre risultati definitivi sull'impatto dell'AI, questa fase mira dunque a validare la struttura dell'esperimento, così da porre le basi per successive implementazioni su scala più ampia. In tal senso, il contributo principale della tesi consiste nel colmare il divario metodologico tra studi condotti su task semplificati e la costruzione di un protocollo sperimentale applicabile a contesti realistici e ad elevata complessità informativa.

## **4.2 Procedura sperimentale**

La procedura sperimentale è stata strutturata in modo da garantire uniformità delle condizioni di svolgimento, controllo delle variabili rilevanti e coerenza con il disegno metodologico descritto nel capitolo precedente. L'esperimento si articola in una fase di formazione preliminare e in una successiva fase di somministrazione del test sperimentale.

In una prima fase, tutti i partecipanti sono stati invitati a svolgere una fase

di formazione preliminare, articolata in una componente asincrona e in un breve momento sincrono online della durata di circa 30 minuti, finalizzata a garantire una comprensione uniforme del contesto sperimentale e delle attività richieste. La componente asincrona è stata erogata attraverso un insieme di materiali di supporto messi a disposizione prima dell'avvio del test, comprendenti un manuale operativo e una serie di video pillole introduttive.

I materiali formativi illustrano il contesto generale dell'esperimento, il ruolo richiesto ai partecipanti, le attività da svolgere e la durata complessiva della prova. In tale fase vengono inoltre presentati i concetti fondamentali necessari per affrontare il test, con particolare riferimento alla nozione di perizia, all'identificazione delle cause e degli effetti del danno, alla lettura delle immagini peritali e alla valutazione della stima economica.

La scelta di adottare una modalità di formazione asincrona consente di standardizzare le informazioni fornite a tutti i partecipanti, riducendo potenziali differenze legate alla fruizione dei contenuti e garantendo una maggiore coerenza metodologica del disegno sperimentale.

A tutti i partecipanti sono stati forniti un manuale sintetico e una serie di video-pillole tematiche. Il manuale, redatto dall'autrice, ha una lunghezza contenuta ed è strutturato in sei capitoli: (i) introduzione all'esperimento e ruolo dei partecipanti; (ii) definizione e struttura generale di una perizia; (iii) cause del danno e criteri per il loro riconoscimento; (iv) tipologie di danno ed effetti osservabili; (v) interpretazione del suggerimento AI; (vi) lettura e utilizzo delle tabelle di stima economica; (vii) criteri generali per la valutazione dell'indennizzabilità della perizia. I capitoli dedicati alle cause e ai danni includono immagini esplicative e indicazioni operative basate sull'esperienza maturata durante l'attività di ricerca. Le video-pillole, anch'esse realizzate dall'autrice, hanno una durata media di circa due minuti ciascuna e riprendono la struttura e i contenuti del manuale, utilizzando le stesse immagini e slide di supporto. Dopo la consegna di tali materiali, i partecipanti sono stati lasciati liberi di gestire autonomamente la lettura del manuale e la visione

delle video-pillole, potendo consultare il materiale anche durante lo svolgimento del test.

Successivamente, prima dell'avvio della prova sperimentale, è stato organizzato un breve incontro sincrono online della durata di circa 30 minuti. Durante tale sessione sono stati richiamati sinteticamente i contenuti principali della formazione asincrona, è stata illustrata la struttura operativa del test su piattaforma e sono state fornite indicazioni pratiche sulle modalità di compilazione delle risposte.

Al termine della spiegazione è stato previsto uno spazio dedicato a eventuali domande di chiarimento, limitate esclusivamente ad aspetti organizzativi e procedurali. Non sono state fornite indicazioni relative alla correttezza delle risposte né anticipazioni sui contenuti delle perizie oggetto del test, al fine di preservare la neutralità informativa e la validità interna dell'esperimento.

La fase sperimentale vera e propria è stata successivamente avviata tramite la somministrazione del test su piattaforma Moodle. Al fine di garantire che ciascuna perizia fosse visualizzata in tutte le condizioni sperimentali senza introdurre effetti di apprendimento, sono stati predisposti tre test distinti, differenziati unicamente per l'assegnazione delle perizie alle condizioni sperimentali. I partecipanti sono stati suddivisi in tre gruppi, ciascuno dei quali ha ricevuto una sola versione del test. A ogni partecipante è stato consentito un unico tentativo di svolgimento. Il test è stato reso disponibile in modalità asincrona e i partecipanti hanno avuto una settimana di tempo per completarlo, potendo svolgere la prova da casa e consultare liberamente il manuale durante l'esecuzione. L'intera procedura sperimentale è sintetizzata nel diagramma di flusso riportato in Figura 4.1, che riassume le principali fasi dell'esperimento e la loro sequenza temporale.

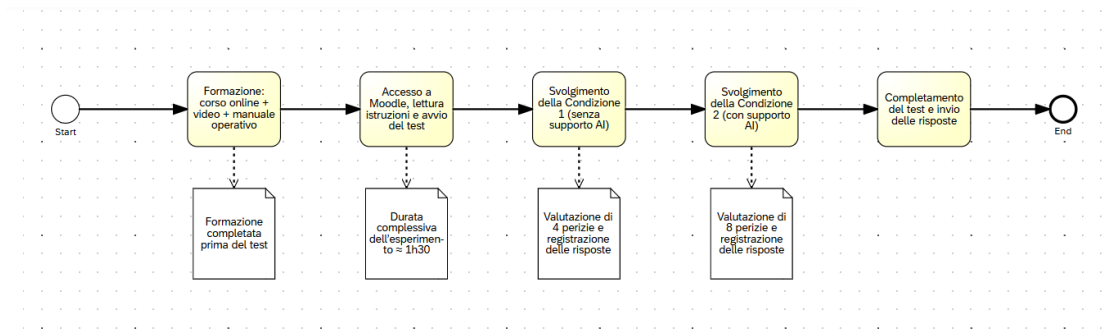


Figura 4.1: Diagramma di flusso della procedura sperimentale

Il diagramma consente di visualizzare in modo immediato la struttura complessiva dell'esperimento e il ruolo dei diversi materiali e condizioni sperimentali, facilitando la comprensione del disegno adottato.

Durante il test, le perizie sono state presentate una alla volta esclusivamente attraverso il materiale visivo. Per ciascuna perizia, ai partecipanti venivano mostrate unicamente le immagini peritali, senza alcun accesso alla documentazione testuale originale. La valutazione non si basava quindi sulla lettura di una perizia scritta, ma sull'osservazione delle immagini e sulle affermazioni proposte. In seguito alla visualizzazione delle immagini, venivano presentate le domande di valutazione associate alla perizia; durante la compilazione delle risposte, le immagini rimanevano sempre visibili. Nelle condizioni con supporto AI, alle immagini era inoltre affiancato il suggerimento prodotto dal sistema, sottoforma testuale e indicazioni relative alla stima economica; nella condizione senza AI tali informazioni non erano presenti.

Per ciascuna perizia, i partecipanti hanno risposto a tre affermazioni valutate su scala Likert a cinque punti, con estremi da “totalmente in disaccordo” (1) a “totalmente d'accordo” (5), e una domanda a risposta binaria. Le affermazioni riguardavano: (i) l'identificazione della causa del danno, (ii) l'identificazione del danno osservabile, (iii) la valutazione dell'indennizzabilità della perizia (risposta sì/no) e (iv) la correttezza della stima economica proposta. La scala Likert è stata adottata per consentire ai partecipanti di esprimere un giudizio graduato, evitando forzature tipiche delle risposte puramente binarie e permettendo un'analisi più fine

delle dinamiche decisionali, in linea con approcci analoghi presenti in letteratura.

Prima dell'inizio del test, a tutti i partecipanti è stato richiesto di indicare la propria media accademica. Al termine della prova è stata inoltre raccolta una misura della percezione soggettiva dell'esperienza di utilizzo attraverso quattro domande relative alla difficoltà percepita del compito, al livello di sicurezza nelle proprie risposte, alla fiducia nelle valutazioni effettuate e alla chiarezza complessiva dell'attività sperimentale. Le risposte sono state raccolte mediante scala Likert, consentendo ai partecipanti di esprimere il proprio grado di accordo rispetto alle affermazioni proposte. L'inclusione di queste domande ha lo scopo di integrare le misure oggettive di performance con una valutazione percettiva dell'esperienza dell'utente, permettendo di cogliere eventuali differenze tra prestazioni osservate e percezione soggettiva del compito. Il tempo di svolgimento è stato registrato automaticamente dalla piattaforma Moodle, consentendo di misurare in modo oggettivo la durata complessiva della prova per ciascun partecipante.

### 4.3 Svolgimento delle sessioni sperimentali

Lo svolgimento delle sessioni sperimentali avviene in **modalità sincrona tramite piattaforma Moodle**. A ciascun partecipante viene assegnato un test sperimentale specifico, coerente con lo schema di controbilanciamento definito in fase di progettazione, e reso disponibile per un unico tentativo di svolgimento entro una finestra temporale predefinita. Il tempo di esecuzione viene registrato automaticamente dalla piattaforma, consentendo una misura oggettiva della durata complessiva della sessione.

Una volta avviata la sessione, il partecipante visualizza le perizie una alla volta. Per ciascuna perizia vengono mostrate esclusivamente le immagini peritali associate; il testo descrittivo della perizia non è mai reso disponibile. Questa scelta metodologica è finalizzata a rendere ancora più realistico il contesto sperimentale, in quanto i periti svolgono prima la fase di visione dell'evento e successivamente

redigono il testo peritale. Per tali ragioni, si è preferito concentrare il processo decisionale sull'interpretazione delle evidenze visive e sull'uso delle informazioni fornite durante la fase di formazione.

Per ogni perizia, dopo la visualizzazione delle immagini, al partecipante viene richiesto di rispondere a quattro **item valutativi**. Tre item sono formulati come affermazioni e valutati su scala Likert a cinque punti (da “totalmente in disaccordo” a “totalmente d'accordo”) e riguardano, rispettivamente, **l'identificazione della causa del danno, l'identificazione del danno e la stima economica associata**. Il quarto item riguarda **l'indennizzabilità della perizia** ed è espresso in forma binaria (sì/no). Durante la compilazione delle risposte, il partecipante può in qualsiasi momento rivedere le immagini della perizia e consultare il manuale operativo fornito nella fase preliminare.

Nelle condizioni sperimentali che prevedono il supporto basato su intelligenza artificiale, insieme alle immagini viene mostrato anche il suggerimento prodotto dal sistema, sotto forma testuale e indicazioni relative alla stima economica. Nella condizione senza supporto AI tali informazioni non sono presenti.

Al termine della valutazione di tutte le perizie assegnate, al partecipante viene richiesto di fornire una valutazione soggettiva della complessità complessiva dell'esperimento. Una volta completata questa fase, il test viene automaticamente chiuso e le risposte vengono salvate per le successive analisi statistiche.

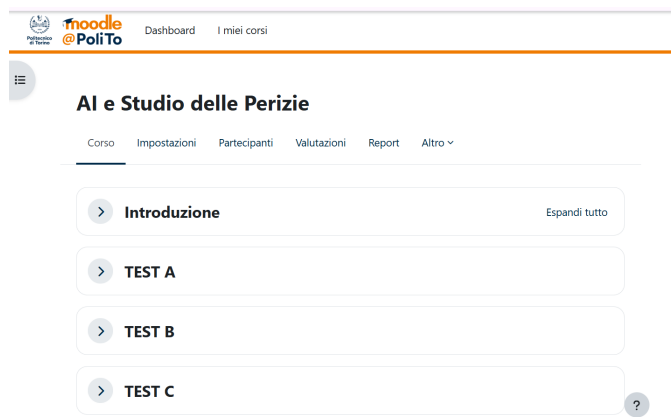
### 4.3.1 Esempio di struttura del test sperimentale

Al fine di rendere più chiara la modalità di interazione del partecipante con il test sperimentale, in Figura 4.2 e 4.3. Esse riportano un esempio della schermata proposta ai partecipanti rispettivamente nella condizione senza supporto AI (a) e nella condizione con supporto AI (b). È riportato un esempio di schermata così come visualizzata durante lo svolgimento dell'esperimento.

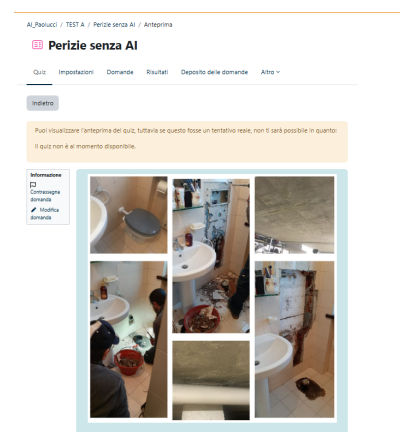
Per ciascuna perizia, il partecipante visualizza esclusivamente le immagini peritali associate al caso, eventualmente corredate dal suggerimento fornito dal sistema

di supporto basato su intelligenza artificiale, quando previsto dalla condizione sperimentale. Il testo descrittivo della perizia non è mai mostrato. Al di sotto delle immagini sono presentati gli item valutativi relativi alla perizia, formulati come affermazioni e valutati su scala Likert a cinque punti, ad eccezione dell'item relativo all'indennizzabilità, che richiede una risposta binaria.

L'interfaccia consente al partecipante di consultare le immagini in qualsiasi momento durante la compilazione delle risposte, garantendo coerenza tra le informazioni visive osservate e il processo decisionale. Questo esempio illustra la struttura standard adottata per tutte le perizie del test, che differiscono esclusivamente per il contenuto delle immagini, delle affermazioni proposte e dell'eventuale suggerimento AI.



(a)



(b)

Figura 4.2: Esempio schermate test

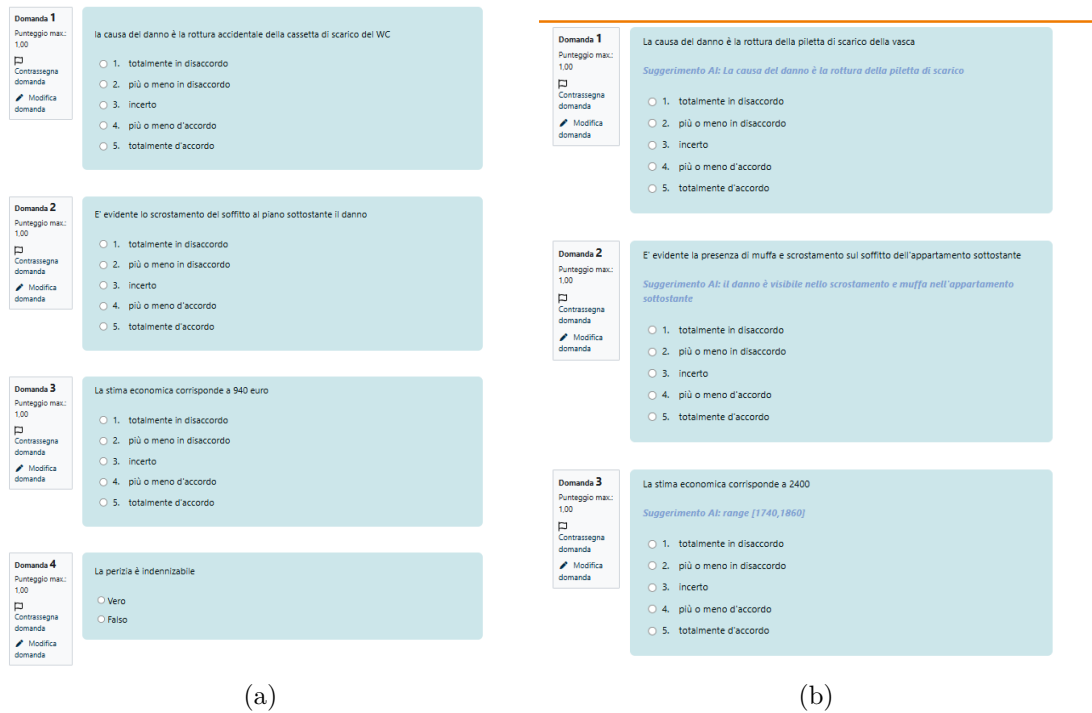


Figura 4.3: Esempio schermate test

## 4.4 Raccolta dei dati

La raccolta dei dati avviene attraverso la piattaforma Moodle, utilizzata come ambiente di somministrazione dei test sperimentali. Per ciascun partecipante viene registrato un unico tentativo di svolgimento del test assegnato, in modo da evitare ripetizioni e garantire l'univocità delle osservazioni raccolte.

I dati esportati da Moodle includono, per ogni partecipante, le risposte fornite a ciascuna domanda associata alle singole perizie, nonché le informazioni temporali relative all'inizio e alla fine del test. In particolare, per ogni tentativo sono disponibili:

- un identificativo anonimo del partecipante, generato automaticamente dalla piattaforma;

- le risposte alle domande relative a ciascuna perizia, codificate secondo la struttura del questionario;
- il tempo complessivo di svolgimento del test, misurato automaticamente da Moodle come differenza tra tempo di inizio e tempo di completamento.

Le perizie sono identificate in modo univoco attraverso un codice alfanumerico incorporato nella denominazione delle domande (ad esempio P03, P07, ecc.), che consente di associare ciascuna risposta alla specifica perizia valutata. Analogamente, la struttura del test consente di ricondurre ogni perizia alla relativa condizione sperimentale (assenza di supporto AI, supporto AI con suggerimento corretto, supporto AI con suggerimento errato), in funzione del test assegnato al partecipante.

L'output fornito da Moodle è esportato in formato `.csv` e presenta una struttura in formato *wide*, nella quale ciascuna riga corrisponde a un partecipante e ciascuna colonna rappresenta una specifica domanda del questionario. Tale formato riflette l'organizzazione nativa dei dati nella piattaforma di e-learning ed è idoneo alla successiva ristrutturazione per fini analitici.

Prima dell'analisi statistica, i dati vengono riorganizzati in modo da ottenere un dataset in formato *long*, nel quale ogni osservazione corrisponde a una singola valutazione perizia-partecipante. Questa ristrutturazione consente di associare in modo esplicito a ciascuna risposta le seguenti informazioni:

- identificativo del partecipante;
- identificativo della perizia;
- test;
- condizione sperimentale;
- risposta fornita alle singole domande;
- tempo di esecuzione.

Tale struttura dei dati risulta coerente con il disegno sperimentale adottato e consente di effettuare analisi comparative tra condizioni, nonché di calcolare le metriche di interesse (accuratezza, discrezionalità decisionale, affidamento al sistema e tempi di esecuzione), come descritto nel paragrafo successivo. L'analisi dei dati viene condotta utilizzando il software statistico *Stata*, che consente una gestione flessibile dei dati ristrutturati e l'applicazione dei test statistici appropriati in funzione delle variabili considerate.

# Capitolo 5

## Analisi dei risultati

### 5.1 Descrizione del campione e dei dati raccolti

L'analisi empirica è stata condotta su un campione di 17 partecipanti inesperti coinvolti in un esperimento volto a valutare l'impatto del supporto basato su intelligenza artificiale nel processo decisionale relativo all'analisi di perizie assicurative. I partecipanti sono stati assegnati a uno dei tre test sperimentali previsti dal disegno di ricerca, ciascuno caratterizzato dalle stesse domande con differenti modalità di supporto decisionale.

Complessivamente, il dataset finale include le decisioni espresse dai partecipanti su un insieme di dodici perizie, per un totale di diverse centinaia di osservazioni a livello partecipante–perizia–domanda. Per ciascuna osservazione sono state raccolte informazioni relative alla valutazione su scala Likert, alla decisione binaria di indennizzabilità, al tempo di esecuzione del compito e al comportamento di adesione al suggerimento algoritmico nelle condizioni con supporto AI.

I dati esportati dalla piattaforma sperimentale sono stati sottoposti a un processo di pulizia e riorganizzazione volto a garantire la comparabilità tra le condizioni. In particolare, sono state escluse le risposte neutrali (di incertezza) nelle analisi di accuratezza, sono stati rimossi i tempi di completamento non plausibili riconducibili

a inattività prolungata e sono state costruite variabili derivate funzionali alle analisi successive, quali la correttezza della decisione rispetto al *ground truth*, la discrezionalità tra valutatori e il tasso di follow del suggerimento AI.

Il dataset così ottenuto costituisce la base delle analisi presentate nelle sezioni successive, finalizzate a esaminare l'impatto del supporto algoritmico su discrezionalità decisionale, accuratezza, affidamento degli utenti ed efficienza del processo decisionale.

## 5.2 Confronto tra le due condizioni

Il disegno sperimentale prevede il confronto tra decisioni assunte in assenza di supporto algoritmico e decisioni formulate in presenza di suggerimenti generati da un sistema di intelligenza artificiale. L'obiettivo principale del confronto è valutare se l'introduzione dell'AI produca modifiche osservabili nel comportamento decisionale dei partecipanti.

Le due condizioni differiscono esclusivamente per la presenza del suggerimento algoritmico, mentre il contenuto informativo delle perizie e la struttura del compito rimangono invariati. Tale impostazione consente di isolare l'effetto del supporto AI sul processo decisionale, riducendo l'influenza di fattori confondenti legati al contesto informativo o alla complessità del compito.

Il confronto tra condizioni viene sviluppato lungo quattro dimensioni principali: (i) discrezionalità tra valutatori, (ii) accuratezza delle decisioni rispetto al *ground truth*, (iii) grado di affidamento degli utenti al suggerimento algoritmico e (iv) tempi di esecuzione del compito. L'analisi congiunta di tali dimensioni permette di valutare non soltanto se l'AI migliori la performance decisionale, ma anche in che modo essa influenzi il comportamento e l'esperienza decisionale degli utenti.

## 5.3 Impatto del supporto AI sulla discrezionalità decisionale

L'obiettivo di questa analisi è verificare se la presenza del supporto AI riduca la discrezionalità decisionale tra i valutatori. Per discrezionalità si intende la **variabilità tra i giudizi espressi dai partecipanti sulla stessa perizia**. Maggiore è la dispersione delle valutazioni, maggiore è la discrezionalità; al contrario, una riduzione della dispersione indica maggiore convergenza tra decisori.

### 5.3.1 Costruzione della misura

Per ciascuna perizia e per ciascuna condizione sperimentale (assenza vs presenza di AI) è stata calcolata la deviazione standard delle valutazioni medie espresse dai partecipanti.

In particolare:

- Per ogni partecipante è stata calcolata la media delle risposte su scala Likert relative alla medesima perizia.
- Per ogni perizia  $i$  e per ciascuna condizione  $c$  è stata quindi calcolata la deviazione standard:

$$SD_{ic} = \sqrt{\frac{1}{n-1} \sum_{j=1}^n (x_{ijc} - \bar{x}_{ic})^2}$$

dove  $x_{ijc}$  rappresenta la valutazione media del partecipante  $j$  sulla perizia  $i$  nella condizione  $c$ .

La discrezionalità media per condizione è stata quindi ottenuta calcolando la media delle deviazioni standard sulle perizie.

Per valutare **l'effetto del supporto AI** è stato calcolato, per ciascuna perizia, il differenziale:

$$\Delta_i = SD_{i,0} - SD_{i,1}$$

dove:

- $SD_{i,0}$  = deviazione standard senza AI
- $SD_{i,1}$  = deviazione standard con AI

Un valore positivo di  $\Delta_i$  indica una riduzione della discrezionalità in presenza dell'AI.

### 5.3.2 Risultati

L'analisi è stata condotta sulle 12 perizie visualizzate dai partecipanti durante l'esperimento. Le statistiche descrittive delle deviazioni standard sono riportate nella Tabella 5.1.

Tabella 5.1: Confronto della discrezionalità (SD) tra condizioni

	Senza AI	Con AI	Differenza media ( $\Delta$ )
Media SD	0.659	0.671	-0.011
Deviazione standard	0.242	0.166	0.256
Numero perizie			12

Il paired t-test sulle deviazioni standard perizia-perizia fornisce:

$$t(11) = -0.15, \quad p = 0.882$$

L'intervallo di confidenza al 95% della differenza media è:

$$[-0.174; 0.152]$$

### 5.3.3 Interpretazione

I risultati indicano che la discrezionalità media, misurata attraverso la deviazione standard delle valutazioni Likert per ciascuna perizia, risulta sostanzialmente

invariata tra le due condizioni sperimentali. In assenza di AI la deviazione standard media è pari a 0.659, mentre in presenza del supporto AI è pari a 0.671. La **differenza** osservata ( $\Delta = -0.011$ ) è prossima allo zero e **statisticamente non significativa** ( $t(11) = -0.15, p = 0.882$ ).

Dal punto di vista sostanziale, ciò implica che **l'introduzione del supporto algoritmico non ha prodotto una maggiore convergenza nei giudizi espressi dai partecipanti**. Se l'AI avesse ridotto la discrezionalità, ci si sarebbe attesi una diminuzione sistematica della deviazione standard nella condizione con supporto. Al contrario, la lieve variazione osservata risulta non solo trascurabile in termini di entità, ma anche priva di direzionalità coerente.

In altri termini, la variabilità tra valutatori sembra rappresentare una componente strutturale del processo decisionale. L'esposizione al suggerimento dell'AI non appare sufficiente a comprimere tale dispersione, suggerendo che i partecipanti continuano a interpretare le informazioni disponibili in modo autonomo, anche in presenza di un'indicazione algoritmica.

La Figura 5.1 fornisce una rappresentazione sintetica della discrezionalità media nelle due condizioni. Le barre risultano quasi sovrapponibili, evidenziando visivamente l'assenza di una riduzione apprezzabile della dispersione in presenza dell'AI. Anche considerando la variabilità tra perizie (barre di errore), non emerge una separazione sistematica tra le condizioni (condizione 0= assenza di AI; Condizione 1= presenza del supporto AI).

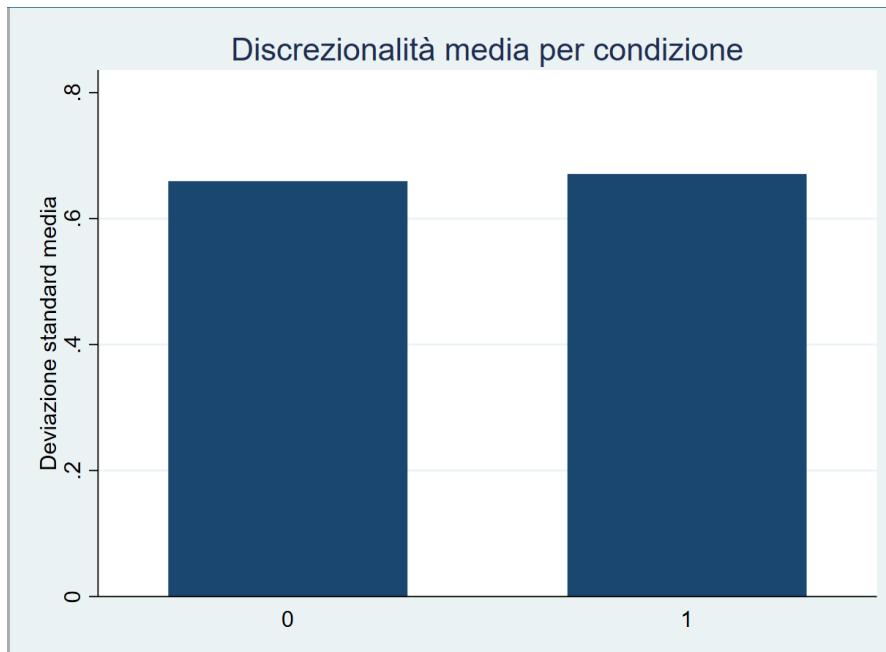
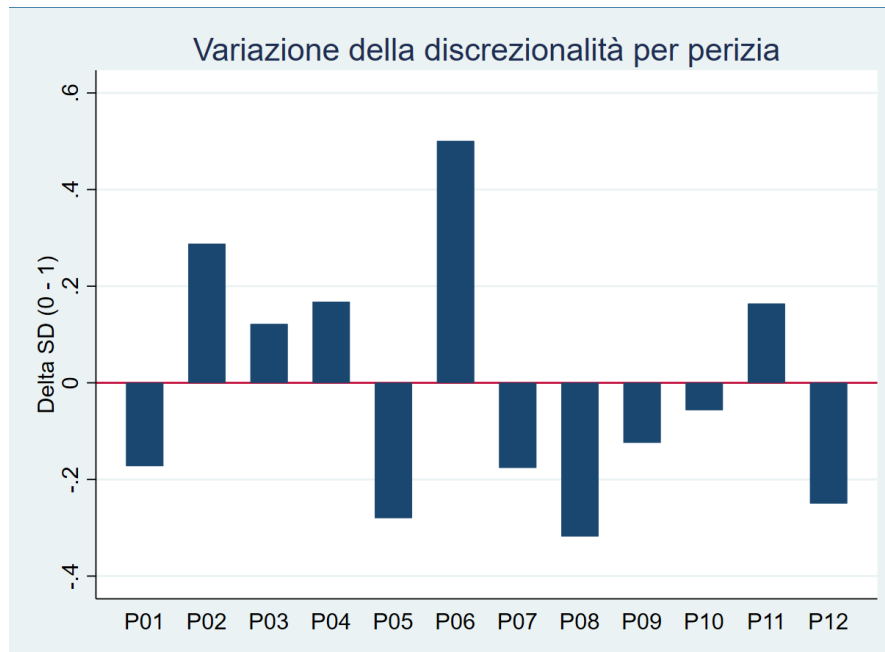


Figura 5.1: Discrezionalità media per condizione

La Figura 5.2 riporta invece il differenziale  $\Delta_i$  per ciascuna perizia. La distribuzione dei valori attorno allo zero mostra chiaramente che l'effetto non è uniforme: alcune perizie registrano una lieve riduzione della discrezionalità con l'AI, mentre altre evidenziano un incremento. L'assenza di un pattern coerente conferma che l'intervento algoritmico non produce un impatto stabile sulla convergenza decisionale.

Figura 5.2: Differenza di discrezionalità per perizia ( $\Delta_i$ )

È importante sottolineare che l'unità di analisi è rappresentata dalle perizie ( $n = 12$ ). Tale numerosità relativamente contenuta limita la potenza statistica del test e potrebbe attenuare la capacità di intercettare effetti di entità moderata. Tuttavia, la coerenza tra risultati numerici e rappresentazione grafica rafforza l'interpretazione complessiva.

Nel complesso, l'evidenza empirica non supporta l'ipotesi che il supporto AI riduca la discrezionalità decisionale nelle valutazioni su scala Likert. I risultati suggeriscono piuttosto che la dispersione tra decisori costituisca una caratteristica intrinseca del processo valutativo, non facilmente modificabile mediante l'introduzione di suggerimenti automatizzati. I risultati ottenuti risultano coerenti con le evidenze riportate da Kolkman et al. (2024) [11], che mostrano come l'introduzione di un sistema di supporto algoritmico possa influenzare significativamente il processo decisionale dei valutatori umani. Analogamente a quanto osservato nel presente esperimento, anche nel loro studio l'utilizzo dell'AI modifica le decisioni dei paralegali. Inoltre, la limitata concordanza tra valutatori osservata nei paired cases (7

decisioni concordi su 18) evidenzia una rilevante discrezionalità nel giudizio umano, fenomeno che trova riscontro anche nella dispersione delle valutazioni osservata tra i partecipanti del presente studio. Nel loro esperimento, l'introduzione di un sistema di AI porta a una riduzione dell'accuratezza delle decisioni dei paralegali (dal 67% al 48%). I risultati del presente studio suggeriscono che un possibile meccanismo alla base di questo fenomeno sia la tendenza degli utenti a seguire il suggerimento algoritmico anche quando questo è errato. In presenza di sistemi caratterizzati da livelli di accuratezza non elevati, tale dinamica può quindi portare a un peggioramento complessivo della qualità delle decisioni umane.

### 5.3.4 Discrezionalità nelle risposte binarie

L'analisi della discrezionalità è stata replicata considerando la variabile decisionale binaria (`binary_answer`), che assume valore 1 in caso di risposta corretta e 0 in caso contrario. Tale variabile è stata doperata per le domande relative all'indennizzabilità della perizia, in quanto non è stato ritenuto possibile rispondere a tale domanda sulla base di una scala ordinale.

Anche in questo caso, la discrezionalità è stata operazionalizzata come variabilità tra partecipanti sulla medesima perizia. Per ciascuna perizia  $i$  e per ciascuna condizione sperimentale  $c$  è stata quindi calcolata la deviazione standard delle risposte binarie:

$$SD_{ic} = \sqrt{\frac{1}{n-1} \sum_{j=1}^n (x_{ijc} - \bar{x}_{ic})^2}$$

dove  $x_{ijc} \in \{0,1\}$  rappresenta la risposta del partecipante  $j$  alla perizia  $i$  nella condizione  $c$ .

Nel caso di variabile binaria, la deviazione standard è funzione della proporzione di risposte positive ( $p$ ), secondo la relazione:

$$SD = \sqrt{p(1-p)}$$

Essa raggiunge il valore massimo (0.5) quando le risposte sono equamente divise tra 0 e 1, e si annulla quando vi è perfetta convergenza tra valutatori. Anche in questo caso, pertanto, la deviazione standard costituisce una misura diretta della discrezionalità.

Per ciascuna perizia è stato quindi calcolato il differenziale:

$$\Delta_i = SD_{i,0} - SD_{i,1}$$

dove un valore positivo indica una riduzione della discrezionalità in presenza del supporto AI.

### 5.3.5 Risultati

L'analisi è stata condotta su 12 perizie. Le statistiche descrittive sono riportate nella Tabella 5.2.

Tabella 5.2: Confronto della discrezionalità (SD) – Risposte binarie

	Senza AI	Con AI	Differenza media ( $\Delta$ )
Media SD	0.444	0.435	0.009
Deviazione standard	0.146	0.079	0.164
Numero perizie			12

Il paired t-test sulle deviazioni standard perizia-perizia fornisce:

$$t(11) = 0.195, \quad p = 0.849$$

L'intervallo di confidenza al 95% della differenza media è:

$$[-0.095 ; 0.113]$$

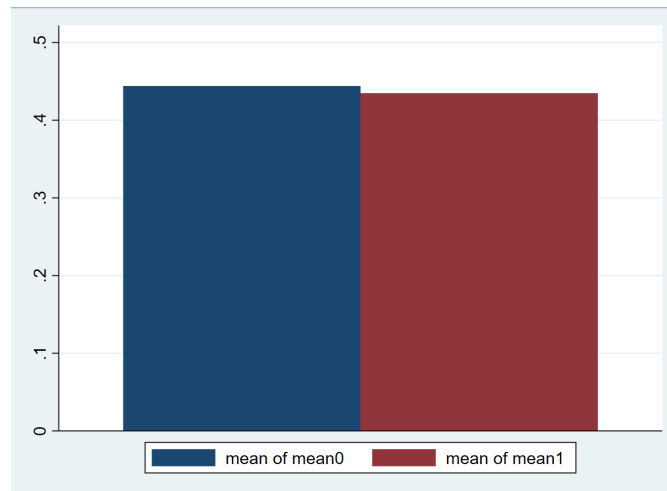


Figura 5.3: Discrezionalità media nelle risposte binarie per condizione

### 5.3.6 Interpretazione

I risultati mostrano che la discrezionalità media nelle risposte binarie è leggermente più elevata in assenza di AI ( $SD = 0.444$ ) rispetto alla condizione con AI ( $SD = 0.435$ ). Tuttavia, la **differenza** osservata è estremamente contenuta ( $\Delta = 0.009$ ) e **statisticamente non significativa**.

L'ampiezza dell'intervallo di confidenza, che include ampiamente lo zero, conferma l'assenza di un effetto sistematico del supporto AI sulla convergenza dei giudizi binari.

In termini sostanziali, i valori di deviazione standard osservati (prossimi al valore massimo teorico di 0.5) indicano una marcata eterogeneità nelle risposte dei partecipanti in entrambe le condizioni. **L'introduzione del supporto AI non appare dunque associata a una riduzione significativa della dispersione decisionale.**

Analogamente a quanto osservato per le scale Likert, l'unità di analisi è rappresentata dalle perizie ( $n = 12$ ), elemento che limita la potenza statistica dell'inferenza e suggerisce cautela nell'interpretazione dei risultati.

Nel complesso, l'evidenza empirica non supporta l'ipotesi che il supporto AI produca una riduzione significativa della discrezionalità decisionale nelle valutazioni binarie.

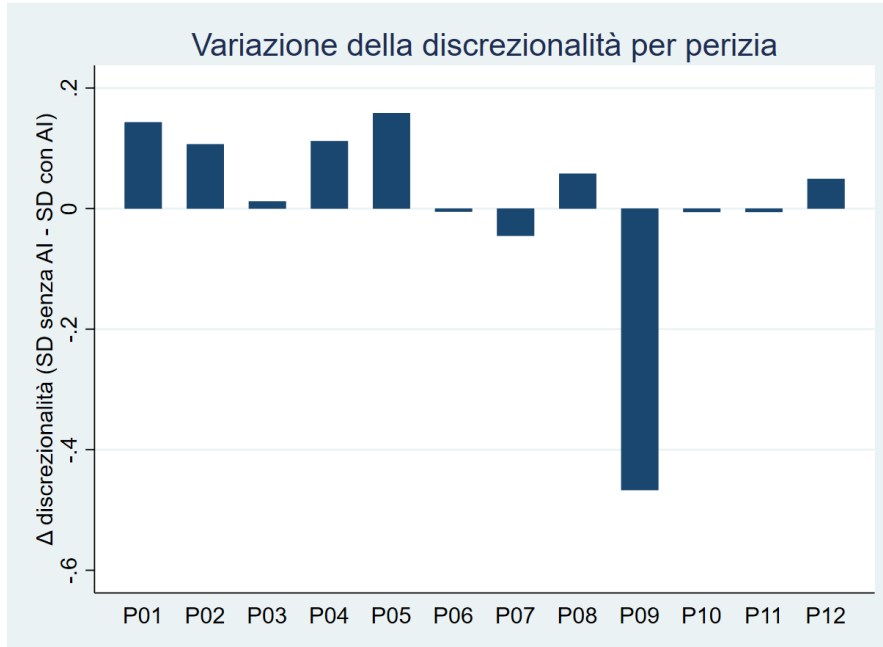


Figura 5.4: Differenza di discrezionalità per singola perizia ( $\Delta_i$ )

### 5.3.7 Discrezionalità decisionale a livello di partecipante

Al fine di analizzare la variabilità delle valutazioni a livello individuale, è stata calcolata una misura di discrezionalità decisionale per ciascun partecipante. In particolare, la discrezionalità è stata misurata come deviazione standard delle valutazioni espresse dal partecipante sulle domande relative a causa, effetto e stima economica del danno (Q01–Q03) nelle dodici perizie considerate. Questa misura consente di cogliere il grado di dispersione delle valutazioni individuali: valori più elevati indicano una maggiore variabilità delle risposte e quindi una maggiore discrezionalità decisionale.

Poiché ogni partecipante ha svolto l'esperimento in entrambe le condizioni sperimentali, con e senza supporto algoritmico, la discrezionalità è stata calcolata

separatamente per ciascuna condizione. Successivamente, per confrontare la variabilità delle valutazioni tra le due condizioni, è stato applicato un test t per campioni appaiati, che consente di confrontare le medie della discrezionalità osservata negli stessi partecipanti nelle due condizioni sperimentali. Va tuttavia considerato che, nel disegno sperimentale adottato, ciascun partecipante ha valutato un numero diverso di perizie nelle due condizioni (otto con supporto algoritmico e quattro senza AI). Tale asimmetria è stata compensata a livello di campione attraverso la rotazione delle perizie nei tre test sperimentali (A, B e C), così che ogni perizia fosse valutata sia con supporto algoritmico sia senza AI da gruppi diversi di partecipanti.

I risultati mostrano una discrezionalità media pari a 1.379 nella condizione senza supporto algoritmico e pari a 1.462 nella condizione con AI. La differenza tra le due medie risulta pari a -0.083. Il test t appaiato non evidenzia una differenza statisticamente significativa al livello convenzionale del 5% ( $t = -1.97$ ,  $p = 0.067$ ), sebbene emerga una lieve tendenza verso una maggiore variabilità delle valutazioni nella condizione con supporto algoritmico.

Nel complesso, questi risultati suggeriscono che **l'introduzione del sistema di supporto algoritmico non ha prodotto una riduzione significativa della discrezionalità decisionale dei partecipanti**. La variabilità delle valutazioni individuali rimane infatti sostanzialmente comparabile tra le due condizioni sperimentali.

Tabella 5.3: Discrezionalità decisionale per condizione sperimentale

Condizione	Media	Deviazione standard	N
Senza AI	1.379	0.249	17
Con AI	1.462	0.229	17

La Figura 5.5 rappresenta graficamente il confronto tra la discrezionalità media osservata nelle due condizioni sperimentali.

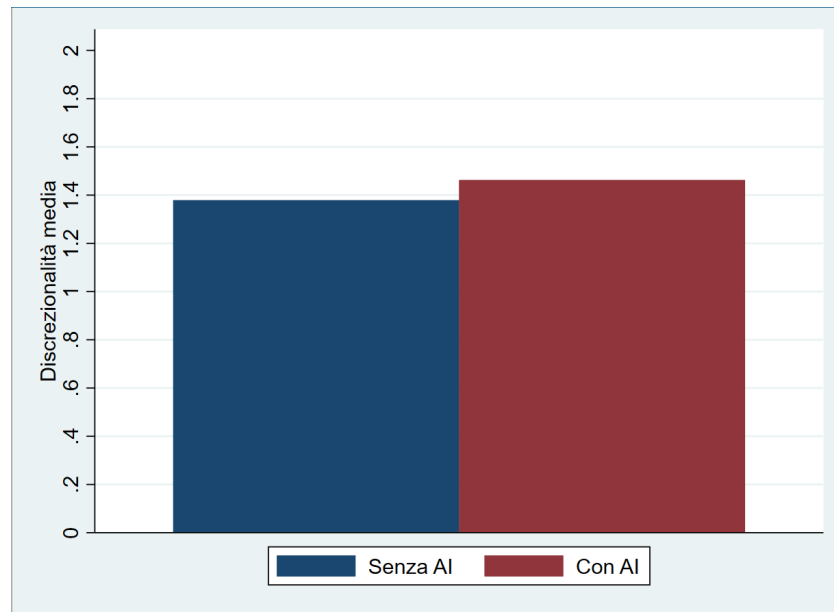


Figura 5.5: Discrezionalità decisionale media dei partecipanti nelle condizioni con e senza supporto algoritmico.

### 5.3.8 Sintesi comparativa dei risultati

L'analisi della discrezionalità è stata condotta separatamente per le risposte su scala Likert e su scala binaria: (i) valutazioni medie su scala Likert e (ii) risposte binarie corretto/errato. Tale duplice approccio consente di verificare la robustezza dei risultati rispetto alla natura della variabile decisionale.

Nel caso delle scale Likert, la discrezionalità media risulta sostanzialmente invariata tra le due condizioni sperimentali, con differenze di entità minima e statisticamente non significative. Analogamente, considerando le risposte binarie, la deviazione standard media per perizia passa da 0.444 (senza AI) a 0.435 (con AI), con una differenza media pari a 0.009 e non significativa dal punto di vista statistico.

La coerenza dei risultati ottenuti attraverso due misure differenti rafforza l'evidenza empirica secondo cui l'introduzione del supporto AI, nel contesto analizzato, non determina una riduzione sistematica della discrezionalità tra valutatori. In

altre parole, **la presenza del suggerimento algoritmico non sembra indurre una maggiore convergenza decisionale né nelle valutazioni ordinali né nelle decisioni dicotomiche.**

Nel complesso, i risultati suggeriscono che la variabilità tra decisori costituisce una componente strutturale del processo valutativo e che il supporto AI, pur potendo incidere su altre dimensioni (ad esempio accuratezza o affidamento), non appare in grado di ridurre in modo significativo la dispersione dei giudizi tra partecipanti. Questa evidenza risulta coerente con quanto osservato da Kolkman et al. (2024) [11], i quali rilevano una limitata concordanza tra valutatori anche quando esposti agli stessi casi, con accordo tra paralegali in soli 7 casi su 18.

Tale evidenza deve tuttavia essere interpretata alla luce del numero limitato di unità di analisi (12 perizie), che riduce la potenza statistica del confronto e potrebbe attenuare la capacità di intercettare effetti di entità contenuta.

## 5.4 Accuratezza delle decisioni

L'accuratezza decisionale è stata analizzata al fine di valutare l'effetto del supporto algoritmico sulla performance dei partecipanti nelle tre condizioni sperimentali: assenza di AI, AI con suggerimento corretto e AI con suggerimento errato. In particolare, l'analisi mira a verificare se e in che misura la presenza del suggerimento algoritmico influenzi la capacità dei partecipanti di fornire risposte coerenti con il *ground truth* definito per ciascuna domanda.

L'accuratezza rappresenta una dimensione centrale nella valutazione dei sistemi di supporto decisionale basati su intelligenza artificiale, poiché consente di osservare se il supporto algoritmico contribuisca a migliorare la qualità delle decisioni umane oppure se, al contrario, possa introdurre effetti distorsivi nel processo di valutazione.

### 5.4.1 Costruzione della variabile di accuratezza

I dati esportati dalla piattaforma Moodle sono stati preliminarmente riorganizzati e convertiti in modo da ottenere un'unità di analisi corrispondente alla singola combinazione partecipante–perizia–domanda, come riportato in Appendice B .

Per ciascuna osservazione è stato effettuato un merge con il file contenente il *ground truth*, ossia la risposta corretta per ogni domanda. Le risposte neutrali (valore centrale della scala Likert originaria) sono state escluse dall'analisi, in quanto non classificabili in modo dicotomico.

È stata quindi costruita una variabile binaria di correttezza, pari a 1 quando la risposta del partecipante risultava coerente con il *ground truth* (totalmente d'accordo e più o meno d'accordo) e 0 in caso contrario (totalmente in disaccordo e più o meno in disaccordo). L'accuratezza delle decisioni è stata quindi calcolata confrontando la risposta del partecipante con il *ground truth* attraverso l'impostazione della *confusion matrix*. Tale approccio consente di classificare le osservazioni in quattro categorie: veri positivi (TP), veri negativi (TN), falsi positivi (FP) e falsi negativi (FN). In particolare, TP rappresenta i casi in cui il partecipante fornisce correttamente una risposta positiva, TN i casi in cui fornisce correttamente una risposta negativa, FP i casi in cui viene fornita una risposta positiva errata e FN i casi in cui viene fornita una risposta negativa errata.

Sulla base di tali componenti, l'accuratezza complessiva è stata calcolata come la proporzione di decisioni corrette sul totale delle decisioni osservate:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

dove il numeratore rappresenta il numero totale di decisioni corrette, mentre il denominatore corrisponde al numero complessivo di decisioni valutate.

Una volta calcolata l'accuratezza per ciascuna delle tre condizioni sperimentali (senza AI, AI con suggerimento corretto e AI con suggerimento errato), le differenze tra le proporzioni di decisioni corrette sono state confrontate mediante test del chi-quadro applicati a coppie di condizioni. Questo approccio consente di verificare se le

eventuali differenze di accuratezza osservate tra le condizioni siano statisticamente significative.

### 5.4.2 Risultati descrittivi

Come riportato nella Tabella 5.4, nella condizione senza supporto AI le osservazioni valide risultano pari a  $N = 260$ , con un'accuratezza media dell'87,69%.

Nella condizione con suggerimento AI corretto ( $N = 272$ ), l'accuratezza raggiunge il 100%, con assenza totale di errori.

Al contrario, nella condizione con suggerimento AI errato ( $N = 263$ ), l'accuratezza si riduce al 75,66%, evidenziando un incremento consistente delle risposte errate rispetto alla decisione autonoma.

Tabella 5.4: Accuratezza decisionale e confronti tra condizioni sperimentali

Condizione	N	Accuratezza (%)
No AI	260	87,69
AI corretta	272	100,00
AI errata	263	75,66

Confronto	$\chi^2(1)$	p-value
No AI vs AI corretta	35,62	< .001
No AI vs AI errata	12,62	< .001
AI corretta vs AI errata	75,18	< .001

Nota. I confronti sono stati effettuati mediante test del chi-quadro di Pearson a due code.

### 5.4.3 Confronti statistici

Al fine di verificare la significatività delle differenze osservate tra le condizioni, sono stati effettuati test del chi-quadro di Pearson a due a due.

Il confronto tra **la condizione senza AI e la condizione con AI corretta** evidenzia una differenza statisticamente significativa ( $\chi^2(1) = 35,62$ ,  $p < .001$ ),

indicando che **la presenza di un suggerimento corretto incrementa in modo significativo l'accuratezza rispetto alla decisione autonoma.**

Il confronto tra **la condizione senza AI e la condizione con AI errata** risulta anch'esso significativo ( $\chi^2(1) = 12,62, p < .001$ ), mostrando che **un suggerimento scorretto riduce significativamente la performance.**

Infine, il confronto tra **AI corretta e AI errata** evidenzia una differenza altamente significativa ( $\chi^2(1) = 75,18, p < .001$ ), confermando che **la qualità del suggerimento algoritmico costituisce un fattore determinante nel processo decisionale.**

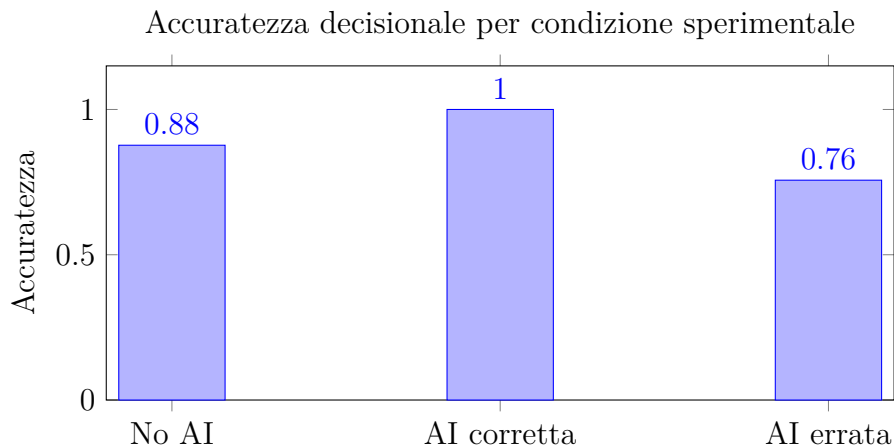


Figura 5.6: Accuratezza decisionale nelle tre condizioni sperimentali

La Figura 5.6 rappresenta graficamente il livello di accuratezza decisionale nelle tre condizioni sperimentali. Come evidenziato, la performance raggiunge il valore massimo quando il suggerimento dell'AI è corretto, mentre diminuisce sensibilmente nella condizione con suggerimento errato. La condizione senza supporto AI si colloca in una posizione intermedia tra le due.

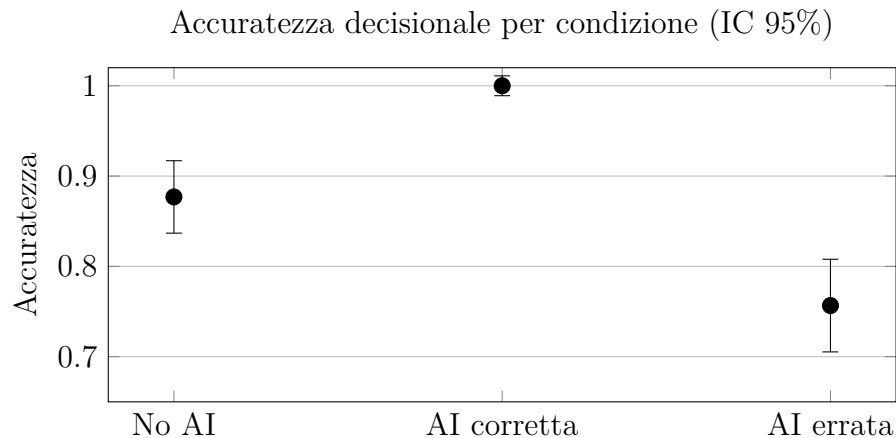


Figura 5.7: Accuratezza media nelle tre condizioni sperimentali con intervalli di confidenza al 95%.

La Figura 5.7 sintetizza graficamente l'accuratezza nelle tre condizioni con intervalli di confidenza al 95%. L'accuratezza aumenta fino al valore massimo in presenza di suggerimento corretto, mentre si riduce significativamente quando il suggerimento è errato, confermando la forte dipendenza della performance dalla qualità dell'indicazione algoritmica.

#### 5.4.4 Interpretazione dei risultati

Nel complesso, i risultati mostrano un effetto marcato del supporto algoritmico sull'accuratezza decisionale. In presenza di suggerimento corretto, i partecipanti si allineano completamente all'indicazione fornita dall'AI, raggiungendo un livello massimo di performance. Quando il suggerimento è errato, invece, l'accuratezza si riduce significativamente rispetto alla condizione autonoma.

La condizione senza AI si colloca a un livello intermedio: la performance è elevata, ma inferiore a quella ottenuta con AI corretta e superiore a quella osservata con AI errata.

**Il pattern osservato suggerisce una forte influenza del suggerimento algoritmico sul comportamento decisionale, evidenziando come il sistema**

migliori la performance quando è accurato ma la peggiori quando fornisce un'indicazione errata. Questo risultato appare coerente con le evidenze riportate da Jussupow et al. (2021) [16], i quali osservano che l'accuratezza delle decisioni aumenta in presenza di suggerimenti algoritmici corretti ma diminuisce quando l'AI fornisce indicazioni errate, mostrando come l'effetto del supporto algoritmico dipenda in modo cruciale dalla qualità del suggerimento fornito.

Un risultato analogo emerge anche nello studio di Kolkman et al. (2024) [11], nel quale l'introduzione di un sistema di supporto algoritmico influenza significativamente le decisioni dei valutatori umani. Nel loro esperimento, i paralegali identificano correttamente la decisione del giudice nel 67% dei casi senza AI, mentre l'accuratezza scende al 48% quando le decisioni sono prese con il supporto del sistema. Tale evidenza suggerisce che l'assistenza algoritmica può modificare in modo rilevante il processo decisionale umano, in particolare quando gli utenti tendono ad affidarsi alle indicazioni fornite dal sistema.

#### 5.4.5 Confronto dell'accuratezza tra domande di causa ed effetto

Per verificare se la performance decisionale dei partecipanti variasse in funzione della tipologia di domanda, è stato effettuato un confronto tra l'accuratezza delle risposte relative alla causa del danno (Q01) e quelle relative all'effetto del danno (Q02).

A tal fine è stata costruita una tabella di contingenza tra la correttezza della risposta (corretta vs errata) e il tipo di domanda (causa vs effetto). L'associazione tra le due variabili è stata valutata mediante test del chi-quadro di Pearson.

I risultati mostrano un'accuratezza pari al 66.9% per le domande relative alla causa e al 69.2% per le domande relative all'effetto. Sebbene la percentuale di risposte corrette risulti leggermente più elevata per le domande sull'effetto, **il test del chi-quadro non evidenzia una differenza statisticamente significativa tra le due tipologie di domanda** ( $\chi^2(1) = 0.22$ ,  $p = 0.639$ ).

Nel complesso, questi risultati suggeriscono che i partecipanti non mostrano differenze sostanziali nella capacità di identificare correttamente la causa del danno rispetto alla valutazione dei suoi effetti.

Tabella 5.5: Accuratezza delle risposte per tipologia di domanda

Tipo di domanda	Errate	Corrette
Causa (Q01)	56 (33.1%)	113 (66.9%)
Effetto (Q02)	57 (30.8%)	128 (69.2%)

La Figura 5.8 rappresenta graficamente l'accuratezza media osservata per le due tipologie di domanda.

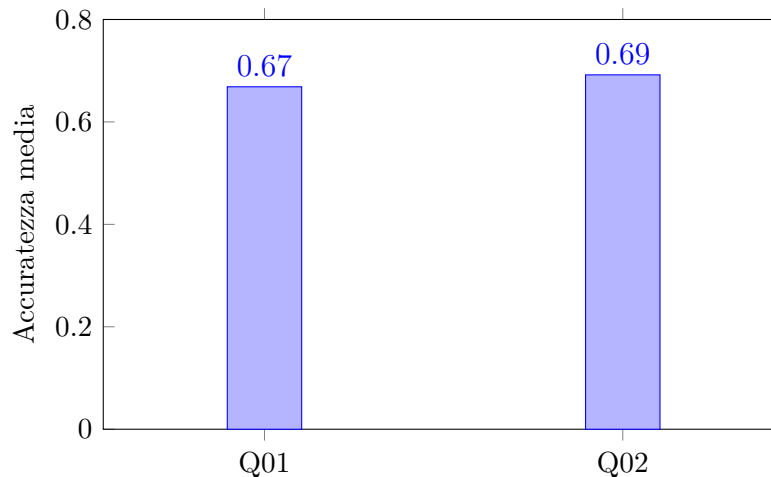


Figura 5.8: Accuratezza media delle risposte nelle domande relative alla causa (Q01) e all'effetto (Q02).

#### 5.4.6 Accuratezza della stima economica

Oltre alle valutazioni relative al nesso di causa ed effetto, è stata analizzata anche l'accuratezza delle decisioni riguardanti la stima economica del danno (domanda Q03). L'obiettivo dell'analisi è verificare se la presenza del supporto algoritmico influenzi la capacità dei partecipanti di fornire una valutazione coerente con il *ground truth* definito per ciascuna perizia.

Per rendere confrontabili le risposte con il riferimento corretto, le valutazioni originariamente espresse su scala Likert a cinque punti sono state ricodificate in forma binaria. In particolare, le risposte pari a 4 o 5 sono state classificate come 1, mentre le risposte pari a 1 o 2 sono state classificate come 0. Le eventuali risposte neutrali (valore 3) sono state escluse dall'analisi.

Il *ground truth* è stato definito in modo differenziato per le diverse perizie: per le perizie da P01 a P06 la risposta corretta corrisponde al valore 1, mentre per le perizie da P07 a P12 la risposta corretta corrisponde al valore 0. Sulla base di questa codifica è stata costruita una variabile binaria di accuratezza, pari a 1 quando la risposta del partecipante coincide con il *ground truth* e pari a 0 in caso contrario.

La Tabella 5.6 riporta il confronto tra le due condizioni sperimentali, distinguendo tra decisioni corrette e scorrette.

Tabella 5.6: Accuratezza della stima economica per condizione sperimentale

	Senza AI	Con AI	Totale
Decisioni scorrette	29	70	99
Decisioni corrette	27	55	82
Totale	56	125	181

L'accuratezza complessiva nella condizione senza supporto AI risulta pari al 48.2% (27 decisioni corrette su 56), mentre nella condizione con supporto AI l'accuratezza è pari al 44.0% (55 decisioni corrette su 125).

La Figura 5.9 rappresenta graficamente il confronto tra le due condizioni sperimentali.

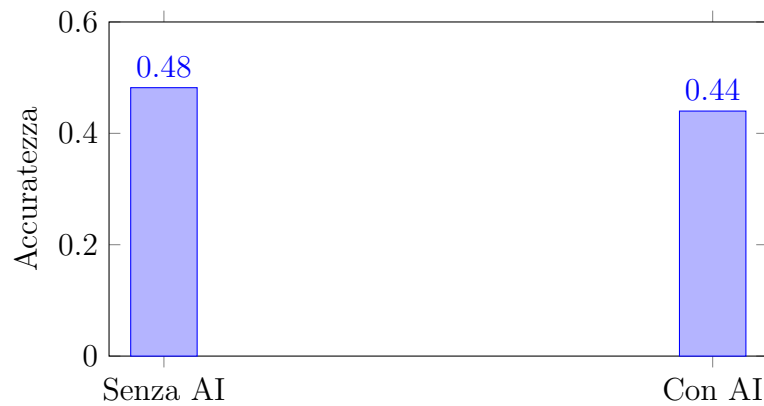


Figura 5.9: Accuratezza della stima economica nelle due condizioni sperimentali

Per approfondire la distribuzione delle decisioni corrette e degli errori, sono state costruite le *confusion matrix* tra le risposte fornite dai partecipanti e il *ground truth*.

Tabella 5.7: Confusion matrix – Condizione senza AI

	Risposta = 0	Risposta = 1
Ground truth = 0	16	16
Ground truth = 1	13	11

Tabella 5.8: Confusion matrix – Condizione con AI

	Risposta = 0	Risposta = 1
Ground truth = 0	40	24
Ground truth = 1	46	15

Per verificare se la differenza tra le due condizioni fosse statisticamente significativa è stato applicato un test del chi-quadro su tabella di contingenza tra accuratezza e condizione sperimentale. Il risultato del test non evidenzia una differenza significativa tra le due condizioni ( $\chi^2(1) = 0.277, p = 0.599$ ).

Nel complesso, i risultati indicano che la presenza del supporto algoritmico non produce un miglioramento significativo dell'accuratezza nella valutazione della stima economica. La distribuzione delle decisioni corrette e scorrette risulta infatti

simile nelle due condizioni sperimentali, suggerendo che l'utilizzo del suggerimento algoritmico non ha influenzato in modo significativo la performance decisionale dei partecipanti in questo specifico compito.

### 5.4.7 Coincidenza delle risposte con il ground truth nella condizione AI

Al fine di approfondire ulteriormente il comportamento dei partecipanti in presenza di supporto algoritmico, è stata analizzata la quota di risposte che coincidono con il *ground truth* limitatamente alla condizione con AI.

Considerando esclusivamente le osservazioni relative alla presenza del sistema di supporto decisionale, si osservano complessivamente 125 valutazioni relative alla stima economica. Di queste, 55 risultano coerenti con il *ground truth*, mentre 70 non coincidono con il riferimento corretto.

Tabella 5.9: Coincidenza delle risposte con il ground truth nella condizione con AI

Accuratezza	Frequenza	Percentuale
Non coincidente con GT	70	56.0%
Coincidente con GT	55	44.0%
Totale	125	100%

L'accuratezza complessiva nella condizione con supporto AI risulta pertanto pari al 44.0%. In altri termini, meno della metà delle valutazioni formulate con il supporto algoritmico risulta coerente con il *ground truth* definito per le perizie considerate. Questo risultato suggerisce che la presenza del suggerimento algoritmico non conduce necessariamente a decisioni più accurate, ma può essere seguita anche quando non conduce alla risposta corretta.

### 5.4.8 Comportamento rispetto alla correttezza del suggerimento algoritmico

Per approfondire ulteriormente il comportamento dei partecipanti in presenza di supporto algoritmico, è stata analizzata la relazione tra la correttezza del suggerimento AI e la risposta fornita dagli utenti.

Nel disegno sperimentale, i suggerimenti algoritmici risultano corretti per le perizie da P01 a P06, mentre risultano deliberatamente errati per le perizie da P07 a P12. Questa distinzione consente di osservare separatamente due dinamiche: da un lato la capacità dei partecipanti di seguire il suggerimento quando esso è corretto, dall'altro la loro capacità di resistere al suggerimento quando esso è errato.

La Tabella 5.10 riporta la distribuzione delle risposte corrette e scorrette nella condizione con supporto AI in funzione della correttezza del suggerimento algoritmico.

Tabella 5.10: Risposte dei partecipanti in funzione della correttezza del suggerimento AI

	AI corretta	AI sbagliata	Totale
Risposta errata	46	24	70
Risposta corretta	15	40	55
Totale	61	64	125

Quando il suggerimento algoritmico è corretto (perizie P01–P06), solo 15 decisioni su 61 risultano coerenti con il *ground truth*, pari al 24.6% dei casi. Ciò suggerisce che i partecipanti tendono frequentemente a non seguire il suggerimento algoritmico anche quando esso è corretto.

Quando invece il suggerimento algoritmico è errato (perizie P07–P12), 40 decisioni su 64 risultano corrette, pari al 62.5% dei casi. In questi casi i partecipanti mostrano quindi una maggiore capacità di discostarsi dal suggerimento algoritmico e di formulare una valutazione autonoma.

Nel complesso, i risultati indicano un comportamento selettivo nei confronti del suggerimento AI: i partecipanti non seguono sistematicamente le raccomandazioni

del sistema, ma mostrano una certa capacità di discostarsene quando le ritengono non plausibili. Allo stesso tempo, la presenza di 24 decisioni errate in presenza di suggerimenti algoritmici sbagliati suggerisce che in alcuni casi il sistema può comunque influenzare negativamente la decisione degli utenti.

Nel complesso, i risultati ottenuti risultano coerenti con quanto evidenziato nella letteratura sul comportamento degli utenti nei confronti dei sistemi di supporto decisionale basati su AI. In particolare, Kolkman et al. (2024) [11] osservano come i decisori tendano a non seguire automaticamente le raccomandazioni algoritmiche, ma adottino piuttosto un atteggiamento selettivo nei confronti del suggerimento fornito dal sistema. Anche nel presente esperimento emerge una dinamica analoga: i partecipanti non si affidano sistematicamente al suggerimento dell'AI, ma mantengono un certo grado di discrezionalità decisionale. Allo stesso tempo, la presenza di decisioni errate in corrispondenza di suggerimenti algoritmici sbagliati suggerisce che il sistema può comunque influenzare il processo decisionale in alcuni casi. Nel complesso, questi risultati indicano che il supporto algoritmico non sostituisce il giudizio umano, ma tende piuttosto a interagire con esso, producendo un equilibrio tra autonomia decisionale e potenziale condizionamento da parte dell'AI.

## 5.5 Affidamento degli utenti inesperti al sistema di supporto

Per analizzare il grado di affidamento degli utenti inesperti al sistema di supporto basato su intelligenza artificiale, è stata costruita una misura comportamentale denominata *follow rate*, definita come la probabilità che il partecipante aderisca al suggerimento fornito dall'AI.

Per ogni item (perizia–domanda) è stata definita una variabile binaria:

$$follow_i = \begin{cases} 1 & \text{se } risposta_i = suggerimento_{AI_i} \\ 0 & \text{altrimenti} \end{cases}$$

Il suggerimento AI è stato ricostruito coerentemente con la manipolazione sperimentale:

- perizie P01–P06: suggerimento corretto (coincidente con la risposta corretta);
- perizie P07–P12: suggerimento errato (non coincidente con la risposta corretta).

L'analisi è stata condotta considerando esclusivamente le decisioni in presenza di AI (N = 770), di cui 382 con AI corretta e 388 con AI sbagliata.

### 5.5.1 Risultati aggregati

Tabella 5.11: Distribuzione del follow per condizione AI

	AI corretta	AI sbagliata	Totale
Non segue (0)	216	269	485
Segue (1)	166	119	285
Totale	382	388	770

In termini percentuali, il suggerimento viene seguito nel 43,8% dei casi quando l'AI è corretta e nel 31,0% dei casi quando l'AI è sbagliata. Come mostrato in Figura 5.10 e in Figura 5.11, il suggerimento dell'AI viene seguito con maggiore probabilità quando è corretto (43,8%) rispetto alla condizione in cui è errato (31,0%). Il follow rate complessivo risulta pari al 37,0%.

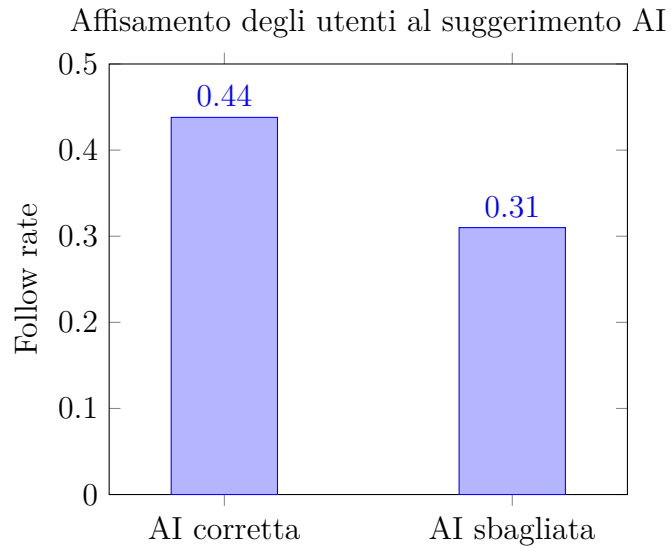


Figura 5.10: Follow rate per condizione di suggerimento AI

La Figura 5.10 rappresenta graficamente il follow rate nelle due condizioni sperimentali. Come evidenziato, il suggerimento dell'AI viene seguito con maggiore probabilità quando è corretto (43,8%) rispetto a quando è errato (31,0%). Per una rappresentazione più informativa della stima, la Figura 5.11 riporta il follow rate medio con intervalli di confidenza al 95%.

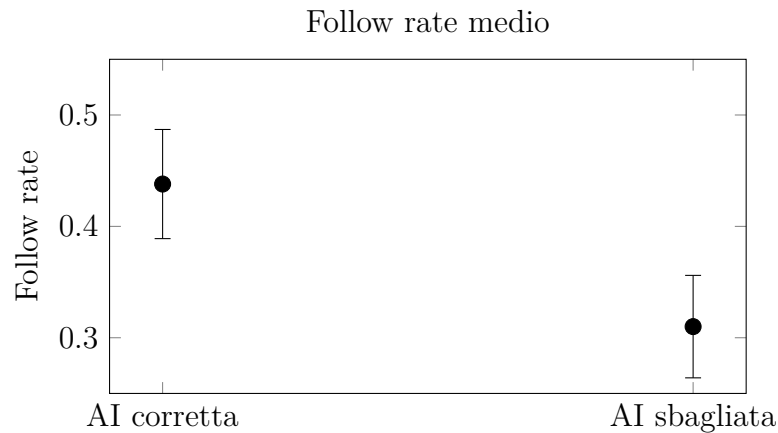


Figura 5.11: Follow rate medio con intervalli di confidenza al 95% per condizione AI

L'analisi dei risultati evidenzia come la presenza del suggerimento fornito dal

sistema di supporto influenzi in maniera significativa il comportamento decisionale degli utenti inesperti. In particolare, il follow rate risulta più elevato nella condizione in cui il suggerimento dell'AI è corretto (43,8%) rispetto alla condizione in cui il suggerimento è errato (31,0%). Questo risultato suggerisce che, pur non affidandosi completamente al sistema, i partecipanti tendono a considerare l'indicazione fornita dall'AI come un elemento informativo rilevante nel processo decisionale.

Il fatto che il follow rate non raggiunga valori prossimi al 100% nella condizione di AI corretta indica inoltre che gli utenti non adottano un comportamento di affidamento cieco al sistema, ma mantengono un certo grado di autonomia decisionale. Allo stesso tempo, la presenza di un follow rate non trascurabile anche nella condizione di AI sbagliata evidenzia come il suggerimento dell'AI possa comunque influenzare le decisioni degli utenti, generando in alcuni casi un fenomeno di affidamento eccessivo.

Nel complesso, questi risultati suggeriscono che il sistema di supporto svolge effettivamente un ruolo nel guidare le scelte degli utenti inesperti, ma che tale influenza non è assoluta. Piuttosto, il comportamento osservato appare coerente con una forma di affidamento moderato, in cui gli utenti utilizzano il suggerimento dell'AI come supporto informativo, senza rinunciare completamente al proprio giudizio.

### 5.5.2 Analisi a livello individuale

Le osservazioni sono state aggregate per partecipante ( $N = 17$ ) calcolando il follow medio nelle due condizioni e la differenza intra-soggetto  $\Delta follow$ .

Tabella 5.12: Statistiche descrittive individuali

Variabile	N	Media	Dev. Std.	Min	Max
Follow_correct	17	0,438	0,230	0,167	0,850
Follow_wrong	17	0,310	0,182	0,087	0,625
$\Delta follow$	17	0,128	0,395	-0,458	0,759

**Test t sulla differenza intra-soggetto** È stato effettuato un t-test a un campione sulla differenza individuale rispetto a zero.

Tabella 5.13: Risultati t-test su  $\Delta follow$

Variabile	N	Media	Std. Err.	t	df	p-value
$\Delta follow$	17	0,128	0,096	1,336	16	0,200

Intervallo di confidenza al 95%: [-0,075 ; 0,331].

La Figura 5.12 mostra la distribuzione delle differenze individuali nel follow rate tra le due condizioni. Sebbene la media della differenza sia positiva, indicando una maggiore adesione quando l'AI è corretta, si osserva una notevole dispersione tra i partecipanti. Alcuni soggetti presentano differenze marcate, mentre altri mostrano valori prossimi allo zero o addirittura negativi, evidenziando comportamenti decisionali eterogenei.

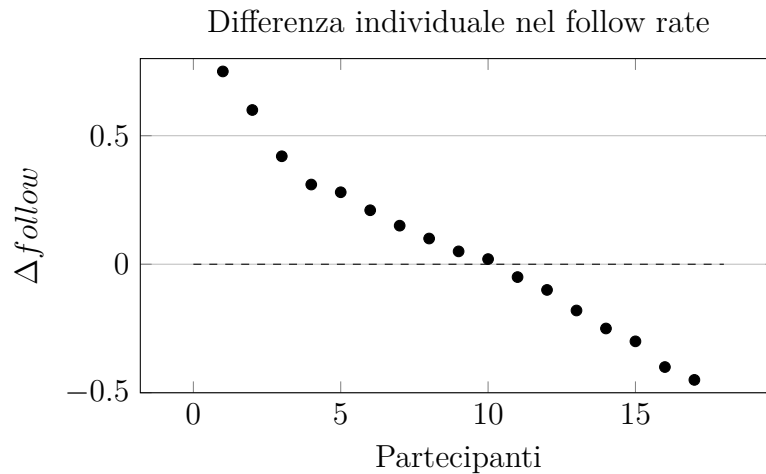


Figura 5.12: Differenza individuale nel follow rate ( $\Delta follow = Follow\_correct - Follow\_wrong$ )

### 5.5.3 Risultati aggregati

L'analisi delle 770 decisioni effettuate in presenza di suggerimento algoritmico evidenzia una differenza sistematica nel tasso di adesione in funzione della correttezza del suggerimento.

Come riportato in Tabella 5.11, quando l'AI fornisce un suggerimento corretto i partecipanti lo seguono nel 43,8% dei casi (166 su 382 osservazioni). Quando invece il suggerimento è errato, il tasso di adesione scende al 31,0% (119 su 388 osservazioni).

La differenza osservata tra le due condizioni è pari a 12,8 punti percentuali. Questo scarto indica che, a livello aggregato, **i partecipanti mostrano una maggiore propensione a seguire l'AI quando essa è allineata alla risposta corretta rispetto a quando propone un'indicazione errata.**

Il **follow rate complessivo**, calcolato considerando congiuntamente entrambe le condizioni sperimentali, risulta pari al 37,0%. Questo valore suggerisce che il sistema di supporto esercita un'influenza concreta sulle decisioni: oltre un terzo delle risposte fornite in presenza di AI coincide con il suggerimento algoritmico.

Tuttavia, il fatto che l'adesione non superi il 50% in nessuna delle due condizioni indica che i partecipanti non adottano un comportamento di accettazione automatica del suggerimento. L'AI viene seguita con maggiore frequenza quando è corretta, ma continua a essere seguita in una quota non trascurabile di casi anche quando è errata, evidenziando una forma di influenza algoritmica non trascurabile.

Questo risultato appare coerente con quanto emerso nell'analisi dell'accuratezza decisionale. In presenza di suggerimenti corretti, l'adesione all'AI contribuisce infatti ad aumentare la probabilità di fornire una risposta corretta; al contrario, quando il suggerimento è errato, il follow dell'AI tende a ridurre la performance dei partecipanti. Il comportamento osservato suggerisce quindi che il suggerimento algoritmico agisce come un fattore di orientamento decisionale, con effetti che dipendono dalla qualità dell'informazione fornita dal sistema.

Per comprendere se tale differenza fosse sistematica anche a livello intra-soggetto, le osservazioni sono state aggregate per partecipante ( $N = 17$ ), calcolando il tasso medio di follow nelle due condizioni e la differenza individuale:

$$\Delta follow = Follow\_correct - Follow\_wrong$$

Le statistiche descrittive mostrano che:

- il follow medio in presenza di AI corretta è pari a 0,438 (DS = 0,230);
- il follow medio in presenza di AI sbagliata è pari a 0,310 (DS = 0,182);
- la differenza media intra-soggetto è pari a 0,128 (DS = 0,395).

La deviazione standard relativamente elevata della differenza ( $DS = 0,395$ ) segnala un'ampia variabilità individuale nella tendenza a modulare l'adesione in funzione della qualità del suggerimento.

Per verificare se la differenza media fosse statisticamente diversa da zero, è stato effettuato un t-test a un campione. Il risultato ( $t(16) = 1,336$ ,  $p = 0,200$ ; IC 95% [-0,075 ; 0,331]) indica che la differenza osservata non è statisticamente significativa al livello convenzionale del 5%.

Sebbene la direzione dell'effetto sia coerente con un comportamento selettivo, maggiore adesione quando l'AI è corretta, **l'evidenza statistica non consente di concludere che tale differenza sia sistematica nella popolazione di riferimento**. Il risultato appare influenzato sia dalla dimensione contenuta del campione sia dall'elevata eterogeneità nei comportamenti individuali.

Nel complesso, i dati suggeriscono un modello di affidamento parziale: gli utenti inesperti non si affidano in modo automatico al sistema, ma neppure mantengono una piena indipendenza decisionale. L'AI esercita un'influenza concreta sul processo valutativo, ma tale influenza risulta distribuita in modo eterogeneo tra i partecipanti.

Questa dinamica appare coerente con quanto osservato da Kolkman et al. (2024) [11], i quali rilevano che i valutatori umani non seguono sistematicamente le raccomandazioni del sistema algoritmico. Nel loro esperimento, i paralegali ignorano in diversi casi il suggerimento dell'AI, mentre in altri casi sembrano essere influenzati dalle predizioni fornite dal sistema. Gli autori sottolineano infatti che i partecipanti “do not follow recommendations blindly”, evidenziando un atteggiamento selettivo nei confronti dell'informazione algoritmica.

Analogamente a quanto osservato nel presente studio, anche nello studio di Kolkman l'adesione al suggerimento algoritmico appare variabile e non uniforme tra i partecipanti, suggerendo che l'utilizzo di sistemi di supporto basati su AI non determina automaticamente un comportamento di affidamento totale, ma piuttosto un processo decisionale in cui il suggerimento algoritmico viene integrato — o talvolta ignorato — all'interno del giudizio umano.

## 5.6 Impatto del supporto AI sui tempi di esecuzione

Un'ulteriore dimensione analizzata riguarda l'efficienza del processo decisionale, valutata attraverso il tempo di esecuzione del compito. In particolare, è stato esaminato il tempo impiegato dai partecipanti per completare la valutazione delle perizie nelle condizioni con e senza supporto AI.

### 5.6.1 Preparazione e normalizzazione dei dati

I tempi di completamento sono stati esportati dalla piattaforma sperimentale e convertiti in secondi al fine di consentire l'elaborazione statistica. Poiché le condizioni sperimentali prevedevano un diverso numero di perizie da valutare (4 nella condizione senza AI e 8 nella condizione con AI), il tempo totale non risultava direttamente confrontabile. Per garantire la comparabilità tra le condizioni, il tempo è stato normalizzato calcolando il *tempo medio per perizia* per ciascun partecipante:

$$T_{\text{senza AI}}^* = \frac{T_{\text{senza AI}}}{4} \quad T_{\text{con AI}}^* = \frac{T_{\text{con AI}}}{8}$$

Questa trasformazione consente di ottenere una misura comparabile dell'efficienza decisionale individuale, espressa in secondi medi necessari per valutare una singola perizia.

## 5.6.2 Analisi esplorativa e identificazione dei valori anomali

Prima dell'analisi inferenziale è stata condotta un'analisi esplorativa della distribuzione dei tempi mediante diagrammi boxplot, riportati nelle Figure 5.13a e 5.13b. Il boxplot consente di rappresentare la mediana, l'intervallo interquartile e i valori estremi, facilitando l'individuazione di osservazioni anomale.

Il boxplot costruito sui dati completi ha evidenziato la presenza di valori estremi particolarmente elevati (superiori alle due ore), incompatibili con la durata reale del compito sperimentale. Tali valori sono riconducibili a inattività prolungata o interruzioni della sessione e non rappresentano il processo decisionale oggetto di studio.

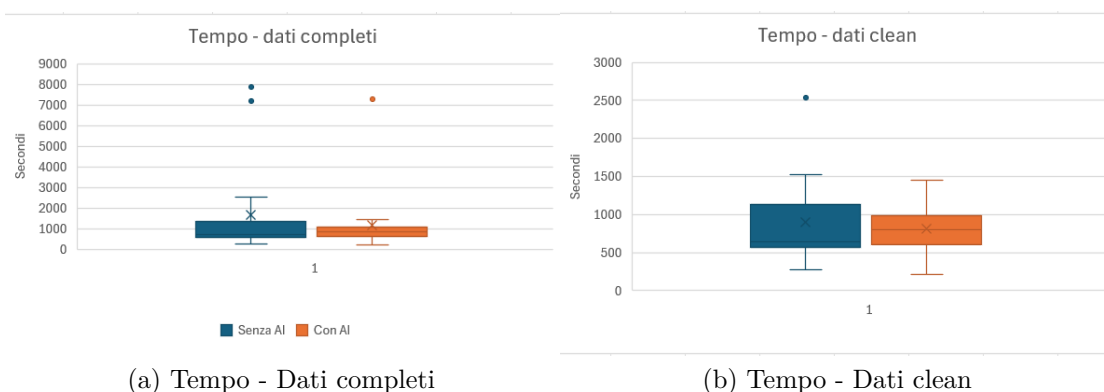


Figura 5.13: Distribuzione dei tempi di esecuzione prima e dopo l'esclusione degli outlier

Queste osservazioni sono state escluse dalle analisi in quanto non valide dal punto di vista sperimentale. Successivamente, è stato prodotto un secondo boxplot sui dati depurati dai valori estremi, al fine di visualizzare correttamente la distribuzione dei tempi effettivi.

È importante sottolineare che valori elevati ma plausibili (ad esempio tempi prossimi ai 40 minuti) sono stati mantenuti, poiché rappresentano la variabilità individuale del comportamento decisionale e non errori di misurazione.

### 5.6.3 Analisi statistica

Dopo la pulizia e normalizzazione dei dati, l'analisi inferenziale è stata condotta mediante il software Stata, che consente l'applicazione di test statistici parametrici appropriati per dati appaiati.

Poiché ciascun partecipante ha completato entrambe le condizioni sperimentali, è stato utilizzato un **paired t-test** per confrontare il tempo medio per perizia nelle due condizioni. Questo test consente di verificare se la differenza media tra le due misure appaiate sia significativamente diversa da zero.

L'analisi è stata effettuata su 15 osservazioni valide, dopo l'esclusione dei casi con dati mancanti.

Tabella 5.14: Effetto del supporto AI sul tempo medio per perizia

Condizione	Media (s)	Dev. Std.
Senza AI	225.77	140.30
Con AI	100.52	42.81

$$t(14) = 4.05, p = 0.0012$$

$$\text{Differenza media} = 125.25 \text{ s}$$

$$\text{IC } 95\% [58.99, 191.51]$$

### Risultati

I risultati del paired t-test evidenziano una **riduzione significativa del tempo medio necessario per la valutazione delle perizie in presenza del supporto AI.**

In particolare:

- senza AI:  $M = 225.77$  secondi per perizia;
- con AI:  $M = 100.52$  secondi per perizia.

La differenza media osservata è pari a:

$$\Delta = 125.25 \text{ secondi}$$

Il test t per campioni appaiati mostra che tale differenza è statisticamente significativa:

$$t(14) = 4.05, \quad p = 0.0012$$

L'intervallo di confidenza al 95% indica che la riduzione del tempo medio è compresa tra circa 59 e 192 secondi.

#### 5.6.4 Interpretazione

I risultati indicano che l'utilizzo del supporto AI determina una significativa riduzione del tempo necessario per la valutazione delle perizie. **In media, i partecipanti impiegano circa due minuti in meno per perizia quando dispongono del supporto del sistema.**

Questo risultato suggerisce che l'AI:

- aumenta l'efficienza del processo decisionale;
- riduce il carico cognitivo richiesto per l'analisi delle informazioni;
- accelera l'identificazione degli elementi rilevanti;
- supporta la formulazione di giudizi in modo più rapido.

In termini relativi, il tempo medio per perizia si riduce di oltre il 50%, evidenziando un impatto sostanziale del supporto AI sulla velocità decisionale.

L'analisi dei tempi di esecuzione evidenzia che il supporto AI non solo migliora la qualità decisionale, ma produce anche un significativo incremento dell'efficienza operativa. La riduzione del tempo necessario per la valutazione delle perizie rappresenta un elemento rilevante in contesti professionali caratterizzati da elevati carichi di lavoro e necessità di decisioni rapide e accurate.

## 5.7 Percezione soggettiva del compito sperimentale

Al fine di valutare come i partecipanti abbiano percepito il compito sperimentale, è stata condotta un'analisi delle risposte fornite al questionario di valutazione soggettiva somministrato al termine dell'esperimento. Tale analisi ha l'obiettivo di verificare se il design sperimentale, il materiale informativo e il supporto fornito risultassero adeguati dal punto di vista degli utenti, consentendo di interpretare i risultati comportamentali alla luce dell'esperienza percepita dai partecipanti.

Il questionario era composto da quattro item valutati su scala Likert a cinque punti (da 1 = totalmente in disaccordo a 5 = totalmente d'accordo), volti a misurare differenti dimensioni della percezione del compito:

- difficoltà complessiva dell'esperimento;
- complessità della valutazione delle perizie;
- adeguatezza del materiale informativo fornito;
- utilità percepita del suggerimento AI nel processo decisionale.

Le risposte raccolte nei tre gruppi sperimentali sono state aggregate e analizzate mediante statistiche descrittive. Per ciascun item sono state calcolate la media, la deviazione standard e la percentuale di partecipanti che ha espresso accordo con l'affermazione (valori pari a 4 o 5 sulla scala Likert).

Tabella 5.15: Percezione soggettiva del compito sperimentale

Item	Media	Dev. Std.	% Accordo
Ho trovato l'esperimento complessivamente difficile	3.47	1.01	58.82%
Valutare le perizie è stato complesso	3.65	1.06	70.59%
Il materiale fornito è stato sufficiente	3.35	0.93	29.41%
Il suggerimento AI ha semplificato la valutazione	4.18	0.88	82.35%

I risultati indicano che il compito sperimentale è stato generalmente percepito come cognitivamente impegnativo, ma non eccessivamente difficile. In particolare, l'esperimento nel suo complesso presenta un livello di difficoltà moderato ( $M = 3.47$ ), mentre la fase di valutazione delle perizie risulta l'aspetto maggiormente complesso del processo decisionale ( $M = 3.65$ ), con oltre il 70% dei partecipanti che concorda sull'elevato grado di complessità richiesto.

Per quanto riguarda l'adeguatezza del materiale informativo, la media pari a 3.35 e la relativamente bassa percentuale di accordo (29.41%) suggeriscono che una parte dei partecipanti abbia percepito le informazioni disponibili come non pienamente sufficienti per formulare le proprie valutazioni. Tale risultato indica la presenza di un certo livello di incertezza informativa, coerente con la natura interpretativa del compito e con la variabilità decisionale osservata nelle analisi precedenti.

Infine, emerge con particolare evidenza il ruolo del supporto algoritmico: il suggerimento AI è stato percepito come elemento di semplificazione del processo valutativo ( $M = 4.18$ ), con oltre l'80% dei partecipanti che dichiara un effetto facilitante nella valutazione delle perizie. Questo dato suggerisce che, pur in presenza di un compito percepito come complesso, l'intervento dell'AI contribuisce a ridurre lo sforzo decisionale richiesto agli utenti.

Nel complesso, l'analisi della percezione soggettiva evidenzia che l'esperimento si colloca in un contesto decisionale realisticamente impegnativo, caratterizzato da un certo grado di incertezza informativa, all'interno del quale il supporto AI viene percepito come uno strumento utile nel facilitare il processo di valutazione.

## **5.8 Risultati e key findings**

L'analisi integrata delle evidenze empiriche consente di formulare una valutazione complessiva dell'impatto del supporto basato su intelligenza artificiale sul processo decisionale degli utenti inesperti. Considerando congiuntamente discrezionalità,

accuratezza rispetto al ground truth, affidamento comportamentale ed efficienza temporale, emerge un risultato generale coerente: **il supporto AI non modifica il grado di autonomia decisionale tra individui, ma incide in modo sostanziale sugli esiti e sulla rapidità delle decisioni.**

Un primo *key finding* riguarda l'assenza di effetti sulla discrezionalità decisionale. In entrambe le valutazioni effettuate, su scala Likert e decisioni binarie, la variabilità tra valutatori rimane sostanzialmente invariata tra condizioni con e senza AI. Le deviazioni standard medie risultano quasi sovrapponibili e le differenze osservate non raggiungono significatività statistica. Questo risultato indica che **l'introduzione del suggerimento algoritmico non produce una convergenza dei giudizi**, suggerendo che la dispersione tra decisori costituisca una componente strutturale del processo valutativo e non venga compressa dalla presenza dell'AI.

Un secondo *key finding*, in netto contrasto con il precedente, riguarda invece l'accuratezza decisionale. I risultati mostrano che la performance dei partecipanti varia drasticamente in funzione della qualità del suggerimento algoritmico: l'accuratezza raggiunge il **100%** in presenza di AI corretta, mentre **diminuisce significativamente** quando il suggerimento è errato rispetto alla condizione autonoma. Ne deriva che **l'AI esercita un'influenza direzionale molto forte sugli esiti decisionali**, amplificando gli effetti informativi del suggerimento fornito.

Il terzo *key finding* emerge dall'analisi dell'affidamento al sistema. Il *follow rate* complessivo pari al **37%** evidenzia che i partecipanti non adottano un comportamento di accettazione automatica del suggerimento algoritmico. L'AI viene seguita più frequentemente quando è corretta (**43,8%**) rispetto a quando è errata (**31,0%**), indicando una tendenza a un utilizzo selettivo del supporto. Tuttavia, l'elevata variabilità individuale e la non significatività statistica della differenza intra-soggetto suggeriscono che **l'affidamento all'AI non rappresenta un comportamento uniforme**, ma dipende da strategie decisionali eterogenee tra utenti.

Il quarto *key finding*, e quello empiricamente più robusto, riguarda l'efficienza del processo decisionale. La presenza del supporto AI determina una **riduzione**

**significativa** del tempo medio necessario per la valutazione delle perizie, con un risparmio superiore al **50%** del tempo per decisione. Questo risultato indica che **la funzione principale dell'AI nel contesto analizzato è di natura cognitivo-strumentale**, consentendo ai partecipanti di raggiungere una decisione più rapidamente senza eliminare la variabilità interpretativa individuale.

Considerati nel loro insieme, questi risultati delineano un modello interpretativo unitario dell'interazione uomo-AI. Il supporto algoritmico **non standardizza le decisioni**, ma **ne aumenta l'efficienza e rende gli esiti maggiormente dipendenti dalla qualità dell'informazione algoritmica disponibile**. In altre parole, l'AI non sostituisce il giudizio umano né riduce la discrezionalità tra decisori; piuttosto, essa agisce come un acceleratore decisionale che orienta le scelte mantenendo intatta l'autonomia interpretativa individuale.

Il quadro complessivo suggerisce quindi una trasformazione del processo decisionale più che una sua automatizzazione: i decisori restano eterogenei nei giudizi, ma diventano più rapidi e più sensibili ai segnali forniti dal sistema. L'impatto principale del supporto AI non consiste dunque nella riduzione della discrezionalità, bensì nella **riallocazione del peso cognitivo della decisione**, con un aumento simultaneo di efficienza operativa e dipendenza dalla qualità del suggerimento algoritmico.

Pertanto, l'evidenza empirica identifica quattro risultati centrali:

1. **assenza di riduzione della discrezionalità** (Likert e binaria);
2. **significativo effetto dell'AI sull'accuratezza**, positivo con suggerimento corretto e negativo con suggerimento errato;
3. **affidamento parziale e non automatico** degli utenti, con ampia eterogeneità individuale;
4. **significativa riduzione dei tempi decisionali** in presenza di supporto AI.

Nel loro insieme, tali evidenze indicano che il supporto AI agisce principalmente come **strumento di amplificazione decisionale**, capace di migliorare o peggiorare gli esiti senza eliminare la componente umana di variabilità e giudizio.

Al fine di sintetizzare le evidenze empiriche emerse dalle analisi presentate nelle sezioni precedenti, la Tabella 5.16 riassume i principali risultati in relazione alle research question formulate nel Capitolo 1.

Tabella 5.16: Sintesi delle evidenze empiriche rispetto alle research question

RQ	Evidenza empirica	Risposta alla RQ
RQ1	La discrezionalità media, misurata tramite deviazione standard delle valutazioni, risulta sostanzialmente invariata tra le condizioni con e senza AI. Il test t non evidenzia differenze statisticamente significative.	L'introduzione del supporto AI non modifica in modo significativo la discrezionalità decisionale degli utenti inesperti.
RQ2	Il confronto tra condizione senza AI e condizione con AI corretta mostra un aumento significativo dell'accuratezza ( $\chi^2(1) = 35.62, p < .001$ ). Al contrario, suggerimenti errati riducono significativamente la performance.	L'AI migliora l'accuratezza quando il suggerimento è corretto, mentre suggerimenti errati peggiorano la performance.
RQ3	Il suggerimento AI viene seguito più frequentemente quando è corretto (43.8%) rispetto a quando è errato (31.0%), con un follow rate complessivo del 37.0%.	Gli utenti inesperti mostrano un affidamento selettivo verso il sistema AI, seguendo più spesso i suggerimenti corretti.
RQ4	Il tempo medio di valutazione delle perizie si riduce significativamente con il supporto AI ( $t(14) = 4.05, p = 0.0012$ ), passando da 225.77 a 100.52 secondi per perizia.	Il supporto AI riduce significativamente il tempo necessario per svolgere il compito decisionale.

## 5.9 Analisi degli scostamenti

Un ulteriore elemento interpretativo emerge dall'analisi degli scostamenti decisionali, che consente di osservare non soltanto il livello finale di accuratezza, ma anche la direzione del cambiamento introdotto dal supporto algoritmico rispetto alla decisione autonoma dei partecipanti. Mentre le analisi precedenti mostrano come l'AI influenzi performance ed efficienza, l'esame degli scostamenti permette di comprendere *come* tale influenza si realizzi nel processo decisionale.

I risultati evidenziano che la presenza del suggerimento algoritmico è associata a uno spostamento sistematico delle decisioni individuali. In particolare, quando l'AI fornisce un suggerimento corretto, gli scostamenti tendono a orientarsi verso il ground truth, contribuendo all'incremento dell'accuratezza osservato. Al contrario, in presenza di suggerimenti errati, una quota rilevante di decisioni si sposta nella direzione indicata dall'AI, determinando un peggioramento della performance rispetto alla condizione autonoma.

Questo risultato suggerisce che il supporto AI non agisce primariamente riducendo la discrezionalità tra valutatori, bensì modificando la traiettoria delle decisioni individuali. L'algoritmo introduce infatti un punto di riferimento aggiuntivo che può orientare il giudizio umano, inducendo revisioni della valutazione iniziale indipendentemente dalla correttezza del suggerimento fornito.

Nel complesso, l'analisi degli scostamenti rafforza l'interpretazione emersa dai key findings: **l'AI opera come meccanismo di orientamento decisionale piuttosto che come strumento di standardizzazione**. Gli effetti osservati sull'accuratezza risultano quindi riconducibili agli spostamenti cognitivi indotti dall'interazione uomo-AI, attraverso i quali il suggerimento algoritmico diventa un fattore capace di riallineare, o deviare, il processo decisionale umano.

## **5.10 Sintesi dei risultati empirici**

Le analisi condotte consentono di delineare un quadro complessivo dell'impatto del supporto basato su intelligenza artificiale sul processo decisionale dei partecipanti. Considerando congiuntamente le diverse dimensioni analizzate, emerge che l'introduzione del suggerimento algoritmico produce effetti differenziati sulle varie componenti della decisione.

In primo luogo, i risultati mostrano che la presenza dell'AI non determina una riduzione significativa della discrezionalità tra valutatori. La variabilità dei giudizi rimane sostanzialmente stabile tra le condizioni sperimentali, suggerendo che i partecipanti continuano a interpretare autonomamente le informazioni disponibili anche in presenza di un'indicazione algoritmica.

Diversamente, l'accuratezza decisionale risulta fortemente influenzata dalla qualità del suggerimento fornito: il supporto AI migliora significativamente la performance quando l'indicazione è corretta, mentre produce un peggioramento quando il suggerimento risulta errato. Tale evidenza evidenzia il ruolo centrale dell'informazione algoritmica nell'orientare gli esiti decisionali.

L'analisi dell'affidamento mostra inoltre un comportamento di adesione parziale al sistema: i partecipanti non seguono automaticamente il suggerimento AI, ma ne risultano comunque influenzati in una quota rilevante di decisioni. Parallelamente, il supporto algoritmico determina una significativa riduzione dei tempi necessari alla valutazione delle perizie, indicando un incremento sostanziale dell'efficienza decisionale.

Infine, l'analisi della percezione soggettiva del compito evidenzia che l'esperimento è stato percepito come cognitivamente impegnativo, mentre il suggerimento AI viene chiaramente riconosciuto come elemento di semplificazione del processo valutativo.

Nel complesso, i risultati empirici suggeriscono che il supporto AI non agisce come strumento di standardizzazione delle decisioni, ma come fattore in grado di accelerare il processo decisionale e di orientarne gli esiti, mantenendo tuttavia un

significativo grado di autonomia interpretativa da parte dei decisori.

## Capitolo 6

# Conclusioni e sviluppi futuri

### 6.1 Risultati principali

Il presente lavoro non si configura come una sperimentazione su larga scala finalizzata a produrre evidenze definitive sull'impatto dell'AI nei processi decisionali peritali, bensì come una fase di verifica e validazione del framework sperimentale progettato. L'obiettivo principale è stato quello di testare la coerenza metodologica dell'impianto, l'adeguatezza delle misure individuate e la funzionalità operativa del protocollo, attraverso un primo test pilota condotto su un campione ristretto.

In tale contesto, l'analisi empirica ha riguardato tre dimensioni centrali del processo valutativo (accuratezza delle decisioni, affidamento degli utenti inesperti al suggerimento algoritmico e discrezionalità tra valutatori) non tanto per trarre conclusioni generalizzabili, quanto per verificare la capacità del disegno sperimentale di intercettare tali fenomeni in modo coerente e misurabile.

Con riferimento all'accuratezza, i risultati preliminari indicano che la struttura del test è effettivamente in grado di rilevare variazioni nel comportamento decisionale in funzione della qualità del suggerimento ricevuto. In presenza di suggerimenti corretti si osserva un miglioramento delle performance, mentre suggerimenti errati possono indurre forme di allineamento che incidono negativamente

sull'esito decisionale. Questo dato, pur riferito a un campione limitato, conferma che il framework consente di isolare l'effetto del supporto AI e di distinguere tra condizioni sperimentali differenti.

L'analisi dell'affidamento mostra che la misura adottata è idonea a intercettare il grado di allineamento dei partecipanti al suggerimento algoritmico. I risultati suggeriscono che il supporto AI venga considerato un'informazione rilevante, ma non automaticamente vincolante, evidenziando un'interazione uomo-AI caratterizzata da influenza e non da sostituzione del giudizio umano. Anche in questo caso, l'interesse principale risiede nella validazione dello strumento di misurazione più che nell'entità dell'effetto osservato.

Per quanto riguarda la discrezionalità, le analisi mostrano che la deviazione standard per perizia rappresenta un indicatore operativo adeguato per misurare la convergenza tra valutatori. Nel campione pilota non emerge una riduzione significativa della variabilità tra le condizioni, ma il dato assume valore soprattutto in termini di verifica della sensibilità dell'indicatore adottato.

Nel complesso, il test preliminare conferma la funzionalità del disegno sperimentale, la coerenza delle variabili operative e la possibilità di rilevare effetti differenziati sulle dimensioni analizzate. I risultati devono essere interpretati come evidenze esplorative, utili alla validazione della progettazione e alla preparazione di una successiva implementazione su scala più ampia, che sarà oggetto di sviluppi futuri.

## **6.2 Contributo della ricerca**

Il presente lavoro offre un contributo prevalentemente metodologico, affiancato da implicazioni teoriche ed empiriche preliminari.

Sul piano teorico, la ricerca si inserisce nel dibattito sull'interazione uomo-AI nei processi decisionali complessi, proponendo un contesto sperimentale che si discosta dai task altamente semplificati frequentemente adottati in letteratura. L'impianto

progettato mira a riprodurre una situazione valutativa caratterizzata da pluralità di elementi informativi e margini interpretativi non trascurabili, avvicinandosi maggiormente a scenari applicativi realistici.

Il contributo principale risiede tuttavia nella progettazione e nella prima validazione di un framework sperimentale replicabile. Il disegno bilanciato, la distinzione tra suggerimenti corretti ed errati, l'operazionalizzazione della discrezionalità tramite misure di dispersione e l'utilizzo combinato di variabili Likert e binarie costituiscono elementi metodologici strutturati per consentire analisi articolate nelle successive fasi di ricerca.

Il test pilota ha consentito di verificare la coerenza interna dell'impianto, la comprensibilità delle istruzioni, la funzionalità tecnica della piattaforma e l'adeguatezza delle metriche di misurazione. In questo senso, il contributo empirico non consiste tanto nei risultati in sé, quanto nella dimostrazione che il framework è operativo, misurabile e suscettibile di estensione su campioni più ampi.

La ricerca pone quindi le basi per future sperimentazioni su larga scala, offrendo una struttura metodologica validata che potrà essere ulteriormente raffinata e applicata in contesti più estesi.

### **6.3 Sviluppi futuri**

Il presente lavoro si configura come una fase di progettazione e prima validazione del framework sperimentale, culminata in un test pilota condotto su un campione ristretto. In tale prospettiva, i risultati ottenuti non devono essere interpretati come conclusivi, ma come evidenze preliminari funzionali alla verifica della solidità metodologica dell'impianto e alla preparazione di successive implementazioni su scala più ampia.

In primo luogo, appare opportuno replicare l'esperimento su un campione numericamente più esteso e maggiormente eterogeneo. L'ampliamento della popolazione di riferimento consentirebbe di aumentare la potenza statistica delle analisi e la

generalizzabilità dei risultati, permettendo di stimare con maggiore precisione l'effetto del supporto AI sulle diverse dimensioni considerate, in particolare sulla discrezionalità decisionale. Una fase di raccolta dati su larga scala rappresenta il naturale sviluppo del framework validato in questa sede e potrà essere oggetto di ulteriori attività di ricerca.

In secondo luogo, un aspetto metodologico emerso durante il test pilota riguarda lo strumento utilizzato per la somministrazione, ossia la piattaforma Moodle. Le configurazioni disponibili non hanno consentito una tracciabilità puntuale delle singole domande ai fini dell'analisi dei risultati, limitando in parte la granularità delle informazioni estraibili e la possibilità di associare in modo immediato ciascun item alle relative condizioni sperimentali in fase di esportazione dei dati. In future implementazioni potrebbe risultare opportuno sperimentare strumenti alternativi di somministrazione che garantiscano una maggiore tracciabilità delle risposte a livello di singolo item, una più chiara identificazione delle condizioni sperimentali e una più agevole esportazione dei dati in formato analizzabile. Tale accorgimento permetterebbe di rafforzare ulteriormente la robustezza metodologica del disegno sperimentale.

Un ulteriore sviluppo riguarda il rafforzamento della fase di formazione preliminare. Sebbene siano stati forniti manuali e materiali introduttivi, future implementazioni dovrebbero prevedere momenti formativi più strutturati e interattivi, finalizzati a una più profonda internalizzazione dei criteri valutativi. Un miglioramento della preparazione consentirebbe di distinguere con maggiore chiarezza tra discrezionalità intrinseca al processo interpretativo e variabilità derivante da una comprensione incompleta del compito.

Un'estensione particolarmente rilevante concerne inoltre l'introduzione di ulteriori indicatori comportamentali, quali l'*Indecision Rate*, definito come la percentuale di risposte neutrali (valore 3 nella scala Likert). Tale misura consentirebbe di analizzare il grado di incertezza del giudizio e di distinguere tra effetti di allineamento al suggerimento algoritmico ed effetti di chiarificazione o confusione indotti dal

supporto AI.

Ulteriori sviluppi potrebbero esplorare il ruolo dell'esperienza professionale, confrontando utenti inesperti ed esperti al fine di verificare se l'AI agisca principalmente come strumento compensativo o come amplificatore di competenze già consolidate. Infine, ricerche future potrebbero adottare un disegno longitudinale per valutare se l'interazione ripetuta con il sistema produca effetti di apprendimento nel tempo, modificando progressivamente accuratezza, affidamento e convergenza tra valutatori.

Nel complesso, tali sviluppi consentirebbero di consolidare e ampliare il framework validato in questa sede, trasformando il presente studio pilota in una piattaforma metodologica per analisi successive più estese e approfondite.

# Appendice A

## Manuale per la valutazione di perizie fotografiche in contesto sperimentale

### Sommario

1. Introduzione e finalità del manuale .....	1
2. Che cos'è una perizia: definizione generale .....	2
3. La causa del danno: come riconoscerla .....	2
4. Il danno (effetto): cosa osservare nelle immagini .....	5
5. Lettura del suggerimento AI .....	9
6. La stima economica: cosa rappresenta e come leggerla .....	9
7. Criteri generali per la valutazione dell'indennizzabilità della perizia .....	12

## A.1 Introduzione e finalità del manuale

Il presente manuale ha lo scopo di fornire agli utenti una guida essenziale per l'osservazione e l'interpretazione di immagini peritali acquisite nel corso di perizie assicurative reali relative a **danni da bagnamento in ambito domestico**. Le immagini, nel contesto di una perizia, sono acquisite dal perito al fine di documentare i danni e gli interventi connessi a un sinistro. Esse costituiscono un supporto decisionale fondamentale nella fase di accertamento della compatibilità tra la dinamica dell'evento e le condizioni di polizza, nonché nella valutazione economica dei danni rimborsabili dall'assicurazione al cliente.

Il manuale è rivolto a partecipanti **non esperti**, privi di una formazione tecnica specifica in ambito peritale, che prendono parte a un esperimento di valutazione basato sull'analisi di immagini fotografiche. Durante l'esperimento, agli utenti verranno mostrate serie di immagini relative a singole perizie e sarà richiesto di rispondere a un insieme di affermazioni esprimendo il proprio **grado di accordo** su una scala di valutazione a *5 punti o su scala binaria*.

Nel corso dell'esperimento, alcune perizie verranno presentate senza alcun suggerimento, mentre altre saranno accompagnate da un **suggerimento** fornito da un sistema di supporto basato su intelligenza artificiale (AI). Il suggerimento ha lo scopo di simulare un supporto decisionale, ma **non deve essere considerato necessariamente corretto**: in alcuni casi può essere accurato, in altri parzialmente corretto o non corretto.

Il partecipante non è chiamato a svolgere una perizia tecnica né a formulare valutazioni professionali, ma a osservare attentamente le immagini fornite ed esprimere il proprio giudizio sulla base delle informazioni fornite attraverso il webinar, le vide-pillole e il suddetto manuale. Anche in presenza di un suggerimento AI, il partecipante è invitato a formulare la propria valutazione in modo autonomo, utilizzando il suggerimento come un'informazione aggiuntiva e non come una risposta da accettare automaticamente.

Lo scopo del manuale è fornire le informazioni minime necessarie per rispondere

in modo consapevole e coerente alle domande proposte, supportando l'utente nell'interpretazione delle immagini senza suggerire risposte corrette o giudizi predefiniti. Le indicazioni contenute nel manuale sono pensate per fornire una linea guida allineata alle reali attività peritali in modo da ridurre valutazioni casuali o puramente intuitive, mantenendo al contempo la libertà di giudizio del partecipante.

## A.2 Che cos'è una perizia: definizione generale

Una perizia è un **insieme strutturato di informazioni** che ha lo scopo di descrivere un evento dannoso, individuarne una possibile *causa*, documentarne gli *effetti* visibili e fornire una stima indicativa del danno economico compatibile con le regole della polizza assicurativa stipulata dall'utente

In particolare, una perizia consente di:

- definire e certificare lo scenario del sinistro;
- individuare l'origine del danno;
- osservare le conseguenze materiali dell'evento (i danni visibili);
- valutare l'entità economica del danno.

Le immagini fotografiche rappresentano un elemento centrale della perizia, in quanto consentono di osservare direttamente i danni e di collocarli spazialmente, ad esempio all'interno di un ambiente o di un'abitazione. Attraverso l'analisi delle immagini è possibile individuare sia le manifestazioni del danno sia una serie di evidenze visive utili a verificare la coerenza tra la causa ipotizzata del sinistro e gli effetti visivamente riscontrabili.

Nel contesto di questo esperimento, all'utente non è richiesto di analizzare una perizia in modo tecnico o professionale. L'attenzione è invece rivolta alla capacità di:

- osservare le immagini fornite;

- riconoscere quali elementi sono riconducibili alla possibile causa del danno (vedi capitolo 3);
- riconoscere le manifestazioni visibili del danno stesso (vedi capitolo 4);
- stimare il valore economico dell'evento (vedi capitolo 6);
- esprimere il proprio giudizio utilizzando le scale di valutazione proposte.

### A.3 La causa del danno: come riconoscerla

Nel contesto di questo manuale, per causa del danno si intende l'**origine fisica dell'evento** che ha generato il danno documentato nelle immagini peritali. Nelle perizie considerate nell'esperimento, la causa è riconducibile a fenomeni legati alla fuoriuscita di acqua in ambienti domestici, generalmente dovuti alla rottura o al danneggiamento di componenti dell'impianto idrico.

La causa del danno è sempre documentata all'interno della perizia attraverso le immagini fornite. Tuttavia, essa può essere mostrata in **momenti diversi rispetto al verificarsi dell'evento**: in alcuni casi la causa è visibile direttamente, mentre in altri casi le immagini possono documentarla dopo un intervento di riparazione già avvenuto oppure in una fase successiva di accertamento.

Le principali tipologie di causa che possono essere riscontrate nelle immagini sono le seguenti:

- rottura di *tubazioni di carico (tubi in pressione)* dell'acqua (vedi figure A1);
- rottura di *tubazioni di scarico* (vedi figure A2.);
- perdite d'acqua dovute a *guarnizioni, giunti o raccordi* (vedi figure A3.);
- rottura o danneggiamento di componenti dell'impianto come *cassette WC o pilette* (vedi figura A4);

- cause non immediatamente visibili, documentate solo in una fase successiva o a seguito di ulteriori interventi.

NB: Nel caso di guarnizioni, giunti o raccordi il danno deriva da una perdita di tenuta; nel caso di componenti dell'impianto il danno deriva dalla rottura del componente stesso.

Nel caso di tubi è possibile distinguere quelli di carico da quelli di scarico in base al loro diametro e colore:

- diametro più piccolo + colore bianco/materiale ferroso tubo di carico
- diametro più ampio + colore nero/colore grigio/ colore arancione tubo di scarico

Durante l'osservazione delle immagini, il partecipante è invitato a riconoscere quale tipologia di causa è rappresentata, basandosi esclusivamente sulle informazioni visive disponibili. Non è richiesto distinguere aspetti tecnici o costruttivi dell'impianto, ma identificare la categoria di causa mostrata nelle immagini.

Qualora la causa sia documentata in modo parziale o successivo alla riparazione, il partecipante è comunque invitato a esprimere il proprio giudizio sulla base di quanto osservabile, utilizzando le scale di valutazione proposte.



(a) Rottura tubo carico acqua



(b) Rottura tubo scarico doccia

Figura A.1: Esempi rottura tubo carico e rottura tubo scarico



(a) Rottura tubo scarico wc



(b) Rottura colonna di scarico condominiale

Figura A.2: Esempi rottura tubo scarico



(a) Rottura guarnizione



(b) Rottura canotto wc

Figura A.3: Esempi di rottura guarnizioni e canotto

## A.4 Il danno (effetto): cosa osservare nelle immagini

Nel contesto di questo manuale, per danno (o effetto) si intende la **conseguenza visibile dell'evento dannoso**, osservabile direttamente nelle immagini peritali. Il danno rappresenta ciò che è stato materialmente compromesso a seguito dell'evento, indipendentemente dalla causa che lo ha generato.

A differenza della causa, che riguarda l'origine del problema, il danno è sempre visibile nelle immagini fornite. Tuttavia, il suo grado di evidenza può variare: in alcuni casi le manifestazioni del danno risultano chiare ed immediatamente riconoscibili, mentre in altri possono presentarsi in modo meno evidente, parziale o ambiguo. Le fotografie mostrano le superfici, i materiali o gli elementi dell'edificio che hanno subito un'alterazione o un deterioramento, con livelli differenti di chiarezza visiva.

Tra le principali manifestazioni di danno che possono essere osservate nelle immagini rientrano, ad esempio:

- macchie o muffe sulle superfici (vedi figure A4-A5.);
- *segni di bagnamento o umidità* (vedi figure A5-A6.);
- *rigonfiamenti, distacchi o scrostamenti* di pavimenti, pareti o soffitti (vedi figure A6-A7.);
- danneggiamenti a finiture o rivestimenti (*vedi figura A7-A8*).

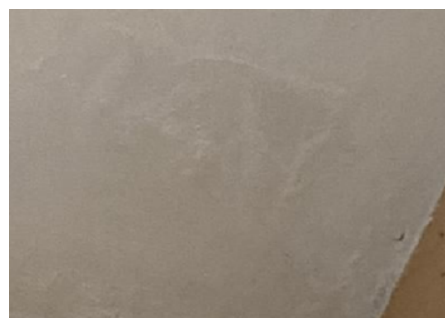
Durante l'osservazione delle immagini, il partecipante è invitato a concentrarsi su ciò che è direttamente visibile, senza formulare ipotesi su elementi non mostrati. Il giudizio richiesto riguarda esclusivamente l'entità e le caratteristiche del danno osservabile, così come appare nelle immagini.

In presenza di danni chiaramente evidenti o, al contrario, meno immediatamente interpretabili, il partecipante è comunque tenuto a esprimere il proprio giudizio valutativo facendo riferimento alle informazioni fornite nel presente manuale e alle proprie considerazioni, sulla base delle evidenze visive disponibili.

Qualora il danno risulti limitato, esteso o distribuito su più superfici o ambienti, il partecipante è invitato a tenerne conto nella valutazione, utilizzando le scale di valutazione proposte.



(a) Macchia



(b) Muffa Chiara

Figura A.4: Esempio macchia e muffa chiara



(a) Muffa scura



(b) Gocciolamento

Figura A.5: Esempio muffa scura e gocciolamento



(a) Perdita d'acqua



(b) Rialzamento

Figura A.6: Esempio perdita d'acqua e rialzamento



(a) Scrostamento 1



(b) Scrostamento 2

Figura A.7: Esempi di scrostamenti

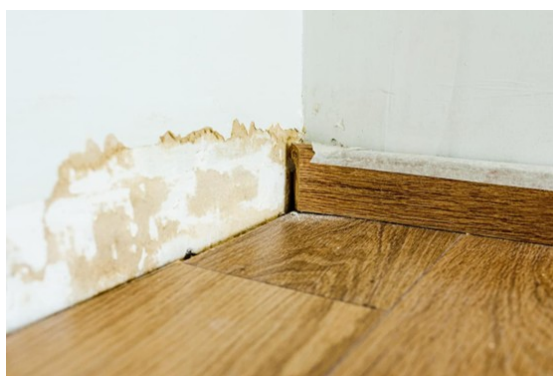


Figura A.8: Danneggiamento finiture

## A.5 Lettura del suggerimento AI

Nel corso dell'esperimento, in alcune fasi, insieme alle immagini peritali potrà essere fornito un suggerimento generato da un sistema di supporto basato su intelligenza artificiale (AI). Il suggerimento viene presentato **in forma testuale** e sintetizza l'output del sistema in relazione al caso peritale analizzato.

Il suggerimento AI rappresenta un'informazione aggiuntiva a supporto dell'osservazione e dell'analisi, ma **non costituisce una risposta prescrittiva** né una soluzione automatica del caso. Il suo contenuto può riguardare, ad esempio, una possibile interpretazione della dinamica del danno o un'indicazione generale sulla compatibilità del caso con determinate tipologie di evento.

È importante sottolineare che i suggerimenti forniti dal sistema AI **non sono necessariamente corretti**: in alcuni casi possono risultare accurati, in altri parzialmente corretti o non corretti. Per questo motivo, il partecipante è invitato a utilizzare il suggerimento AI in modo critico, integrandolo – se ritenuto opportuno – con le informazioni ricavabili dalle immagini peritali e dalle indicazioni contenute nel presente manuale.

Il giudizio finale rimane interamente a carico del partecipante, che è chiamato a formulare la propria valutazione in maniera autonoma, senza fare affidamento automatico sul suggerimento fornito dal sistema di supporto AI.

## A.6 La stima economica: cosa rappresenta e come leggerla

La stima economica effettuata dal perito rappresenta una valutazione dell'impatto economico complessivo dell'evento dannoso osservato nelle immagini. Tale stima comprende sia i **costi** legati alla **riparazione della causa** dell'evento, sia i costi necessari per il **ripristino dei danni** visibili sulle superfici. La determinazione dei danni è fondamentale per decidere quale cifra economica è corretto rimborsare al cliente assicurativo.

Nel contesto di questo esperimento **non è richiesto stimare un importo preciso**, né effettuare calcoli dettagliati o utilizzare competenze tecniche specifiche. In alcune domande verrà proposto un intervallo di valori economici, che rappresentano un'ipotesi di inquadramento del costo complessivo. Il compito del partecipante consiste nel valutare quanto tale **intervallo risulti adeguato**, esprimendo il proprio grado di accordo sulla scala di valutazione fornita.

Per supportare questa valutazione, nel manuale sono fornite due tabelle di riferimento: una relativa alle cause e una relativa ai danni.

La **tabella delle cause** (vedi tabella A1) riporta le principali tipologie di causa dell'evento (ad esempio rottura di tubazioni, componenti dell'impianto, guarnizioni) e un costo indicativo di ricerca e ripristino della causa stessa. I valori riportati hanno una funzione orientativa e servono a comprendere quanto possa essere complesso e oneroso intervenire sulla causa, anche in relazione alla modalità di accesso e al tipo di impianto coinvolto.

La **tabella dei danni** (vedi tabella A2) riporta invece, per ciascun tipo di superficie o elemento coinvolto (ad esempio muro, pavimento, soffitto, zoccolo), i principali tipi di danno osservabili (come macchie, muffa, scrostamenti, rialzamenti) e un costo indicativo di ripristino, espresso in relazione a una determinata estensione (ad esempio per un certo numero di metri quadrati o metri lineari). Questa tabella consente di orientarsi nel valutare quanto il danno osservato possa incidere

economicamente, tenendo conto sia della tipologia di danno sia della superficie interessata.

Nel formulare il proprio giudizio sulla stima economica complessiva, il partecipante è invitato a considerare in modo integrato:

- la **complessità dell'intervento sulla causa**, ad esempio se l'intervento appare semplice o richiede operazioni più invasive (come lavori sottotraccia);
- l'**estensione dei danni**, ossia se il danno risulta limitato a una piccola area o diffuso su più superfici o ambienti;
- le **superfici o finiture coinvolte**, alcune delle quali comportano interventi di ripristino più complessi o costosi.

Le tabelle **NON** devono essere utilizzate per calcolare un **valore esatto**, ma come strumenti di supporto per valutare la plausibilità dell'intervallo economico proposto, mettendo in relazione ciò che è visibile nelle immagini con l'ordine di grandezza dei costi riportati.

Anche in presenza di un suggerimento fornito dal sistema di supporto basato su intelligenza artificiale (AI), l'intervallo economico proposto **NON** deve essere considerato **automaticamente corretto**.

Il partecipante è invitato a utilizzare tutte le informazioni disponibili – immagini, manuale e tabelle – come supporto al proprio giudizio, che deve rimanere **autonomo e critico**.

Dettaglio causa	Costo totale Ricerca e Ripristino [€]
Rottura braga	600–750
Rottura tubo carico	650–700
Rottura tubo scarico acque grigie bidet/lavelli (immurate)	350–400
Rottura tubo scarico acque grigie doccia	750 (+200 se vasca da sostituire)
Rottura tubo scarico acque grigie vasca	600 (+200 manodopera +200 piatto doccia base)
Rottura tubo scarico acque nere (indoor, rottura tratto privato)	500–600
Rottura tubo scarico acque nere (indoor, tra più piani)	1000–1200 (quota parte se condominio)
Rottura tubo riscaldamento	650–750
Rottura tubo riscaldamento a pavimento	1100–1200
Rottura cassetta WC incasso	600–650
Rottura cassetta WC esterna	300–400
Rottura guarnizione	60
Rottura saracinesca bagno (valvole singole sanitari)	150–200

Tabella A.1: Costi indicativi di ricerca e ripristino per le principali cause di danno idrico

Tabella A.2: Tabella dei danni

Elemento	Finitura	Tipo di danno	Costo indicativo
Muro	Intonaco	Macchia / Gocciolamento	200–300 €/20 mq
Muro	Intonaco	Muffa	50 €/10 mq
Muro	Piastrelle	Gocciolamento	250–300 €/10 mq
Pavimento	Parquet	Rialzamento	300–600 €/5 mq
Zoccolo	Legno	Rialzamento	100 €/10 ml

## A.7 Criteri generali per la valutazione dell'indennizzabilità della perizia

Ai fini di questo studio, al partecipante è richiesto di esprimere un giudizio binario (**sì/no**) sull'indennizzabilità della perizia, basandosi esclusivamente sulle informazioni fornite nel presente manuale e sulle evidenze visibili nelle immagini

peritali. La valutazione non richiede competenze tecniche specialistiche, ma si fonda sull'osservazione di alcuni criteri generali, comunemente utilizzati nella pratica peritale.

In primo luogo, affinché una perizia possa essere considerata indennizzabile, **deve essere presente un danno visibile**. Il danno si manifesta attraverso segni osservabili nelle immagini (vedi capitolo 4), quali macchie su pareti o soffitti, scrostamenti dell'intonaco, pavimentazioni rialzate o deformate, spargimento o perdita di liquidi, oppure la presenza di lavorazioni edili o idrauliche in corso (ad esempio pareti o pavimenti demoliti per accedere a una tubazione). In assenza di un danno visibile, la perizia non è indennizzabile.

Il danno visibile deve inoltre essere **riconcucibile alla fuoriuscita di acqua o al malfunzionamento di un impianto idrico**. In questa categoria rientrano, ad esempio, la rottura accidentale di tubazioni, la fuoriuscita d'acqua da elettrodomestici collegati all'impianto idrico, il gelo delle tubazioni o, se previsto dalla garanzia, l'occlusione o il rigurgito degli impianti. Il partecipante è chiamato a valutare se, sulla base delle immagini fornite, il danno osservato possa essere plausibilmente ricondotto a una di queste dinamiche.

Un ulteriore criterio fondamentale riguarda la **manutenzione dell'immobile**. Il danno è indennizzabile solo se non risulta chiaramente riconducibile a cattiva manutenzione o a un deterioramento prolungato nel tempo. Segni quali guarnizioni usurate, sigillature ammalorate o muffa diffusa e persistente sono generalmente indicativi di problemi preesistenti e non danno luogo a indennizzo.

È inoltre necessario che vi sia **coerenza tra la causa dichiarata e il danno osservabile**. La tipologia, la posizione e l'estensione del danno devono essere compatibili con la causa indicata. Ad esempio, la localizzazione delle macchie, il loro colore o la presenza di interventi di riparazione devono risultare plausibili rispetto al tipo di perdita o rottura ipotizzata. In caso di incoerenza evidente tra causa e danno, la perizia non è considerata indennizzabile.

Infine, sono indennizzabili solo le **spese direttamente conseguenti al danno**,

ossia quelle strettamente necessarie per individuare la causa, riparare la tubazione danneggiata e ripristinare le parti effettivamente compromesse. Non rientrano nell'indennizzo interventi migliorativi, lavori estesi non giustificati dal danno osservato o rifacimenti completi non supportati dalle evidenze visive.

In sintesi, la perizia può essere considerata indennizzabile se sono contemporaneamente soddisfatte tutte le seguenti condizioni: presenza di un danno visibile, riconducibilità del danno alla fuoriuscita di acqua o al malfunzionamento di un impianto idrico, assenza di cattiva manutenzione, coerenza tra causa ed effetti e proporzionalità degli interventi rispetto al danno. In mancanza anche di una sola di queste condizioni, la risposta corretta è “no, la perizia non è indennizzabile”.

# Appendice B

## Tabelle risultati

### Struttura delle tabelle e codifica delle variabili

In questa appendice sono riportate le tabelle riorganizzate e riformattate al fine di garantire una maggiore chiarezza nella lettura dei risultati relativi ai tre test somministrati tramite piattaforma Moodle, nelle diverse condizioni sperimentali (con e senza supporto AI).

La variabile di condizione è codificata come segue: la condizione *senza AI* è identificata con valore 0, mentre la condizione *con AI* è identificata con valore 1.

È inoltre riportata una tabella esplicativa utile all'interpretazione delle risposte su scala Likert (valori da 1 a 5) relative alle prime tre domande di ciascuna perizia. Tali item riguardano rispettivamente: (i) l'identificazione della causa del danno, (ii) l'identificazione dell'effetto/danno osservabile e (iii) la valutazione della stima economica.

L'ultima domanda di ciascuna perizia, relativa all'indennizzabilità, è invece espressa in forma binaria: il valore 0 indica una risposta non corretta, mentre il valore 1 indica una risposta corretta.

Infine, è inclusa una tabella di riferimento contenente, per ciascuna perizia e per ciascun item, la classificazione delle risposte corrette (*ground truth*), utilizzata

come base per il calcolo delle metriche di accuratezza e per le analisi comparative tra condizioni.

Le perizie sono identificate tramite un codice progressivo da P01 a P12 (per un totale di 12 perizie), mentre le domande sono codificate da Q01 a Q04. In particolare:

- **Q01:** Identificazione della causa.
- **Q02:** Identificazione del danno/effetto.
- **Q03:** Valutazione della stima economica.
- **Q04:** Indennizzabilità (variabile binaria).

<b>Legenda risposte</b>					
<b>Risposte</b>	Totalmente in disaccordo	Più o meno in disaccordo	Incerto	Più o meno d'accordo	Totalmente d'accordo
<b>Punteggio associato</b>	1	2	3	4	5

Tabella B.1: Legenda delle risposte

Perizie	Domande	Risposte	Perizie	Domande	Risposte
P01	Q01	4, 5	P07	Q01	1, 2
P01	Q02	4, 5	P07	Q02	1, 2
P01	Q03	4, 5	P07	Q03	4, 5
P01	Q04	4, 5	P07	Q04	4, 5
P02	Q01	4, 5	P08	Q01	1, 2
P02	Q02	4, 5	P08	Q02	1, 2
P02	Q03	4, 5	P08	Q03	4, 5
P02	Q04	4, 5	P08	Q04	4, 5
P03	Q01	4, 5	P09	Q01	1, 2
P03	Q02	4, 5	P09	Q02	1, 2
P03	Q03	4, 5	P09	Q03	4, 5
P03	Q04	1, 2	P09	Q04	4, 5
P04	Q01	4, 5	P10	Q01	1, 2
P04	Q02	4, 5	P10	Q02	1, 2
P04	Q03	4, 5	P10	Q03	4, 5
P04	Q04	4, 5	P10	Q04	4, 5
P05	Q01	4, 5	P11	Q01	1, 2
P05	Q02	4, 5	P11	Q02	1, 2
P05	Q03	4, 5	P11	Q03	4, 5
P05	Q04	4, 5	P11	Q04	4, 5
P06	Q01	4, 5	P12	Q01	1, 2
P06	Q02	4, 5	P12	Q02	1, 2
P06	Q03	4, 5	P12	Q03	4, 5
P06	Q04	4, 5	P12	Q04	4, 5

Tabella B.2: Risposte alle domande per ciascuna perizia

## B.1 TEST A SENZA AI

ID	Test	Cond.	Tempo	P01				P02				P03				P04			
				Q01	Q02	Q03	Q04	Q01	Q02	Q03	Q04	Q01	Q02	Q03	Q04	Q01	Q02	Q03	Q04
s362605	A	0	18m55s	5	3	4	0	2	1	1	1	2	5	4	0	5	4	1	1
s339561	A	0	10m43s	4	3	4	1	4	5	2	1	4	4	4	0	5	5	3	1
s305000	A	0	17m46s	3	4	3	0	5	4	2	1	4	5	5	1	3	5	3	0
s333191	A	0	10m04s	1	5	4	1	5	2	4	1	5	5	5	1	5	1	1	0
s348924	A	0	42m18s	3	4	4	0	5	3	1	0	1	5	4	1	5	5	1	1
s331777	A	0	16m02s	4	3	2	1	4	4	2	1	2	5	3	1	4	5	2	1

Tabella B.3: Valutazioni dei partecipanti nel Test A (condizione 0) per le perizie P01–P04.

## B.2 TEST A CON AI

ID	Test	Cond.	Tempo	P05				P06				P07				P08			
				Q01	Q02	Q03	Q04	Q01	Q02	Q03	Q04	Q01	Q02	Q03	Q04	Q01	Q02	Q03	Q04
s339561	A	1	12m30s	5	4	2	1	3	5	2	1	3	2	1	1	5	5	4	1
s362605	A	1	12m01s	3	5	1	1	4	5	2	1	2	1	1	1	2	1	1	0
s333191	A	1	16m37s	5	1	1	0	5	5	1	1	1	1	0	4	4	1	1	
s305000	A	1	20m20s	4	5	2	1	2	5	2	1	2	2	1	0	4	2	2	1
s331777	A	1	15m49s	4	5	2	1	3	5	3	0	3	1	1	0	4	5	2	1
s348924	A	1	20m12s	5	5	2	1	3	5	3	0	3	1	4	0	5	1	4	0

Tabella B.4: Valutazioni dei partecipanti nel Test A (condizione 1) per le perizie P05–P08.

ID	Test	Cond.	Tempo	P09				P10				P11				P12			
				Q01	Q02	Q03	Q04	Q01	Q02	Q03	Q04	Q01	Q02	Q03	Q04	Q01	Q02	Q03	Q04
s339561	A	1	12m30s	4	4	3	1	3	5	1	0	2	2	4	1	4	4	4	0
s362605	A	1	12m01s	1	1	2	1	2	5	1	0	2	4	4	0	5	5	1	1
s333191	A	1	16m37s	1	5	1	0	5	1	1	0	1	5	4	1	5	2	1	0
s305000	A	1	20m20s	4	4	4	1	4	5	1	1	2	1	3	0	4	3	3	1
s331777	A	1	15m49s	3	1	4	0	3	5	1	1	3	5	4	0	1	5	4	1
s348924	A	1	20m12s	1	1	4	0	3	1	1	0	1	1	1	0	4	5	4	1

Tabella B.5: Valutazioni dei partecipanti nel Test A (condizione 1) per le perizie P09–P12.

## B.3 TEST B SENZA AI

ID	Test	Cond.	Tempo	P06				P07				P08				P09			
				Q01	Q02	Q03	Q04	Q01	Q02	Q03	Q04	Q01	Q02	Q03	Q04	Q01	Q02	Q03	Q04
s330444	B	0	9m27s	3	4	4	1	5	5	3	1	3	1	1	0	3	2	4	0
s317881	B	0	9m32s	5	5	2	1	4	4	4	0	4	2	1	1	5	2	3	1
s334261	B	0	9m15s	4	5	3	1	5	2	2	1	2	1	2	0	5	2	4	1
s333878	B	0	20m35s	5	5	1	0	5	5	5	1	2	2	4	0	4	2	2	1
s337919	B	0	8m42s	4	5	1	1	1	5	2	1	2	1	3	0	2	4	1	1

Tabella B.6: Valutazioni dei partecipanti nel Test B (condizione 0) per le perizie P06–P09.

## B.4 TEST B CON AI

ID	Test	Cond.	Tempo	P01				P02				P03				P04			
				Q01	Q02	Q03	Q04	Q01	Q02	Q03	Q04	Q01	Q02	Q03	Q04	Q01	Q02	Q03	Q04
s330444	B	1	15m	3	1	4	1	4	3	2	0	5	5	4	0	4	3	4	0
s317881	B	1	12m19s	5	5	2	1	5	4	1	1	5	5	4	1	5	4	1	1
s334261	B	1	11m31s	4	1	4	1	3	4	1	1	4	4	2	0	3	4	2	0
s337919	B	1	9m33s	4	3	3	0	5	3	1	1	5	5	3	1	3	5	2	1
s333878	B	1	24m14s	5	2	2	0	1	1	4	1	5	2	2	0	5	5	1	1

Tabella B.7: Valutazioni dei partecipanti nel Test B (condizione 1) per le perizie P01–P04.

Tabelle risultati

ID	Test	Cond.	Tempo	P09				P10				P11				P12			
				Q01	Q02	Q03	Q04	Q01	Q02	Q03	Q04	Q01	Q02	Q03	Q04	Q01	Q02	Q03	Q04
s330444	B	1	15m	2	4	4	1	2	5	1	1	2	2	4	0	3	5	4	1
s317881	B	1	12m19s	5	2	4	1	5	5	1	1	1	1	2	1	1	5	2	0
s334261	B	1	11m31s	1	4	2	1	1	4	1	0	1	2	4	1	1	4	1	1
s337919	B	1	9m33s	3	1	2	1	1	5	1	1	1	4	4	1	1	5	1	1
s333878	B	1	24m14s	1	1	2	1	1	2	1	1	1	2	5	1	1	2	3	1

Tabella B.8: Valutazioni dei partecipanti nel Test B (condizione 1) per le perizie P09–P12.

## B.5 TEST C SENZA AI

ID	Test	Cond.	Tempo	P09				P10				P11				P12			
				Q01	Q02	Q03	Q04	Q01	Q02	Q03	Q04	Q01	Q02	Q03	Q04	Q01	Q02	Q03	Q04
s348936	C	0	10m16s	2	4	2	1	2	4	3	1	4	4	4	1	2	4	3	1
s334187	C	0	4m35s	1	1	2	1	3	5	4	0	2	1	2	0	4	2	4	1
s329886	C	0	12m9s	1	1	3	1	1	1	5	0	1	3	1	1	3	3	1	1
s328182	C	0	2h11m	1	1	2	1	3	1	4	0	3	2	4	0	3	5	4	0
s323015	C	0	25m27s	1	2	2	1	5	5	1	1	4	4	4	1	3	5	3	0
s333683	C	0	2h	5	1	4	1	4	3	1	0	4	1	4	1	1	4	4	1

Tabella B.9: Valutazioni dei partecipanti nel Test C (condizione 0) per le perizie P09–P12.

## B.6 TEST C CON AI

ID	Test	Cond.	Tempo	P01				P02				P03				P04			
				Q01	Q02	Q03	Q04	Q01	Q02	Q03	Q04	Q01	Q02	Q03	Q04	Q01	Q02	Q03	Q04
s334187	C	1	4m10s	5	4	4	1	4	5	1	1	4	4	4	1	5	5	3	1
s348936	C	1	3m35s	4	4	2	1	4	4	2	1	4	4	3	1	2	4	2	1
s329886	C	1	8m50s	5	4	3	1	3	3	2	1	4	5	4	0	3	5	4	1
s328182	C	1	2h1m	4	5	2	1	4	5	4	1	5	5	4	1	5	5	2	1
s333683	C	1	15m25s	5	5	2	1	5	1	1	1	5	5	1	1	5	5	2	1
s323015	C	1	14m21s	4	3	4	1	4	4	2	1	5	5	1	1	5	5	1	1

Tabella B.10: Valutazioni dei partecipanti nel Test C (condizione 1) per le perizie P01–P04.

ID	Test	Cond.	Tempo	P05				P06				P07				P08			
				Q01	Q02	Q03	Q04	Q01	Q02	Q03	Q04	Q01	Q02	Q03	Q04	Q01	Q02	Q03	Q04
s334187	C	1	4m10s	1	3	4	1	2	3	2	1	5	1	1	1	4	2	4	1
s348936	C	1	3m35s	2	4	2	1	2	4	2	1	4	2	1	1	4	4	4	1
s329886	C	1	8m50s	4	5	2	1	1	5	4	1	3	2	2	1	4	3	4	1
s328182	C	1	2h1m	2	5	2	1	3	5	2	0	3	4	1	1	4	2	4	1
s333683	C	1	15m25s	5	5	2	1	2	5	2	1	5	3	1	1	1	5	1	1
s323015	C	1	14m21s	4	5	1	1	2	4	2	1	4	3	1	1	5	4	1	1

Tabella B.11: Valutazioni dei partecipanti nel Test C (condizione 1) per le perizie P05–P08.

# Bibliografia

- [1] Joshua Newman, Michael Mintrom e Deirdre O’Neill. «Digital technologies, artificial intelligence, and bureaucratic transformation». In: *Futures* 136 (2022), p. 102886 (cit. alle pp. 3, 5).
- [2] Max Weber. *Scritti politici*. Donzelli Editore, 1998 (cit. a p. 4).
- [3] Justin B Bullock. «Artificial intelligence, discretion, and bureaucracy». In: *The American Review of Public Administration* 49.7 (2019), pp. 751–761 (cit. a p. 5).
- [4] Justin B Bullock, Hsini Huang e Kyoung-Cheol Kim. «Machine intelligence, bureaucracy, and human control». In: *Perspectives on Public Management and Governance* 5.2 (2022), pp. 187–196 (cit. a p. 5).
- [5] Shakked Noy e Whitney Zhang. «Experimental evidence on the productivity effects of generative artificial intelligence». In: *Science* 381.6654 (2023), pp. 187–192 (cit. alle pp. 17, 34).
- [6] Kyogo Kanazawa, Daiji Kawaguchi, Hitoshi Shigeoka e Yasutora Watanabe. «AI, skill, and productivity: The case of taxi drivers». In: *Management Science* (2025) (cit. a p. 17).
- [7] Jonathan H Choi e Daniel Schwarcz. «Ai assistance in legal analysis: An empirical study». In: *J. Legal Educ.* 73 (2024), p. 384 (cit. a p. 18).

- [8] Fabrizio Dell’Acqua, Edward McFowland III, Ethan R Mollick, Hila Lifshitz-Assaf, Katherine Kellogg, Saran Rajendran, Lisa Krayner, François Candelon e Karim R Lakhani. «Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality». In: *Harvard Business School Technology & Operations Mgt. Unit Working Paper 24-013* (2023) (cit. alle pp. 18, 24, 31).
- [9] Julian Senoner, Simon Schallmoser, Bernhard Kratzwald, Stefan Feuerriegel e Torbjørn Netland. «Explainable AI improves task performance in human–AI collaboration». In: *Scientific reports* 14.1 (2024), p. 31150 (cit. a p. 21).
- [10] Sarah Lebovitz, Hila Lifshitz-Assaf e Natalia Levina. «To engage or not to engage with AI for critical judgments: How professionals deal with opacity when using AI for medical diagnosis». In: *Organization science* 33.1 (2022), pp. 126–148 (cit. alle pp. 21, 31).
- [11] Daan Kolkman, Floris Bex, Nitin Narayan e Manuella van der Put. «Justitia ex machina: The impact of an AI system on legal decision-making and discretionary authority». In: *Big Data & Society* 11.2 (2024), p. 20539517241255101 (cit. alle pp. 22, 27, 33, 75, 82, 87, 93, 99).
- [12] Anoop Mishra, Gopinath Gangiseti e Deepak Khazanchi. «An investigation into the advancements of edge-AI capabilities for structural health monitoring». In: *IEEE Access* 12 (2024), pp. 25325–25345 (cit. alle pp. 23, 27, 39, 53).
- [13] Andres Campero, Michelle Vaccaro, Jaeyoon Song, Haoran Wen, Abdullah Almaatouq e Thomas W Malone. «A test for evaluating performance in human-computer systems». In: *arXiv preprint arXiv:2206.12390* (2022) (cit. alle pp. 24, 31).
- [14] Sida Peng, Eirini Kalliamvakou, Peter Cihon e Mert Demirer. «The impact of ai on developer productivity: Evidence from github copilot». In: *arXiv preprint arXiv:2302.06590* (2023) (cit. alle pp. 24, 31, 34).

- [15] Erik Brynjolfsson, Danielle Li e Lindsey Raymond. «Generative AI at work». In: *The Quarterly Journal of Economics* 140.2 (2025), pp. 889–942 (cit. alle pp. 25, 34).
- [16] Ekaterina Jussupow, Kai Spohrer, Armin Heinzl e Joshua Gawlitza. «Augmenting medical diagnosis decisions? An investigation into physicians’ decision-making process with artificial intelligence». In: *Information Systems Research* 32.3 (2021), pp. 713–735 (cit. alle pp. 25, 32, 51, 54, 87).
- [17] Henry Mintzberg. «The structuring of organizations». In: *Readings in strategic management*. Springer, 1979, pp. 322–352 (cit. a p. 26).
- [18] Andrew Caplin, DJ Deming, Shangwen Li, DJ Martin, Philip Marx, Ben Weidmann e KJ Ye. *The abc’s of who benefits from working with ai: Ability, beliefs*. Rapp. tecn. and calibration. Technical report, National Bureau of Economic Research, 2024 (cit. a p. 31).
- [19] Andrés Suárez-García, María Álvarez-Hernández, Elena Arce e José Roberto Ribas. «Exploring the efficacy of binary surveys versus likert scales in assessing student perspectives using bayesian analysis». In: *Applied Sciences* 14.10 (2024), p. 4189 (cit. alle pp. 33, 53).
- [20] Emma Benz Wiles. «Artificial Intelligence in Labor Market Matching». Tesi di dott. Massachusetts Institute of Technology, 2024 (cit. a p. 34).
- [21] Eli Ben-Michael, D James Greiner, Melody Huang, Kosuke Imai, Zhichao Jiang e Sooahn Shin. «Does AI help humans make better decisions? A statistical evaluation framework for experimental and observational studies». In: *Proceedings of the National Academy of Sciences* 122.38 (2025), e2505106122 (cit. a p. 55).
- [22] Howard T Tokunaga. *Fundamental statistics for the social and behavioral sciences*. Sage Publications, 2018 (cit. a p. 56).