



UNIVERSITÀ DEGLI STUDI DI SALERNO

Dipartimento di Informatica

Corso di Laurea in Informatica

TESI DI LAUREA

Generative AI e Requirements Engineering: dalla Didattica Accademica alla Produttività Aziendale

RELATORI

Prof.ssa Filomena Ferrucci

Dott. Gabriele De Vito

Università degli Studi di Salerno

CANDIDATA

Michela Palmieri

Matricola: 0512116768

Anno Accademico 2024-2025

Questa tesi è stata realizzata nel

sesa^{lab}
SOFTWARE ENGINEERING
SALERNO

*Ventosa d'inverno,
Rovente d'estate,
Perfetta nelle mezze stagioni,
UniSa...*

*Le tue aule mi hanno vista crescere,
Lezione dopo lezione ho consumato prime file,
Risate e interi pacchetti di slides...
Ed esame dopo esame sono diventata più forte, più capace.
L'informatica non mi è stata mai così tanto amica, prima di conoscerti.
Perché UniSa si scopre ogni giorno di più,
Osservandola, ogni volta, da uno scorcio verde diverso,
E non c'è niente di più bello che riscoprire nella propria Università,
Un angolo di mondo in cui potersi sentire accolti.
A te UniSa, che mi hai cullata durante questi tre bellissimi anni.*

A me, che ho trovato in te:

...La mia chiave.

Abstract

Negli ultimi anni l'introduzione della Generative Artificial Intelligence (GenAI) ha trasformato numerosi settori dell'informatica, e l'Ingegneria dei Requisiti non fa eccezione. In particolare, le fasi di requirements elicitation e requirements analysis stanno sperimentando l'uso di modelli di linguaggio di grandi dimensioni (LLM) e di strumenti intelligenti in grado di supportare l'identificazione, la riformulazione e la validazione dei requisiti. Tuttavia, la letteratura scientifica e i report industriali mostrano un panorama ancora frammentato: se da un lato emergono potenziali benefici in termini di riduzione dei tempi e miglioramento della qualità, dall'altro permangono dubbi legati a riproducibilità, affidabilità e gestione dei bias.

Questa tesi propone una **Systematic Multivocal Literature Review (SMLR)** che integra fonti accademiche *peer-reviewed* e contributi di *grey literature*, adottando come riferimento metodologico principale le linee guida di **Garousi et al. (2019)** [1]. L'obiettivo è analizzare in che misura la GenAI possa supportare le attività di Requirements Engineering, con particolare attenzione agli aspetti di produttività e qualità. L'obiettivo è duplice: da un lato comprendere in che misura le aziende possano trarre benefici concreti dall'adozione di tali strumenti, dall'altro delineare implicazioni per la formazione accademica, suggerendo come il corso di Ingegneria del Software possa integrare in modo efficace le nuove tecnologie.

Il lavoro delinea un percorso che collega la didattica accademica allo sviluppo di competenze sulla GenAI, mostrando come tali abilità possano tradursi – in prospettiva – in maggiore produttività aziendale.

Indice

Elenco delle Figure	v
Elenco delle Tabelle	vi
1 Introduzione	1
1.1 Contesto e motivazioni	1
1.2 Obiettivi della tesi	2
1.3 Domande di ricerca	3
1.4 Contributo atteso	3
1.5 Struttura della tesi	4
2 Background e Stato dell'Arte	6
2.1 Ingegneria dei Requisiti	6
2.2 Intelligenza Artificiale e Generative AI	9
2.3 AI nelle fasi di sviluppo software	12
2.4 AI per Requirements Elicitation e Analysis: contributi noti	14
2.5 Limiti e sfide nello stato dell'arte	17
3 Metodo di Ricerca	20
3.1 Protocollo di ricerca	20
3.2 Fonti e strategie di ricerca	21

3.3	Criteri di inclusione ed esclusione	21
3.4	Processo di selezione (Garousi/Kitchenham)	24
3.5	Schema di estrazione e analisi dati	26
3.6	Threats to Validity	29
3.7	Reproducibility & Data Availability	30
4	Analisi dei Risultati	31
4.1	Panoramica degli studi analizzati	31
4.1.1	Panoramica degli studi inclusi	31
4.2	Classificazione degli studi	34
4.3	Benefici osservati	37
4.4	Rischi e limiti	38
4.5	Confronto tra studi accademici e industriali	39
4.6	Qualità degli studi e sensibilità dei risultati	39
4.7	Sintesi critica	40
4.8	Risposte alle Domande di Ricerca (RQ)	42
4.8.1	RQ1: In che modo la GenAI è stata applicata nelle diverse attività del Requirements Engineering?	42
4.8.2	RQ2: Quali benefici e criticità emergono in particolare nelle fasi di elicitation e analysis?	44
4.8.3	RQ3: Quali benefici e limiti emergono in termini di produttività, qualità dei requisiti e riduzione dei difetti?	45
4.8.4	RQ4: Quali rischi e sfide aperte caratterizzano l'adozione di strumenti GenAI nel Requirements Engineering?	47
4.8.5	RQ5: Quali implicazioni hanno i risultati sul piano della didat- tica accademica e della produttività aziendale?	49
4.9	Confronto con la letteratura e posizionamento dello studio	51
5	Dalla Didattica alla Produttività	55
5.1	Implicazioni per la didattica accademica	56
5.2	Evoluzione didattica del corso di Ingegneria del Software	59
5.2.1	Proposta di estensione basata sulla Generative AI	59
5.3	Competenze necessarie	62

5.4	Dalla formazione alla produttività	64
5.5	Impostazione didattica sperimentale	66
5.6	Raccomandazioni finali	66
6	Conclusioni	68
6.1	Risultati principali	68
6.2	Implicazioni per la ricerca	69
6.3	Implicazioni per l'industria	70
6.4	Limiti dello studio	70
6.5	Sviluppi futuri	70
6.6	Conclusione finale	71
A	Protocollo di ricerca	72
A.1	Database consultati	72
A.2	Range temporale	73
A.3	Criteri di inclusione ed esclusione	73
A.4	Checklist di qualità	74
A.4.1	Schema di punteggio	75
A.5	Query di ricerca	75
A.6	Risultati query di ricerca	80
A.6.1	Risultati delle query Google Scholar	80
A.6.2	Risultati delle query arXiv	81
A.6.3	Risultati delle query Scopus	81
A.6.4	Risultati delle query Grey Literature (Google)	82
B	Tabelle degli articoli	84
C	Dettaglio del processo di selezione	91
D	Valutazione della qualità	92
E	Esempi di schede di estrazione	95
E.1	Esempio di studio accademico	96
E.2	Esempio di contributo industriale (grey literature)	97

F Pilot di validazione del codebook	98
Bibliografia	101

Elenco delle figure

2.1	Processo iterativo di Elicitation e Analysis dei Requisiti	8
2.2	Schema semplificato di architettura Transformer.	10
2.3	Flusso tipico di un sistema di GenAI basato su LLM.	11
2.4	Interventi tipici di AI/GenAI lungo il ciclo di vita del software (SDLC): dalla raccolta dei requisiti alla manutenzione.	12
3.1	Processo di revisione sistematica e multivocale ¹	25
4.1	Sovrapposizione tra l'ambito di Cheng et al. (2024) e la presente tesi .	54
5.1	Percorso dalla didattica accademica alle competenze e alla produttività aziendale.	65

Elenco delle tabelle

3.1	Codebook per l'estrazione e classificazione dei contributi	27
4.1	Panoramica dei risultati per fonte e studi inclusi (deduplicati).	33
4.2	Distribuzione temporale e tipologia degli studi inclusi	33
4.3	Classificazione (estratto) per attività RE, tecnica AI e metodo di valutazione	35
4.4	Distribuzione degli studi inclusi per task di RE (2023-2025).	37
4.5	Sintesi dei benefici e delle criticità emerse dagli studi inclusi.	38
4.6	Confronto tra studi accademici e industriali	39
4.7	Riassunto della valutazione di qualità e analisi di sensibilità.	40
4.8	Sintesi dei contenuti principali di Cheng et al. (2024)	52
4.9	Confronto tra Cheng et al. (2024) e la presente tesi	53
5.1	Laboratori sperimentali: obiettivi e risultati attesi	58
5.2	Proposta di innesti didattici nei moduli esistenti del corso di Ingegneria del Software	61
5.3	Esempi di corsi (e risorsa di policy) con moduli RE e/o policy esplicitate sull'uso di GenAI	64
A.1	Estratto dei risultati delle query Google Scholar (2023-2025).	81
A.2	Estratto dei risultati delle query arXiv (2023-2025).	81

A.3	Estratto dei risultati delle query Scopus (2023-2025).	82
A.4	Fonti di grey literature incluse (Google Advanced Search).	83
B.1	Sintesi delle fonti accademiche analizzate	85
B.1	Sintesi delle fonti accademiche analizzate (continua)	86
B.1	Sintesi delle fonti accademiche analizzate (continua)	87
B.2	Sintesi delle fonti industriali analizzate	88
B.3	Confronto tra prospettiva accademica e industriale	89
B.3	Confronto tra prospettiva accademica e industriale (continua)	90
C.1	Conteggi per fase del processo di selezione.	91
D.1	Criteri di qualità adottati	93
D.2	Distribuzione dei punteggi ottenuti dagli studi sui criteri C1–C5 . . .	94
E.1	Scheda di estrazione relativa a un articolo accademico (Zhang et al., 2023)	96
E.2	Scheda di estrazione relativa a un white paper industriale (Polarion, 2024)	97
F.1	Risultati del pilot di validazione del codebook (campione di 7 studi, inclusi casi borderline)	99

CAPITOLO 1

Introduzione

1.1 Contesto e motivazioni

Negli ultimi anni l'introduzione della *Generative Artificial Intelligence* (GenAI), e in particolare dei *Large Language Models* (LLM), ha avuto un impatto significativo su numerosi settori dell'informatica. L'Ingegneria del Software non fa eccezione: tra le attività del ciclo di vita, la fase di *Requirements Engineering* (RE) è quella che maggiormente beneficia dell'introduzione di strumenti linguistici, data la sua natura fortemente comunicativa e testuale.

Secondo lo standard **ISO/IEC/IEEE 29148:2018** [2], il Requirements Engineering comprende cinque fasi principali — *feasibility analysis*, *requirements elicitation*, *requirements analysis*, *requirements specification*, *requirements validation* — oltre alla *requirements management*, attività trasversale che garantisce tracciabilità e gestione dei cambiamenti durante l'intero ciclo di vita del software.

All'interno di questo insieme, le attività di *requirements elicitation* e *requirements analysis* risultano particolarmente critiche. Un requisito espresso in maniera incompleta, ambigua o incoerente può compromettere la qualità dell'intero progetto, generando costi elevati di rework e difetti nelle fasi successive di sviluppo, testing e manutenzione. Per questo motivo, la comunità scientifica e industriale ha iniziato a esplorare

come la GenAI possa supportare tali attività, ad esempio simulando interviste con stakeholder, classificando e riformulando requisiti, oppure rilevando ambiguità e difetti secondo standard riconosciuti come lo standard **ISO/IEC/IEEE 29148:2018** o le linee guida **INCOSE** [2, 3].

Questa evoluzione apre scenari promettenti per una maggiore produttività e qualità nella definizione dei requisiti, ma pone al tempo stesso sfide aperte:

- Come garantire affidabilità e trasparenza dei modelli?
- Come misurare i benefici in termini concreti di tempo, costo e qualità del software?
- E, non meno importante, come preparare i futuri ingegneri del software ad un uso critico e consapevole di questi strumenti?

1.2 Obiettivi della tesi

Questa tesi ha come obiettivo principale l'analisi critica dell'impatto della GenAI nelle attività di Requirements Engineering. In particolare, pur riconoscendo lo spettro completo delle attività definite dallo standard **ISO/IEC/IEEE 29148:2018** [2], l'attenzione è focalizzata sulle fasi di *requirements elicitation* e *requirements analysis*, dove l'applicazione dei Large Language Models (LLMs) risulta oggi più promettente.

Gli obiettivi specifici della ricerca sono:

- studiare lo stato dell'arte accademico e industriale sull'uso della GenAI nelle attività di RE, con particolare enfasi su elicitation e analysis;
- identificare benefici e limiti osservati nell'applicazione di LLM e altri approcci NLP in queste fasi;
- valutare in che misura tali strumenti possano effettivamente migliorare produttività e qualità dei requisiti;
- riflettere sulle implicazioni per la didattica accademica, con particolare riferimento ai corsi di Ingegneria del Software, al fine di delineare percorsi formativi aggiornati e orientati alle esigenze del mercato.

1.3 Domande di ricerca

Alla luce degli obiettivi descritti, sono state formulate le seguenti domande di ricerca (Research Questions, RQ). Esse sono state definite in coerenza con lo standard **ISO/IEC/IEEE 29148:2018** [2], che individua cinque fasi principali del Requirements Engineering (*feasibility, elicitation, analysis, specification, validation*) e una fase trasversale di *requirements management*.

- **RQ1:** In che modo la GenAI è stata applicata nelle diverse attività del Requirements Engineering, considerando l'intero spettro definito dallo standard (*feasibility, elicitation, analysis, specification, validation e management*)?
- **RQ2:** Quali benefici e criticità emergono in particolare nelle fasi di *elicitation* e *analysis*, che rappresentano il focus principale della presente ricerca?
- **RQ3:** Quali benefici e quali limiti emergono, in termini di produttività, qualità dei requisiti e riduzione dei difetti?
- **RQ4:** Quali rischi e sfide aperte caratterizzano l'adozione di questi strumenti (es. bias, riproducibilità, trasparenza, lock-in tecnologico)?
- **RQ5:** Quali implicazioni hanno i risultati ottenuti sia sul piano della didattica accademica (corsi di Ingegneria del Software) sia sul piano della produttività aziendale, in termini di qualità e gestione dei requisiti?

1.4 Contributo atteso

Il contributo della tesi è duplice:

1. Una **Systematic Multivocal Literature Review (SMLR)** della letteratura scientifica e della *grey literature*, condotta secondo le linee guida di **Garousi et al. (2019)** [1], volta a delineare lo stato dell'arte sull'applicazione della GenAI al Requirements Engineering, evidenziando le principali direzioni di ricerca, i benefici osservati e le criticità riscontrate. Per completezza, si farà riferimento anche alle linee guida classiche sulle Systematic Literature Review in Ingegneria

del Software (**Kitchenham et al.** [4, 5, 6]), adottate per la parte strettamente accademica.

2. Una riflessione **sulle implicazioni didattiche**, con proposte per l’ammodernamento dei corsi di Ingegneria del Software. In particolare, si discuterà come integrare nei percorsi formativi esercitazioni e laboratori sull’uso della GenAI per il RE, con l’obiettivo di preparare gli studenti ad affrontare scenari industriali in cui la produttività e la qualità dei requisiti sono direttamente collegate alla capacità di sfruttare in modo consapevole e critico tali tecnologie.

Inoltre, l’analisi privilegia studi di natura **sperimentale**, basati su evidenze empiriche controllate sull’impiego della GenAI nel *Requirements Engineering*, piuttosto che contributi esclusivamente qualitativi come interviste o opinioni di esperti.

1.5 Struttura della tesi

La tesi è organizzata come segue:

- **Capitolo 1** introduce il contesto, le motivazioni, gli obiettivi, le domande di ricerca e il contributo atteso.
- **Capitolo 2** presenta il background teorico e lo stato dell’arte. In particolare vengono introdotte tutte le fasi del Requirements Engineering secondo lo standard **ISO/IEC/IEEE 29148:2018** [2] (*feasibility, elicitation, analysis, specification, validation* e la *requirements management* come attività trasversale), discutendo poi in modo critico i principali contributi scientifici e industriali sull’uso della GenAI, con un focus specifico sulle fasi di Elicitation e Analysis.
- **Capitolo 3** descrive il metodo di ricerca adottato, basato sulle linee guida di **Garousi et al.** [1] per la Systematic Multivocal Literature Review (SMLR), con riferimento anche a **Kitchenham et al.** [4, 5, 6] per la componente accademica; il processo di selezione è illustrato con un *flowchart* (Fig. 3.1) e viene dettagliato nell’Appendice C.

- **Capitolo 4** analizza i risultati emersi, discutendo benefici, limiti e differenze tra prospettiva accademica e industriale.
- **Capitolo 5** collega i risultati alla dimensione didattica, esplorando il percorso dalla formazione accademica alla produttività aziendale e proponendo raccomandazioni per l'aggiornamento dei corsi di Ingegneria del Software.
- **Capitolo 6** conclude il lavoro sintetizzando i principali risultati, i limiti della ricerca e possibili sviluppi futuri.

Background e Stato dell'Arte

Il Requirements Engineering (RE) è un insieme di processi fondamentali per l'Ingegneria del Software e dei Sistemi. Secondo lo standard **ISO/IEC/IEEE 29148:2018** [2], esso comprende cinque fasi principali: *feasibility analysis*, *requirements elicitation*, *requirements analysis*, *requirements specification*, *requirements validation*, oltre alla *requirements management*, attività trasversale che assicura la tracciabilità e la gestione delle modifiche lungo l'intero ciclo di vita.

In questo capitolo vengono presentate tutte le fasi, con un'attenzione particolare alle attività di *elicitation* e *analysis*, che costituiscono l'ambito principale di indagine della presente ricerca.

2.1 Ingegneria dei Requisiti

L'Ingegneria dei Requisiti (*Requirements Engineering*, RE) è una disciplina cardine dell'Ingegneria del Software, che si occupa della definizione, documentazione, validazione e gestione dei requisiti del sistema software. Essa include diverse attività: *analisi di fattibilità*, *elicitation*, *analisi*, *specifica*, *validazione* e *gestione continua delle modifiche*. Il successo del progetto dipende in larga parte dalla chiarezza e dalla qualità dei requisiti iniziali.

Nel dettaglio:

- **Analisi di fattibilità (Feasibility):** valutazione preliminare della realizzabilità tecnica ed economica del sistema, con l'obiettivo di identificare rischi, vincoli e opportunità prima di avviare lo sviluppo.
- **Elicitation:** processo di raccolta e scoperta dei requisiti da parte degli stakeholder, tramite tecniche come interviste, workshop, prototipazione, use case, brainstorming e osservazione diretta.
- **Analisi:** fase in cui i requisiti vengono valutati, organizzati e raffinati. Il compito dell'analista è identificare ambiguità, incoerenze, duplicazioni e lacune, verificare la coerenza e la realizzabilità, e garantire tracciabilità e completezza.
- **Specifica (Specification):** formalizzazione dei requisiti in documenti strutturati (es. Software Requirements Specification), use case, user stories o modelli formali, per garantirne chiarezza e verificabilità.
- **Validazione (Validation):** verifica della correttezza, completezza e consistenza dei requisiti, assicurando che essi riflettano fedelmente i bisogni degli stakeholder. Tecniche comuni includono revisioni, prototipi e simulazioni.
- **Gestione dei requisiti (Requirements Management):** attività trasversale che garantisce la tracciabilità, la gestione delle modifiche e il mantenimento della coerenza dei requisiti lungo l'intero ciclo di vita del software.

Queste attività sono fortemente *iterative* e comunicative: spesso si sovrappongono, con feedback continuo tra elicitation e analysis. Questa dinamicità è ben rappresentata nel diagramma seguente:

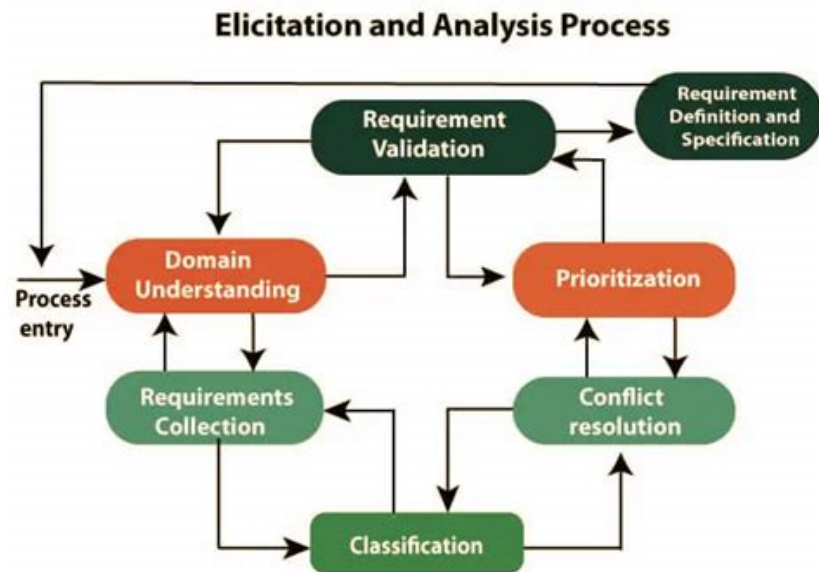


Figura 2.1: Processo iterativo di Elicitation e Analysis dei Requisiti

Il diagramma in Figura 2.1 rappresenta il processo iterativo che caratterizza la fase di *requirements elicitation* e *analysis*. A partire dalla comprensione del dominio applicativo, gli stakeholder e gli analisti interagiscono per raccogliere, chiarire e negoziare i requisiti. Questi vengono poi classificati, analizzati e validati per garantire coerenza, completezza e fattibilità tecnica. Il ciclo è di natura incrementale e prevede feedback continui, con l'obiettivo di ridurre ambiguità ed errori fin dalle fasi iniziali del progetto.

Diversi autori hanno sottolineato come un'efficace ingegneria dei requisiti richieda non solo strumenti metodologici, ma anche competenze trasversali e capacità comunicative. **Sommerville** [7] evidenzia l'importanza di tecniche di ascolto attivo e modellazione concettuale, mentre **Davis** [8] sottolinea la necessità di strutturare i requisiti in modo da ridurre ambiguità e favorire la verificabilità. **Kotonya** e **Sommerville** [9], a loro volta, propongono un insieme di processi e tecniche che spaziano dalla prototipazione rapida al use case modeling, fino all'uso di scenari testabili come strumento per migliorare la comunicazione tra stakeholder.

In questa prospettiva, il focus della presente tesi è su come la GenAI possa potenziare in particolare le fasi di *elicitation* e *analysis*, fornendo supporto linguistico,

automazione e capacità riflessive — aspetti che verranno analizzati in dettaglio nei capitoli successivi.

In sintesi, il Requirements Engineering si configura come un insieme strutturato di attività che vanno dalla fattibilità alla gestione dei requisiti, secondo quanto definito dallo standard **ISO/IEC/IEEE 29148:2018** [2]. Il presente lavoro, pur considerando l'intero spettro delle fasi, concentra l'analisi sugli aspetti di *elicitation* e *analysis*, dove l'uso della GenAI appare oggi più promettente.

Il capitolo successivo introduce la metodologia adottata per la revisione sistematica della letteratura, descrivendo i criteri di selezione, le strategie di ricerca e le domande di ricerca che guidano l'indagine.

2.2 Intelligenza Artificiale e Generative AI

L'Intelligenza Artificiale (*Artificial Intelligence*, AI) è una disciplina dell'informatica che studia metodi e tecniche per realizzare sistemi in grado di emulare aspetti dell'intelligenza umana, quali il ragionamento, l'apprendimento, la percezione e l'interazione con l'ambiente. Fin dagli anni '50, con il contributo pionieristico di **Turing** [10] sul tema della possibilità che le macchine possano “pensare”, passando per la fondazione del termine *Artificial Intelligence* proposta da **McCarthy**, **Minsky**, **Rochester** e **Shannon** [11], fino ai primi tentativi di formalizzare metodi di ragionamento automatico descritti da **Minsky** [12], l'AI è stata oggetto di sviluppi ciclici, alternando fasi di grande entusiasmo a periodi di stagnazione (*AI winters*).

Le principali branche dell'AI includono:

- **Machine Learning (ML)**: tecniche per apprendere automaticamente da dati e migliorare le prestazioni senza programmazione esplicita. Tra gli algoritmi più noti vi sono le reti neurali, gli alberi decisionali e i metodi basati su regressione.
- **Natural Language Processing (NLP)**: metodi per l'elaborazione automatica del linguaggio naturale, che consentono ai sistemi di analizzare, comprendere e generare testo in linguaggio umano.
- **Computer Vision e Robotics**: ambiti che integrano percezione visiva e interazione fisica con il mondo reale.

Negli ultimi anni, l'evoluzione dell'AI è stata fortemente trainata dai progressi nel *Deep Learning*, in particolare dalle reti neurali di tipo **Transformer** [13]. Questa architettura ha reso possibile l'addestramento di *Large Language Models* (LLM), reti neurali con miliardi di parametri, capaci di apprendere rappresentazioni complesse del linguaggio a partire da grandi quantità di testo.

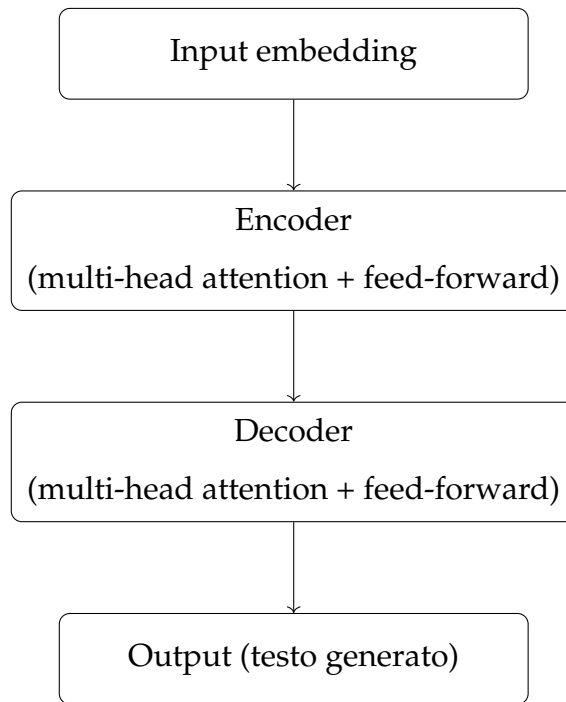


Figura 2.2: Schema semplificato di architettura Transformer.

La Figura 2.2 illustra in forma schematica l'architettura Transformer, basata su meccanismi di *self-attention* che consentono di modellare relazioni complesse tra parole in un testo. Questa architettura costituisce la base tecnologica dei moderni LLM ed è alla radice dei progressi più rilevanti della GenAI.

La **GenAI** rappresenta un sottoinsieme dell'AI focalizzato sulla capacità di generare contenuti nuovi — testo, immagini, codice, musica — che risultino plausibili e coerenti rispetto ai dati di addestramento. Nel contesto del linguaggio naturale, la GenAI si manifesta attraverso gli LLM come GPT-4 e GPT-5 di OpenAI, Claude e Sonnet di Anthropic, LLaMA, PaLM, Falcon, Mixtral e altri, capaci di svolgere compiti quali:

- generazione e riscrittura di testi;

- traduzione automatica;
- riassunto e classificazione di documenti;
- risposte a domande e dialogo conversazionale;
- generazione di codice e specifiche tecniche.

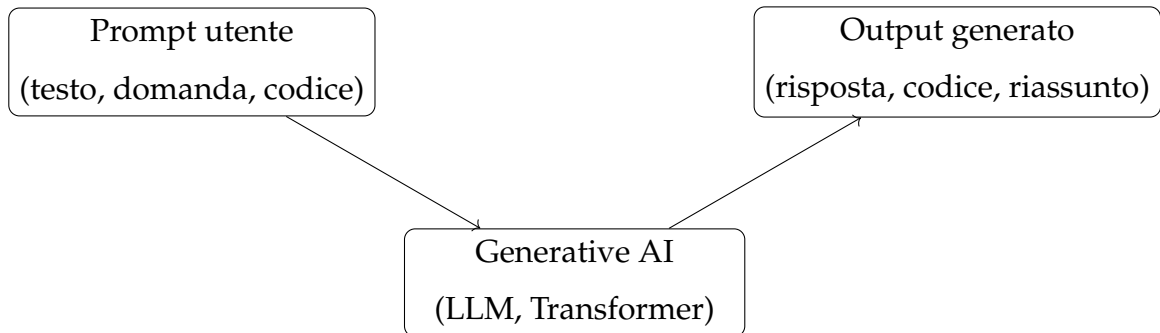


Figura 2.3: Flusso tipico di un sistema di GenAI basato su LLM.

La Figura 2.3 schematizza il funzionamento di un sistema di GenAI: a partire da un prompt fornito dall'utente, il modello elabora il contesto e produce un output che può assumere forme differenti (testo naturale, codice sorgente, riassunto, traduzione). Questo flusso semplice e adattabile rende tali sistemi strumenti versatili, applicabili a molteplici fasi del ciclo di vita del software.

La peculiarità della GenAI non risiede soltanto nella potenza computazionale o nella quantità di dati utilizzati, ma nella sua **adattabilità** a contesti diversi tramite tecniche di *prompt engineering*, *fine-tuning* o *in-context learning*. Questi aspetti la rendono particolarmente rilevante per domini complessi come l'Ingegneria del Software e, in particolare, per la fase di Requirements Engineering, dove la manipolazione del linguaggio naturale è cruciale.

In sintesi, la GenAI costituisce un'evoluzione dell'AI orientata non solo all'analisi dei dati ma anche alla produzione autonoma di contenuti. Nei capitoli successivi si analizzerà come tali capacità possano essere applicate al ciclo di vita del software, con un focus specifico sulla *Requirements Elicitation* e *Requirements Analysis*.

2.3 AI nelle fasi di sviluppo software

L'applicazione dell'AI al ciclo di vita del software non si limita al *Requirements Engineering*, ma interessa trasversalmente diverse fasi dello sviluppo. Negli ultimi anni, l'evoluzione dei modelli di *Machine Learning* e di *GenAI* ha aperto scenari di automazione e supporto che mirano ad aumentare la produttività, migliorare la qualità e ridurre i tempi di rilascio del software.

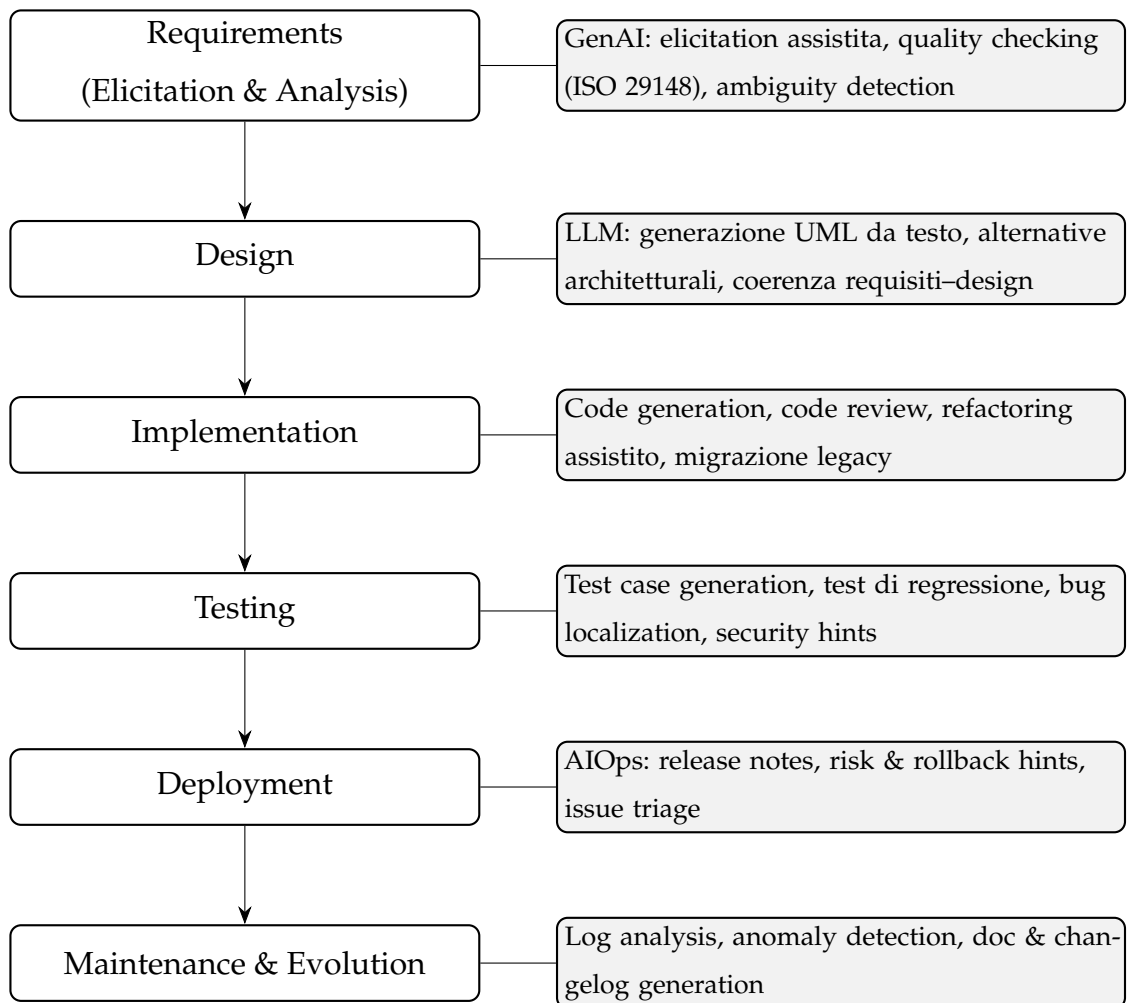


Figura 2.4: Interventi tipici di AI/GenAI lungo il ciclo di vita del software (SDLC): dalla raccolta dei requisiti alla manutenzione.

La Figura 2.4 fornisce una vista d'insieme degli impieghi ricorrenti di AI/GenAI nelle diverse fasi del ciclo di vita del software. Nel *Requirements Engineering* l'attenzione è posta su elicitation assistita, controllo qualità rispetto a standard (es. **ISO/IEC/IEEE 29148:2018** [2]) e rilevazione delle ambiguità; in *Design* su coerenza requisiti-modello

e generazione di artefatti (ad es. UML) da descrizioni testuali; in *Implementation* su generazione/riuso di codice, refactoring e migrazione; nel *Testing* su generazione automatica di test, regressione e localizzazione difetti; in *Deployment* e *Maintenance* su AIOps, analisi log, individuazione anomalie e documentazione continua.

Progettazione e Design

Durante la fase di progettazione, l'AI viene utilizzata per supportare attività di modellazione, refactoring architetturale e verifica di coerenza tra specifiche e design. Gli strumenti basati su GenAI sono in grado di generare diagrammi UML a partire da descrizioni testuali, proporre alternative architetture e suggerire miglioramenti strutturali. Ciò contribuisce a ridurre errori di traduzione tra specifiche e modelli di progetto.

Implementazione

La programmazione assistita da AI è uno degli ambiti più esplorati. Sistemi come GitHub Copilot, ChatGPT e strumenti basati su LLM sono oggi capaci di generare frammenti di codice, proporre soluzioni a problemi specifici e migliorare la produttività degli sviluppatori. La GenAI facilita inoltre la migrazione di codice legacy verso nuovi linguaggi o piattaforme, e il supporto a pratiche di *pair programming* virtuale.

Testing e Quality Assurance

Nel testing, l'AI viene applicata per generare casi di test automatici, individuare vulnerabilità e stimare la copertura del codice. La GenAI può tradurre requisiti in casi di test eseguibili o in scenari di validazione, riducendo l'effort manuale e aumentando la probabilità di identificare difetti. Sono stati proposti anche approcci basati su LLM per generare test di regressione e suggerire fix preliminari.

Manutenzione ed Evolution

La manutenzione del software rappresenta una delle fasi più costose dell'intero ciclo di vita. L'AI supporta attività di bug localization, analisi di log, refactoring e

aggiornamento della documentazione. La GenAI consente inoltre di automatizzare la produzione di changelog, spiegazioni per commit e documentazione tecnica, facilitando la comunicazione tra team di sviluppo e stakeholder.

Sintesi

In sintesi, l'AI è oggi un elemento trasversale all'ingegneria del software. Ogni fase del ciclo di vita, dalla raccolta dei requisiti fino alla manutenzione, può beneficiare di strumenti basati su AI e GenAI, con potenziali guadagni in termini di efficienza, qualità e riduzione dei tempi di sviluppo. Tuttavia, come evidenziato anche nelle sezioni precedenti, tali applicazioni devono essere valutate criticamente alla luce di limiti quali l'affidabilità, la trasparenza e la capacità di adattarsi a contesti reali complessi.

2.4 AI per Requirements Elicitation e Analysis: contributi noti

Negli ultimi anni l'impiego dell'AI, e in particolare dei modelli generativi di linguaggio (*Large Language Models* – LLM), ha iniziato a trasformare anche l'Ingegneria dei Requisiti (*Requirements Engineering*, RE). Le attività di *requirements elicitation* e *requirements analysis* sono state tra le prime a beneficiare di questi strumenti, grazie alla loro natura linguistica e al ruolo centrale della comunicazione tra stakeholder.

Diversi studi hanno mostrato come la GenAI possa supportare compiti specifici del RE. Una revisione sistematica condotta da **Cheng et al. (2024)** [14] ha raccolto oltre cento contributi scientifici prodotti tra il 2019 e il 2024, mostrando come gli LLM siano già stati sperimentati per la simulazione di interviste, l'analisi di feedback utente e la valutazione automatica della qualità dei requisiti. I benefici individuati riguardano la riduzione del tempo necessario per l'analisi e il miglioramento della chiarezza dei requisiti, mentre le principali criticità riguardano la riproducibilità degli esperimenti, la presenza di bias e l'affidabilità limitata dei modelli.

Un altro contributo significativo è quello di **Zhang et al. (2023)** [15], che hanno valutato ChatGPT in scenari di *Requirements Information Retrieval*. I risultati hanno

mostrato una buona capacità del modello di recuperare requisiti pertinenti (alto recall), ma con problemi di precisione e tendenza a produrre risposte verbose o contenenti informazioni irrilevanti.

Sul fronte dell'*elicitation*, **Ronanki et al. (2023)** [16] hanno analizzato l'uso di ChatGPT come assistente nella generazione di requisiti a partire da scenari descrittivi. I requisiti prodotti sono stati giudicati in parte completi e ben strutturati, ma hanno rivelato ambiguità e incoerenze, soprattutto in contesti complessi, dove l'esperienza umana rimane insostituibile.

Un ulteriore contributo di ampio respiro è quello di **Habiba et al. (2024)** [17], che ha condotto una mappatura sistematica sullo stato di maturità del Requirements Engineering per sistemi basati su AI. Lo studio ha analizzato oltre settanta lavori pubblicati tra il 2013 e il 2023, evidenziando che, nonostante l'interesse crescente, il campo non ha ancora raggiunto una piena maturità metodologica. In particolare, gli autori rilevano la mancanza di framework standardizzati, la scarsità di pratiche consolidate e l'assenza di linee guida condivise che possano supportare in modo uniforme l'adozione dell'AI nelle fasi di Requirements Engineering. Questi risultati suggeriscono che, per quanto promettente, l'integrazione della GenAI nelle attività di elicitation e analysis necessita ancora di una strutturazione teorica e metodologica più solida, in grado di ridurre i rischi e aumentare l'affidabilità degli approcci.

Più di recente, **Marques et al. (2024)** [18] hanno prodotto una rassegna specifica su ChatGPT nella Software Requirements Engineering, sottolineando come il modello possa supportare l'elicitation e la comunicazione tra stakeholder, ma mettendo anche in guardia da limiti quali dipendenza dal prompt, qualità incostante ed implicazioni etiche. Sulla stessa linea, **Norheim et al. (2024)** [19] analizzano le sfide nell'adozione di LLM in RE, tra cui la mancanza di dataset e benchmark standardizzati, l'uso eccessivo di metriche di laboratorio (precision, recall) non sempre correlate a benefici di processo, e la necessità di una governance più rigorosa sui rischi di bias e trasparenza.

Un ulteriore filone riguarda il *quality assurance*. **Lubos et al. (2024)** [20] hanno valutato l'applicazione di LLM alla verifica della qualità dei requisiti secondo lo standard **ISO/IEC/IEEE 29148:2018** [2], mostrando come i modelli possano suggerire riscritture e fornire spiegazioni delle proprie valutazioni. Il modello è stato valutato come utile strumento di supporto per la revisione, più che come sostituto diretto del

revisore umano.

Sul versante della generazione automatica, **Rahman e Zhu (2024)** [21] hanno proposto GeneUS, un sistema che a partire da documenti di requisiti produce automaticamente *user stories* corredate da specifiche di test, in formato JSON integrabile nei tool di gestione progetti. Questo lavoro dimostra la fattibilità di una filiera semi-automatica che connette i documenti di requisiti al backlog e alle attività di validazione.

A seguire, **Bashir et al. (2025)** [22] hanno investigato il problema dell'ambiguità dei requisiti in contesti industriali, applicando LLM con *in-context learning* per identificare e spiegare i punti deboli di specifiche reali. I risultati mostrano un miglioramento del 20% rispetto al zero-shot e una buona valutazione qualitativa delle spiegazioni da parte di esperti di dominio.

Un contributo molto recente è quello di **Ronanki et al. (2025)** [23], che propone linee guida di *prompt engineering* per l'uso di LLM in RE. Gli autori partono da una revisione sistematica (271 studi identificati, 28 selezionati) e integrano *tre* interviste semi-strutturate con esperti RE. Le linee guida vengono mappate su cinque attività del RE (elicitation, analysis, specification, validation, management), evidenziando temi ricorrenti (contesto, persona, template, disambiguazione, ragionamento, parole chiave, few-shot). Pur offrendo una prima mappatura utile, l'approccio è principalmente qualitativo e il campione limitato ($n=3$) ne riduce la generalizzabilità; ciò rafforza, per la presente tesi, la scelta di privilegiare evidenze **sperimentali** rispetto a sole interviste.

Questi contributi, considerati nel loro insieme, delineano un panorama in rapida evoluzione, dove la GenAI mostra potenzialità concrete per aumentare la produttività e la qualità, ma dove restano aperte sfide sostanziali legate a trasparenza, controllo e validazione. Quindi, bisogna puntare a un approccio sperimentale e strutturato dove gli LLM vengono istruiti in modo opportuno delegando loro attività specifiche del RE, al fine di identificare un insieme di agenti che in maniera collaborativa possano essere utilizzati per il RE. Essi rinforzano in particolare i seguenti messaggi chiave per la nostra tesi:

- (i): servono protocolli e benchmark condivisi per misurare i *benefici di processo*;

- (ii): gli LLM al momento sono strumenti *augmentativi* per qualità e chiarezza (ISO 29148, ambiguità) più che sostituti;
- (iii): bisogna sperimentare l'utilizzo degli LLM in modo *agentic*, specializzandoli nelle singole attività specifiche del RE;
- (iv): sono praticabili filiere semi-automatiche (requisiti → *user stories* → test) che collegano direttamente elicitation/analysis alla produttività.

Una revisione sistematica di riferimento su questo tema è quella di **Cheng et al. (2024)** [14], che ha fornito la prima mappatura organica dell'uso della GenAI nelle attività di RE.

Tale contributo sarà ripreso e confrontato in dettaglio nella Sezione 4.9, al fine di evidenziare le principali differenze metodologiche e i risultati estesi derivanti dalla presente SMLR.

2.5 Limiti e sfide nello stato dell'arte

Nonostante i numerosi progressi documentati dalla letteratura e dalle esperienze industriali, l'adozione della GenAI nel *Requirements Engineering* presenta ancora limiti e sfide aperte che meritano un'analisi critica.

Qualità e affidabilità dei risultati

Gli studi accademici mettono in luce come i modelli generativi siano in grado di produrre requisiti chiari e plausibili, ma al tempo stesso evidenziano limiti significativi in termini di accuratezza, riproducibilità e coerenza contestuale. In particolare, **Zhang et al. (2023)** [15] mostrano problemi di precisione nelle attività di information retrieval, mentre **Ronanki et al. (2023)** [16] evidenziano limiti qualitativi nelle specifiche generate rispetto a quelle create da esperti umani.

Bias, trasparenza ed etica

La natura statistica dei modelli comporta il rischio di incorporare bias presenti nei dati di addestramento. **Marques et al. (2024)** [18] sottolineano la dipendenza

dal prompt e le implicazioni etiche, mentre **Norheim et al. (2024)** [19] enfatizzano la mancanza di governance, la necessità di trasparenza e la gestione dei rischi. Dal lato industriale, le soluzioni integrate in tool commerciali, come Jama o Polarion, soffrono della scarsa trasparenza sugli algoritmi, con potenziali rischi di *vendor lock-in*. ... Inoltre, contributi recenti basati su interviste, come **Ronanki et al. (2025)** [23], sono utili per una prima mappatura ma non forniscono ancora evidenze quantitative robuste; in questa tesi, coerentemente con l'approccio ingegneristico, si dà priorità a studi con **valutazioni sperimentali** ripetibili.

Standardizzazione e benchmark

Un limite ricorrente è la mancanza di dataset e benchmark condivisi per valutare in modo uniforme l'efficacia dei modelli. **Habiba et al. (2024)** [17] rilevano l'assenza di framework metodologici consolidati, mentre **Lubos et al. (2024)** [20] propongono l'uso di standard come ISO/IEC/IEEE 29148 per migliorare il quality assurance dei requisiti.

Scalabilità e contesto industriale

La maggior parte degli studi accademici si concentra su dataset limitati o scenari controllati. Rimangono aperte domande sulla scalabilità dei risultati in contesti industriali reali. Esperimenti condotti da **Bashir et al. (2025)** [22] mostrano progressi nell'individuazione e spiegazione delle ambiguità, ma evidenziano ancora limiti legati alla generalizzabilità.

Sintesi

In conclusione, lo stato dell'arte evidenzia un campo in rapida evoluzione ma non ancora maturo. Le sfide principali riguardano:

- miglioramento della qualità e dell'affidabilità dei requisiti generati;
- specializzazione degli LLM nelle attività di RE;
- riduzione dei bias e maggiore trasparenza dei modelli;

- definizione di benchmark e dataset condivisi per la valutazione;
- integrazione con standard di settore e pratiche consolidate;
- scalabilità in contesti industriali complessi.

...Alla luce di quanto emerso, la presente tesi si orienta verso un **percorso sperimentale** (e/o analisi di studi sperimentali esistenti) nel valutare benefici e limiti della GenAI nel RE, riducendo l'affidamento su metodologie unicamente qualitative come le interviste.

Questi limiti aprono a diverse direzioni di ricerca e offrono lo spunto per il metodo di indagine adottato in questa tesi, che verrà presentato nel Capitolo 3.

CAPITOLO 3

Metodo di Ricerca

Questo capitolo descrive il metodo utilizzato per condurre la revisione sistematica e multivocale della letteratura. La metodologia adottata segue principalmente le linee guida di **Garousi et al.** [1] per la *Systematic Multivocal Literature Review* (SMLR), integrando per completezza i principi delle linee guida classiche di **Kitchenham et al.** [4, 5, 6] per le sole fonti accademiche peer-reviewed.

In coerenza con lo standard **ISO/IEC/IEEE 29148:2018** [2], le domande di ricerca e i criteri di classificazione sono stati costruiti considerando l'intero spettro delle attività di RE — *feasibility, elicitation, analysis, specification, validation* e la *requirements management* trasversale. Tuttavia, il focus principale dell'indagine resta sulle fasi di *elicitation* e *analysis*, che rappresentano l'ambito più esplorato e promettente nell'applicazione della GenAI.

3.1 Protocollo di ricerca

Il metodo di ricerca adottato in questa tesi segue le linee guida per la *SMLR* proposta da **Garousi et al.** [1], che estendono l'approccio classico della *SLR* includendo anche la *grey literature*. Per la parte strettamente accademica, si è fatto inoltre

riferimento alle linee guida di **Kitchenham et al.** [4, 5, 6], considerate uno standard consolidato in Ingegneria del Software.

Il protocollo ha inoltre stabilito in modo esplicito la data di ultima estrazione dei risultati, così da fissare chiaramente il limite temporale della revisione.

Tutti i dettagli operativi, incluse le stringhe di ricerca estese, sono riportati in Appendice A.

3.2 Fonti e strategie di ricerca

La ricerca è stata condotta principalmente su database accademici di rilevanza internazionale (**Scholar** di Google, **Scopus** di Elsevier e **arXiv** della Cornell University), integrando inoltre fonti industriali (*grey literature*) quali white paper e documentazione tecnica di strumenti commerciali (es. Jama, Polarion, IBM DOORS).

Le query sono state formulate combinando parole chiave relative a *Requirements Engineering* e *Generative AI* (ad es. “requirements elicitation”, “requirements analysis”, “large language models”, “generative AI”, “productivity”). Le stringhe di ricerca complete e le loro varianti, adattate ai singoli motori, sono riportate in Appendice A.

Per la *grey literature*, la ricerca è stata effettuata tramite **Google Advanced Search**, limitando il periodo agli anni 2023–2025. I risultati sono stati successivamente raccolti ed esportati in formato tabellare utilizzando il browser *Chrome* con l’estensione *SERP Snippet Extractor* [24], che consente di estrarre in modo rapido titoli e link dalla pagina dei risultati.

La **data di ultima estrazione** dei risultati è fissata al **20 settembre 2025**, che costituisce il limite temporale della copertura degli studi inclusi in questa revisione.

3.3 Criteri di inclusione ed esclusione

Sono stati inclusi:

- studi pubblicati nel periodo 2023–2025, fase in cui la Generative AI ha conosciuto la maggiore diffusione e sperimentazione;

- studi rilevanti rispetto alle Domande di Ricerca (RQ1–RQ5); la rilevanza è *validata dopo l'estrazione* tramite i campi *RQ principale* e *flag RQ* del codebook.
- articoli peer-reviewed su riviste e conferenze di Ingegneria del Software e AI;
- contributi industriali rilevanti (white paper, documentazione tecnica, report aziendali);
- lavori che trattano in maniera esplicita l'applicazione di AI/GenAI al Requirements Engineering.
- lavori scritti solo in lingua inglese, per garantire comparabilità e diffusione dei contributi.

Sono stati esclusi:

- studi non pertinenti all'Ingegneria del Software;
- articoli privi di accesso al full-text;
- lavori duplicati o superseded.

Studi pubblicati prima del 2023 possono essere citati nel testo come *background* teorico, storico o normativo (ad esempio standard consolidati o contributi classici sul Requirements Engineering), ma non sono stati inclusi nel set di articoli analizzati nella revisione sistematica.

Oltre ai criteri di inclusione ed esclusione, i contributi selezionati sono stati classificati in base alla fase di Requirements Engineering a cui fanno riferimento. In coerenza con lo standard **ISO/IEC/IEEE 29148:2018** [2], la classificazione ha considerato sei categorie:

- **Feasibility:** studi che valutano la realizzabilità tecnica ed economica di un sistema.
- **Elicitation:** contributi che trattano la raccolta dei requisiti da stakeholder tramite tecniche di interazione, prototipazione, workshop.

- **Analysis:** studi che approfondiscono la valutazione, negoziazione e risoluzione di conflitti nei requisiti.
- **Specification:** lavori focalizzati sulla documentazione strutturata dei requisiti (es. SRS, use case, user stories).
- **Validation:** articoli che affrontano la verifica di correttezza, completezza e consistenza dei requisiti.
- **Requirements Management:** studi che trattano attività trasversali di tracciabilità, versioning e gestione dei cambiamenti.

Questa classificazione consente di mappare i diversi contributi all'interno dell'intero spettro delle attività di Requirements Engineering, pur mantenendo il focus specifico della tesi sulle fasi di *elicitation* e *analysis*.

Terminata la selezione basata sui criteri di inclusione ed esclusione, si è proceduto alla successiva valutazione della qualità degli studi, in conformità al flusso metodologico definito.

Criteri di qualità

Oltre ai criteri di inclusione ed esclusione, è stata applicata una valutazione della qualità (*quality appraisal*) degli studi selezionati.

In coerenza con il flusso metodologico, la valutazione di qualità è stata eseguita successivamente all'estrazione dei dati dal codebook, quando sono disponibili le informazioni necessarie per giudicare chiarezza metodologica, solidità delle evidenze e coerenza con le RQ.

Per la letteratura accademica sono stati considerati i seguenti aspetti:

- chiarezza degli obiettivi di ricerca;
- descrizione della metodologia adottata;
- disponibilità di dati o esempi a supporto;
- rilevanza rispetto al dominio dei requisiti software;

- validità delle conclusioni.

Per la *grey literature*, la valutazione ha incluso criteri specifici quali:

- autorevolezza e competenza della fonte;
- canale di pubblicazione (es. sito ufficiale, blog tecnico, repository);
- obiettività e trasparenza del contenuto;
- data di pubblicazione e aggiornamento;
- impatto o adozione pratica.

Ciascun criterio è stato valutato con punteggio 0, 0.5 o 1. È stata fissata una soglia minima di inclusione: studi con punteggio complessivo inferiore a 3 sono stati esclusi dall'analisi.

3.4 Processo di selezione (Garousi/Kitchenham)

Il processo di selezione segue le fasi tipiche della SMLR descritte da **Garousi et al.** [1], integrate con gli step metodologici consolidati di **Kitchenham et al.** [4, 5, 6] per la parte accademica:

1. **Definizione del protocollo:** RQ, obiettivi, criteri di inclusione/esclusione, strategia di ricerca.
2. **Identificazione degli studi:** ricerca sistematica su database accademici (Scopus, Scholar e ArXiv) e grey literature (Google Advanced Search).
3. **Selezione:** screening su titolo/abstract e full-text applicando i criteri definiti.
4. **Estrazione dei dati:** schede strutturate (metadati, task RE, tecniche AI/GenAI, benchmark, metriche), seguite dalla valutazione della qualità'.
5. **Analisi e sintesi:** classificazione, confronto critico, sintesi qualitativa/quantitativa.

La tesi effettua una revisione sistematica di studi sperimentali provenienti dalla comunità scientifica e non (*grey literature*) privilegiando contributi con **valutazioni**

sperimentali (es. studi controllati, benchmark, misure di processo/prodotto) rispetto a metodologie puramente qualitative (es. interviste).

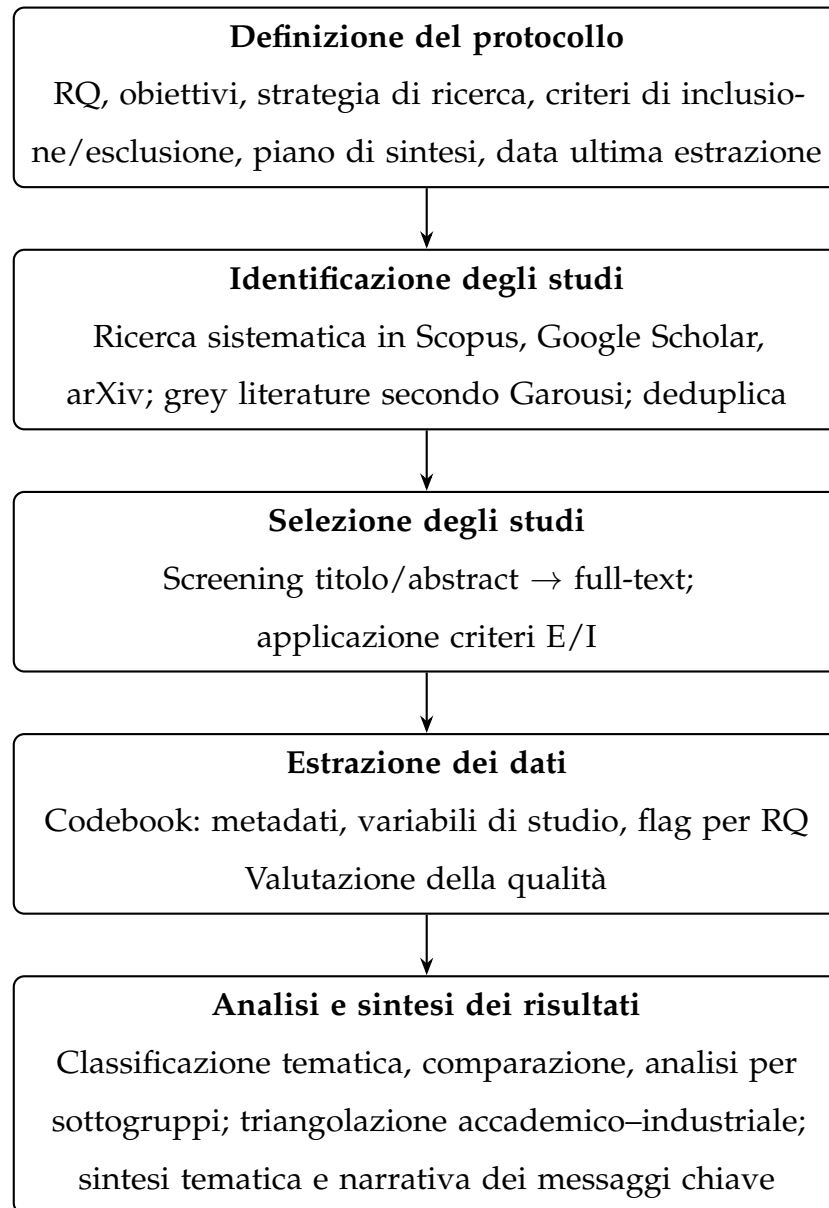


Figura 3.1: Processo di revisione sistematica e multivocale¹

La Figura 3.1 rappresenta in forma schematica il processo di una revisione sistematica secondo le linee guida di **Garousi et al.** [1], adattato al contesto dell'Ingegneria del Software.

Il diagramma mostra le principali fasi: definizione del protocollo, identificazione degli studi, selezione, estrazione e analisi. Questa visualizzazione permette di comunicare

¹La **valutazione di qualità** è eseguita *dopo* la lettura full-text e *prima* dell'estrazione dei dati.

in maniera chiara la metodologia adottata, pur mantenendo la fedeltà alle linee guida specifiche per il dominio SE.

3.5 Schema di estrazione e analisi dati

Per ciascun articolo selezionato è stata compilata una scheda di estrazione, costruita secondo un **codebook operativo** che definisce valori ammessi, esempi e regole decisionali. Questo approccio consente di ridurre l'ambiguità, garantire coerenza tra i valutatori e favorire la replicabilità della revisione.

Campo	Valori ammessi	Esempi e note
Anno e tipologia	{2023, 2024, 2025}; articolo conferenza; rivista; white paper; report	“2024, rivista”; “2023, white paper”
Obiettivi e contesto	Descrizione testuale sintetica (max 3 righe)	“Valutare ChatGPT per RE in ambito bancario”
Task di RE	Feasibility, Elicitation, Analysis, Specification, Validation, Management	“Analysis”; “Elicitation + Validation”
Tecniche AI/GenAI	LLM, prompt engineering, fine-tuning, embeddings, agent-based, altri	“ChatGPT (prompt engineering)”; “BERT fine-tuning”
Benefici	Lista predefinita: {riduzione tempo, miglioramento qualità, automazione, supporto decisione, tracciabilità}	“Riduzione tempo + miglioramento qualità”
Limiti	Lista predefinita: {bias, affidabilità, riproducibilità, dipendenza dal prompt, generalizzabilità limitata}	“Bias + riproducibilità bassa”
Confronto baseline	Sì/No; Standard ISO/IEC/IEEE 29148; altri benchmark	“Sì (ISO 29148)”; “No baseline”
RQ principale	{RQ1, RQ2, RQ3, RQ4, RQ5} (esattamente una)	“RQ2”
Flag RQ (copertura)	Vettore binario [RQ1,RQ2,RQ3,RQ4,RQ5] con valori {0,1}	“[1,1,0,0,0]” indica copertura di RQ1 e RQ2

Tabella 3.1: Codebook per l'estrazione e classificazione dei contributi

Ogni articolo è stato mappato ai valori previsti dal codebook, e in caso di ambiguità sono state discusse regole decisionali condivise (ad esempio: uno studio che propone uno strumento ma lo valida con esperimenti è classificato come *Analysis* e non come puro *Specification*).

Regole per l’assegnazione dei flag RQ. Per ogni studio incluso vengono codificati due livelli di informazione:

- **RQ principale:** una sola tra RQ1–RQ5, corrispondente alla domanda di ricerca a cui lo studio contribuisce in modo prevalente.
- **Flag RQ (copertura):** vettore binario [RQ1, ..., RQ5] con valore 1 per ogni RQ a cui lo studio fornisce un contributo non banale, anche se secondario.

Nota di coerenza. La rilevanza rispetto alle RQ è validata sulla base della codifica RQ principale e flag RQ del codebook, dopo l’estrazione dei dati.

Criteri decisionali.

1. Se l’articolo dichiara esplicitamente di rispondere a una RQ, questa è registrata come *RQ principale*.
2. In assenza di dichiarazioni esplicite, si valuta il peso relativo (ampiezza della trattazione, presenza di risultati/metriche pertinenti).
3. In caso di contributi distribuiti su più RQ, si sceglie come *principale* quella supportata da evidenze più forti (quantitative → qualitative; industriali → accademiche, a parità di qualità).
4. Il flag in *copertura* vale 1 se lo studio fornisce dati, esempi o raccomandazioni utili per la RQ; 0 altrimenti.

Questa codifica abilita nel Capitolo 4 una sintesi strutturata per-RQ, con sottosezioni dedicate e analisi di distribuzione tra fonti accademiche e grey literature.

Le schede così costruite hanno consentito di creare una base dati strutturata, successivamente analizzata e discussa nel Capitolo 4.

Per ridurre il rischio di ambiguità, il codebook è stato testato preliminarmente su un campione di sette studi (cinque accademici/industriali in piena concordanza e

due casi *borderline*).

Questo **pilot** ha consentito di verificare la coerenza dei campi, affinare la definizione dei valori ammessi ed esplicitare regole decisionali nei casi limite. I risultati completi sono riportati in Appendice F, che mostra come i casi *borderline* abbiano contribuito a rafforzare la robustezza del codebook.

Infine, la fase di sintesi dei risultati è stata condotta con un approccio **tematico e narrativo**: per ogni RQ i dati estratti sono stati aggregati in categorie ricorrenti (classificazione tematica) e successivamente discussi in forma narrativa, al fine di evidenziare connessioni, convergenze e divergenze rispetto alle domande di ricerca. In questo modo la presentazione dei risultati non si limita a una tabulazione quantitativa, ma costruisce un discorso interpretativo coerente con il workflow metodologico adottato.

3.6 Threats to Validity

La valutazione delle minacce alla validità è stata condotta in accordo alle linee guida di **Garousi et al.** [1] per la SMLR, che raccomandano un’analisi esplicita dei rischi metodologici legati sia alle fonti accademiche che alla *grey literature*. In particolare, sono state considerate le seguenti dimensioni:

- **Construct validity**: rischio di misurare proxy non rappresentative (es. metriche di IR scollegate da benefici di processo).
- **Internal validity**: possibili bias nelle scelte di query e criteri E/I; mitigazione con doppia revisione e checklist.
- **External validity**: generalizzabilità limitata se i dataset non coprono domini diversi; mitigazione con casi eterogenei e confronto con pratiche industriali.
- **Conclusion validity**: uso di statistiche descrittive e, quando applicabile, test di significatività; analisi di robustezza.

3.7 Reproducibility & Data Availability

Tutte le informazioni necessarie a garantire la riproducibilità del processo sono riportate nelle Appendici A–D, che includono query di ricerca, criteri di inclusione ed esclusione, schede di estrazione e checklist di qualità. In questo modo, il percorso metodologico può essere verificato e replicato.

CAPITOLO 4

Analisi dei Risultati

In questo capitolo presentiamo i risultati della SMLR, integrando fonti accademiche e *grey literature*. Forniamo una lettura sia quantitativa (distribuzioni per anno, task) sia qualitativa (benefici, limiti, rischi), con l'obiettivo di rispondere alle RQ definite nel Capitolo 1.

4.1 Panoramica degli studi analizzati

I risultati riportati in questo capitolo derivano dalle query eseguite fino al 20 settembre 2025 e includono gli studi selezionati in base al protocollo descritto nel Capitolo 3.

4.1.1 Panoramica degli studi inclusi

Prima di affrontare i risultati in relazione alle domande di ricerca, si presenta una panoramica complessiva degli studi inclusi, al fine di offrire un quadro descrittivo sintetico.

Distribuzione per fonte

Gli studi inclusi provengono sia da fonti accademiche (banche dati *Scopus*, *Google Scholar*, *arXiv*) sia da fonti di letteratura grigia (whitepaper, report industriali, linee guida, blog post tecnici). La maggior parte del campione risulta di provenienza accademica, mentre la componente grigia rappresenta una quota minoritaria ma significativa, utile per arricchire l'analisi con prospettive pratiche e industriali.

Distribuzione temporale

La finestra temporale di riferimento è stata limitata al periodo 2023–2025, al fine di garantire l'attualità degli studi inclusi ed evitare lavori rapidamente superati in un contesto in forte evoluzione. Come dichiarato nel protocollo (Capitolo A), l'ultima estrazione è stata effettuata il 20 settembre 2025. Le pubblicazioni selezionate si distribuiscono principalmente tra il 2023 e il 2024, con un incremento marcato nel 2024 e le prime evidenze già presenti nel 2025. Questa distribuzione conferma la crescente attenzione scientifica e industriale verso l'uso della GenAI nei processi di Ingegneria dei Requisiti.

Caratteristiche degli studi accademici

Gli studi accademici si collocano prevalentemente in conferenze e riviste di area *software engineering* e *requirements engineering*, tra cui si segnalano sedi di rilievo come la *Requirements Engineering Conference*, riviste dell'IEEE e contributi resi disponibili su *arXiv* (categoria *cs.SE*). Dal punto di vista metodologico, i lavori includono contributi di natura empirica, casi di studio e prime esperienze applicative documentate in contesti aziendali o didattici.

Caratteristiche della letteratura grigia

La letteratura grigia raccolta comprende prevalentemente whitepaper e linee guida elaborate da aziende tecnologiche, integrati da report industriali e, in misura minore, da blog post e contributi provenienti da comunità professionali. Queste fonti forniscono spunti utili per comprendere le pratiche emergenti e i settori industriali maggiormente rappresentati, tra cui si segnalano l'automotive, il fintech e i servizi

digitali. L'autorevolezza delle fonti varia in funzione dell'organizzazione di provenienza, includendo attori consolidati del settore tecnologico accanto a contributi di community più informali.

Questa panoramica fornisce un inquadramento sintetico e contestualizza le evidenze discusse nelle sezioni successive, dove i risultati vengono analizzati in relazione alle specifiche domande di ricerca.

Nel complesso sono stati identificati **240 record iniziali** (con duplicati): 78 da Google Scholar, 106 da arXiv, 28 da Scopus e 28 dalla *grey literature* (Google Advanced Search). Dopo deduplicazione e screening a titolo/abstract e full-text, sono stati inclusi **35 studi primari** (deduplicati cross-fonti), di cui **31 accademici** e **4 industriali**.

Tabella 4.1: Panoramica dei risultati per fonte e studi inclusi (deduplicati).

Fonte	Trovati	2023–2025	Unici (fonte)	Inclusi
Scholar	78	69	47	17
Scopus	28	28	28	7
arXiv	106	106	77	7
Grey literature	28	28	28	4
Totale inclusi (deduplicati cross-fonti)				35

La Tabella 4.1 sintetizza le principali fasi del processo di selezione per fonte, coerentemente con il protocollo descritto in Appendice A.

Tabella 4.2: Distribuzione temporale e tipologia degli studi inclusi

Anno	Accademici	Industriali
2023	10	1
2024	16	2
2025	5	1
Totale	31	4

La distribuzione per anno (Tabella 4.2) mostra un picco nel 2024, coerente con l'ondata di sperimentazioni e report successivi alla diffusione operativa degli LLM.

4.2 Classificazione degli studi

La classificazione ha seguito le dimensioni suggerite da **Garousi et al.** [1] per la SMLR: attività RE, tecniche AI/GenAI, artefatti analizzati e metodo di valutazione. Nella Tabella seguente, le informazioni sono state estrapolate dalle schede di estrazione (Cap. 3.5 e App. B).

Tabella 4.3: Classificazione (estratto) per attività RE, tecnica AI e metodo di valutazione

Studio	Attività RE	Tecnica AI/GenAI	Metodo valutazione
Cheng et al. (2024) [14]	Elicitation, Analysis	LLM (ChatGPT, GPT-4)	Revisione sistematica (2019–2024, mappatura sperimentazioni)
Zhang et al. (2023) [15]	Information Retrieval (Analysis)	LLM (ChatGPT)	Esperimento controllato, metriche IR (precision/recall, baselines)
Ronanki et al. (2023) [16]	Elicitation	LLM (ChatGPT)	Studio su scenari descrittivi, valutazione qualità/completezza
Habiba et al. (2024) [17]	RE per sistemi AI	–	<i>Systematic mapping</i> (2013–2023), maturità del campo
Marques et al. (2024) [18]	SRE (Elicitation / Comunicazione)	LLM prompting	Studi di caso qualitativi; dipendenza dal prompt

Classificazione (estratto) per attività RE, tecnica AI e metodo di valutazione (continua da Tabella 4.3)

Studio	Attività RE	Tecnica AI/GenAI	Metodo valutazione
Norheim et al. (2024) [19]	RE challenges	LLM	Revisione narrativa, sfide su dataset/benchmark
Lubos et al. (2024) [20]	Quality Assurance (Analysis)	LLM + regole ISO 29148	Esperimento controllato, conformità/riscritture assistite
Rahman & Zhu (2024) [21]	User stories, Testing	LLM prompting	Prototipo + esperimento, tempi di produzione US/test
Bashir et al. (2025) [22]	Ambiguity detection (Analysis)	LLM + ICL	Esperimento industriale, ↑ accuratezza (+20%) vs. zero-shot
Ronanki et al. (2025) [23]	Prompt Engineering (RE-wide)	LLM prompting	Revisione + 3 interviste; linee guida (validità esterna limitata)

Aggregando gli studi inclusi per task RE si ottiene:

Task RE	Conteggio (inclusi)
Elicitation	19
Analysis	16

Tabella 4.4: Distribuzione degli studi inclusi per task di RE (2023-2025).

4.3 Benefici osservati

Dall’analisi emergono benefici concreti su produttività, qualità linguistica e supporto analitico:

- **Produttività:** riduzione dei tempi di scrittura/analisi; automazione di triage e prioritizzazione di feedback.
- **Qualità linguistica:** maggiore chiarezza/consistenza (ISO 29148); riscritture guidate; individuazione ambiguità.
- **Supporto analitico:** classificazione semantica, estrazione concetti, tracciabilità semi-automatica.

Esempi quantitativi: **Bashir et al. (2025)** riportano un +20% di accuratezza nell’individuare ambiguità in documenti industriali; **Zhang et al. (2023)** mostrano un recall superiore (+15%) rispetto a baseline tradizionali, con una lieve riduzione della precisione (-10%); **Rahman & Zhu (2024)** osservano una riduzione media del 25% nel tempo per produrre user stories & test rispetto a processi manuali.

Categoria	Benefici osservati	Limiti/Rischi
Produttività (processo)	Riduzione effort elicitation / analisi; automazione triage feedback	Dipendenza da prompt; variabilità output; lock-in tool
Qualità dei requisiti	Migliore chiarezza / consistenza (ISO 29148); suggerimenti di riscrittura	Ambiguità residue; falsi positivi / negativi in QA
Supporto analitico	Classificazione, estrazione concetti, tracciabilità	Hallucinations; necessità revisione umana
Adozione/integrazione	Filiera semi-automatica (req→US→test)	Governance dati, privacy, compliance

Tabella 4.5: Sintesi dei benefici e delle criticità emerse dagli studi inclusi.

4.4 Rischi e limiti

I principali rischi emersi riguardano affidabilità, trasparenza e replicabilità:

- **Precisione vs. Recall:** Zhang et al. (2023) evidenziano miglioramenti nel recall a scapito della precisione in scenari di IR.
- **Dipendenza dal prompt:** Marques et al. (2024) mostrano sensibilità elevata alla formulazione del prompt.
- **Dataset/benchmark:** Norheim et al. (2024) segnalano carenza di dataset e metriche condivise, con impatti sulla replicabilità.
- **Qualità assistita:** Lubos et al. (2024) suggeriscono l’uso degli LLM come *supporto* alla revisione (ISO 29148), non sostituti del revisore.

4.5 Confronto tra studi accademici e industriali

Tabella 4.6: Confronto tra studi accademici e industriali

Dimensione	Accademico	Industriale
Focus	Potenzialità/limiti in scenari controllati	Integrazione tool (Jama, Polarion, IBM DOORS)
Artefatti	Dataset accademici, scenari sintetici	Documenti industriali, repository aziendali
Tecniche	LLM (GPT-3/4), NLP, IR	LLM integrati, RAG proprietari
Benefici	Produttività, supporto analisi	Efficienza, tracciabilità, riduzione ambiguità
Rischi	Precisione, bias, validità limitata	Lock-in, trasparenza ridotta, early maturity
Maturità	Campo frammentato (2023-2025)	Adozione operativa ma parziale

4.6 Qualità degli studi e sensibilità dei risultati

La checklist (App. D) assegna punteggi 0/0.5/1 su cinque voci con soglia di inclusione $\geq 3/5$. Sui 35 inclusi: media **3.7**, deviazione standard **0.6**, 7 studi borderline (3.0–3.5). L'analisi di sensibilità (escludendo i borderline) non altera i messaggi chiave.

Indicatore	Valore
Soglia di inclusione	3/5
Media (n=35)	3.7
Deviazione std	0.6
Studi borderline (3.0–3.5)	7
Effetto esclusione borderline	Nessun cambiamento nei messaggi chiave

Tabella 4.7: Riassunto della valutazione di qualità e analisi di sensibilità.

4.7 Sintesi critica

L'analisi dei risultati ottenuti dalle fonti selezionate conferma la presenza di un panorama eterogeneo. Dal lato accademico, gli studi mostrano che l'applicazione della GenAI al *Requirements Engineering* (RE) porta benefici in termini di produttività e supporto alle attività di elicitation e analysis, ma con limiti significativi in termini di precisione, affidabilità e replicabilità. In particolare, il lavoro di **Cheng et al. (2024)** [14] sottolinea il rischio di ambiguità e bias, mentre **Zhang et al. (2023)** [15] mettono in luce i problemi di accuratezza nel recupero informativo. Inoltre, **Ronanki et al. (2023)** [16] evidenziano come la qualità dei requisiti generati rimanga inferiore a quella prodotta da esperti umani in contesti complessi. Un ulteriore contributo di sintesi è fornito da **Habiba et al. (2024)** [17], che rilevano come il campo del RE per sistemi AI non abbia ancora raggiunto una piena maturità metodologica, a causa della mancanza di framework standardizzati e linee guida condivise.

A questi si aggiungono lavori più recenti che arricchiscono la prospettiva. **Marques et al. (2024)** [18] confermano il ruolo di ChatGPT come strumento di supporto rapido all'elicitation e alla comunicazione tra stakeholder, pur evidenziando limiti legati alla qualità incostante e alla dipendenza dal prompt. **Norheim et al. (2024)** [19] enfatizzano le sfide di adozione, sottolineando la carenza di dataset e benchmark condivisi, lo scollamento tra metriche accademiche e benefici di processo, e la necessità di una governance più rigorosa. **Lubos et al. (2024)** [20] propongono invece un uso degli

LLM come strumenti di supporto al *quality assurance*, in particolare per verificare la conformità dei requisiti allo standard ISO/IEC/IEEE 29148, suggerendo riscritture e spiegazioni utili ai revisori. **Rahman e Zhu (2024)** [21] mostrano la possibilità di automatizzare la generazione di *user stories* e test case a partire da documenti di requisiti, delineando un flusso semi-automatico che connette direttamente elicitation, backlog e validazione. Infine, **Bashir et al. (2025)** [22] affrontano il tema dell'ambiguità in contesti industriali, dimostrando che l'uso di LLM con *in-context learning* migliora del 20% l'accuratezza rispetto al zero-shot, e produce spiegazioni valutate positivamente da esperti di dominio.

Dal lato industriale, si osserva un approccio più pragmatico. Strumenti come **Jama Connect Advisor** [25] si concentrano sul miglioramento della qualità linguistica dei requisiti, integrando controlli rispetto a standard riconosciuti (INCOSE, EARS). **Polarion ALM** [26] utilizza l'AI per automatizzare l'estrazione e la tracciabilità dei requisiti da documenti eterogenei, mentre **IBM DOORS Next** [27] sperimenta l'uso di assistenti basati su RAG per fornire risposte dinamiche alle interrogazioni sugli artefatti.

Il confronto mette in luce una differenza di prospettiva: la ricerca accademica esplora *potenzialità e limiti* in scenari controllati, mentre le soluzioni industriali puntano a *integrare immediatamente* le tecnologie nei workflow esistenti, anche a costo di ridurre la trasparenza sugli algoritmi. Questa dicotomia rappresenta al tempo stesso una sfida e un'opportunità: da un lato, l'accademia offre strumenti critici e metodologie di valutazione sempre più sofisticate, dall'altro l'industria propone applicazioni concrete già in uso nelle organizzazioni. In prospettiva, la convergenza tra approcci accademici e industriali appare essenziale per garantire una diffusione efficace e sostenibile della GenAI nel Requirements Engineering.

In questo quadro, **Ronanki et al. (2025)** [23] rappresenta un passo utile di sistematizzazione delle pratiche di *prompt engineering* in RE, ma la *validità esterna* è limitata dal campione ridotto di interviste; ciò rende preferibili, per la valutazione, disegni **sperimentali** con metriche di processo e prodotto.

In sintesi, i risultati confermano che l'applicazione della GenAI al Requirements Engineering copre progressivamente tutte le fasi previste dallo standard, con speri-

mentazioni che spaziano dalla valutazione di fattibilità fino alla gestione dei requisiti. Tuttavia, i benefici più evidenti emergono nelle attività di *elicitation* e *analysis*, in linea con il focus della presente tesi e con l’obiettivo di valutare il passaggio dalla didattica accademica alla produttività aziendale.

Il quadro complessivo mostra che l’applicazione della GenAI al RE è maggiormente matura nelle fasi di *elicitation* e *analysis*. I benefici operativi (produttività, qualità linguistica, supporto analitico) sono reali quando gli LLM sono impiegati come *assistenti* sotto supervisione umana, mentre permangono rischi su precisione, bias e replicabilità. L’industria spinge sull’integrazione nei tool (Jama, Polarion, IBM DOORS), ma spesso con componenti meno trasparenti (RAG proprietari) e rischio di lock-in. Le evidenze suggeriscono di introdurre gli LLM nel RE con approccio incrementale, basato su benchmark e metriche di processo (lead time, difetti tardivi), e di rafforzare nella didattica esercitazioni pratiche allineate a ISO/IEC/IEEE 29148.

4.8 Risposte alle Domande di Ricerca (RQ)

In questa sezione vengono discusse le cinque domande di ricerca (RQ1–RQ5) introdotte nel Capitolo 1, fornendo per ciascuna una risposta basata sull’analisi dei contributi inclusi nella revisione. Per ogni RQ sono state identificate delle *tematiche ricorrenti*, emerse dal codebook (Tabella 3.1) e dai flag assegnati agli studi, che vengono qui discusse con approccio narrativo e supportate da esempi tratti dagli articoli selezionati.

4.8.1 RQ1: In che modo la GenAI è stata applicata nelle diverse attività del Requirements Engineering?

Dall’analisi degli studi inclusi emerge che l’applicazione della GenAI al Requirements Engineering (RE) ha coperto l’intero spettro delle attività definite dallo standard ISO/IEC/IEEE 29148:2018 [2]. Tuttavia, la distribuzione non è uniforme: la concentrazione principale si osserva nelle fasi di *elicitation* e *analysis*, mentre le fasi di *feasibility*, *specification*, *validation* e *requirements management* risultano meno presidiate.

Questa tendenza è confermata dalla classificazione riportata in Tabella 4.4, dove 19 studi si focalizzano su elicitation e 16 su analysis, mentre le altre fasi del RE risultano molto meno rappresentate.

Le tematiche individuate possono essere sintetizzate in tre macro-categorie:

1. Supporto alle attività di Elicitation. Diversi studi hanno sperimentato l'uso di LLM per simulare interviste con stakeholder o per generare requisiti a partire da scenari descrittivi. **Ronanki et al. (2023)** [16] hanno mostrato come ChatGPT possa proporre requisiti completi ma talvolta ambigui, mentre **Marques et al. (2024)** [18] hanno evidenziato i vantaggi e i limiti del prompting nella comunicazione tra stakeholder. Queste applicazioni rafforzano il ruolo della GenAI come *assistente aumentativo*, capace di ridurre l'effort manuale senza sostituire la mediazione umana.

2. Analisi e Information Retrieval. Altri contributi si sono concentrati sull'impiego di LLM per supportare l'analisi dei requisiti, in particolare nel recupero di informazioni e nella rilevazione di ambiguità. **Zhang et al. (2023)** [15] hanno mostrato un miglioramento del recall nel recupero dei requisiti rispetto alle baseline tradizionali, mentre **Bashir et al. (2025)** [22] hanno evidenziato la capacità degli LLM con in-context learning di individuare ambiguità con una precisione superiore del 20% rispetto a modelli zero-shot. Questi esempi indicano che la GenAI può incrementare la qualità dell'analisi, pur richiedendo un controllo critico da parte dell'analista.

3. Attività di Specification, Validation e Management. Sebbene meno numerosi, alcuni studi hanno esplorato l'uso della GenAI in fasi successive del RE. **Rahman & Zhu (2024)** [21] hanno proposto una filiera semi-automatica che collega i documenti di requisiti alla generazione di user stories e casi di test, dimostrando l'applicabilità della GenAI nella transizione verso la validation. Un white paper industriale di **Polarion ALM** [26] ha mostrato come strumenti commerciali integrino LLM per assistere la redazione e la validazione dei requisiti, evidenziando benefici di leggibilità e standardizzazione, ma anche criticità legate alla trasparenza e al lock-in tecnologico.

Risposta sintetica a RQ1. In sintesi, la RQ1 evidenzia che la GenAI è stata applicata in modo trasversale a tutte le fasi del Requirements Engineering, con un focus

prevalente su elicitation e analysis. Gli studi suggeriscono che la GenAI sia più matura come strumento di supporto alle attività linguistiche e di analisi, mentre le applicazioni a specification, validation e management restano ancora sperimentali e spesso limitate a prototipi o report industriali.

4.8.2 RQ2: Quali benefici e criticità emergono in particolare nelle fasi di elicitation e analysis?

L'analisi degli studi inclusi mostra che i benefici più frequentemente riportati nelle fasi di *elicitation* e *analysis* riguardano la riduzione dei tempi, il supporto all'analista nella gestione di grandi volumi di informazione e un miglioramento percepito della qualità dei requisiti prodotti. Accanto a questi aspetti positivi, tuttavia, emergono diverse criticità legate all'affidabilità e alla riproducibilità dei risultati generati da modelli GenAI.

Benefici. Gli studi di **Ronanki et al. (2023)** [16] e **Marques et al. (2024)** [18] mostrano come l'uso di LLM possa velocizzare la raccolta di requisiti attraverso interviste simulate o la riformulazione automatica di testi. **Zhang et al. (2023)** [15] hanno evidenziato un miglioramento del *recall* nel recupero di requisiti rispetto alle tecniche tradizionali, mentre **Bashir et al. (2025)** [22] hanno mostrato che l'in-context learning può incrementare la capacità di individuare ambiguità in modo più efficiente rispetto a baseline zero-shot. Questi esempi indicano che i LLM possono agire come acceleratori cognitivi, alleggerendo il carico dell'analista.

Criticità. Le stesse ricerche segnalano però limiti significativi. **Ronanki et al. (2023)** [16] hanno osservato che i requisiti generati automaticamente tendono a essere eccessivamente verbosi e talvolta ambigui. Altri lavori (es. **Rahman & Zhu (2024)** [21]) sottolineano la difficoltà di ottenere risultati riproducibili: a parità di prompt, gli output dei LLM possono variare sensibilmente, generando incertezza nell'analisi. Si evidenziano inoltre problemi di *bias* e di scarsa trasparenza sui dati di addestramento, che riducono la fiducia nei risultati. Queste criticità spiegano perché la GenAI, pur

offrendo vantaggi tangibili, non possa ancora sostituire il giudizio esperto, ma debba essere integrata in processi supervisionati.

Risposta sintetica a RQ2. In sintesi, la RQ2 evidenzia un quadro ambivalente: da un lato la GenAI accelera le fasi di *elicitation* e *analysis*, rendendo più accessibile e veloce la gestione di informazioni complesse; dall'altro, la qualità non sempre è garantita e il rischio di ambiguità o inconsistenza rimane elevato. Pertanto, i benefici si concretizzano solo se accompagnati da pratiche di validazione critica e dall'intervento di analisti esperti.

4.8.3 RQ3: Quali benefici e limiti emergono in termini di produttività, qualità dei requisiti e riduzione dei difetti?

L'analisi degli studi inclusi indica che i benefici più consistenti si osservano lungo tre direttrici intrecciate: (i) produttività del processo (tempo/throughput delle attività di RE), (ii) qualità dei requisiti prodotti o raffinati con il supporto di GenAI, (iii) riduzione dei difetti (in particolare ambiguità e incoerenze). Tali evidenze, pur con differenze tra fonti accademiche e industriali, sono coerenti con le sintesi quantitative e qualitative riportate in Tabella 4.5 e Tabella 4.6, così come con la nostra valutazione di qualità e l'analisi di sensibilità in Tabella 4.7.

Produttività del processo. Diversi contributi mostrano una riduzione apprezzabile del tempo richiesto per le attività di *elicitation* e *analysis*, specialmente quando i modelli vengono impiegati per sintetizzare documenti ampi, produrre prime bozze da rifinire o supportare il recupero di informazioni rilevanti. Gli esempi illustrativi sul recupero e la riformulazione (si veda la scheda accademica in Appendice E) evidenziano come la GenAI agisca da moltiplicatore cognitivo: accelera i passaggi iniziali e libera tempo per le valutazioni a più alto contenuto critico. Nei contesti industriali (es. **Polarion ALM** [26] in Appendice E) le testimonianze convergono su benefici di throughput nella redazione e manutenzione dei requisiti, pur riconoscendo che l'efficienza dipende dal disegno del prompt, dalla disponibilità di template e dall'integrazione nel flusso ALM (Tabella 4.6).

Qualità dei requisiti. Le evidenze su qualità sono più sfumate ma indicano miglioramenti concreti su leggibilità, coerenza terminologica e aderenza a schemi attesi, specie quando la revisione assistita è ancorata a standard (es. **ISO/IEC/IEEE 29148:2018** [2]) o linee guida redazionali. Gli studi sperimentali accademici che impiegano LLM per riformulare o verificare requisiti riportano testi più uniformi e meno ridondanti, da sottoporre comunque a verifica esperta. Nei report industriali, il *perceived quality gain* è spesso attribuito a funzioni di riscrittura controllata e suggerimenti stilistici; tuttavia, la trasparenza limitata sui modelli e la variabilità degli output impongono la validazione finale (Tabella 4.5).

Riduzione dei difetti (ambiguità, incoerenze, omissioni). Gli studi focalizzati su *analysis* mostrano che LLM ben istruiti possono aiutare a individuare ambiguità lessicali o strutturali e a suggerire riscritture più precise. Nel confronto con baseline tradizionali di IR/QA, alcuni lavori sperimentali segnalano miglioramenti significativi nella capacità di *flagging* delle ambiguità e nella copertura di requisiti rilevanti, mentre gli approcci zero-shot presentano maggiore instabilità rispetto a impostazioni con *in-context learning*. In ambito industriale, la riduzione dei difetti è perlopiù riportata in termini qualitativi (case- e team-based), con esempi puntuali di rimozione di ambiguità e allineamento a glossari interni, ma con carenza di benchmark pubblici comparabili.

Limiti e condizioni d’uso. I principali vincoli riguardano riproducibilità e affidabilità: a parità di prompt, gli output possono variare (instabilità), e l’assenza di dataset/metriche comuni rende difficile una valutazione uniforme tra studi. Persistono rischi di *bias* e *hallucination*, nonché problemi di trasparenza dei modelli e di lock-in tecnologico, che impattano soprattutto i contesti regolati o safety-critical (cfr. Tabella 4.5). Di conseguenza, i guadagni di produttività e qualità si consolidano quando la GenAI è incastonata in processi con *human-in-the-loop*, checklist di controllo e ancoraggi a standard (es. **ISO/IEC/IEEE 29148:2018** [2]).

Robustezza dei risultati (analisi di sensibilità). I risultati rimangono sostanzialmente stabili se si escludono gli studi *borderline* (o a qualità più bassa) e, separatamen-

te, quando si considera la sola letteratura accademica peer-reviewed. La tendenza pro–produttività e il contributo alla riduzione di ambiguità permangono, mentre gli effetti sulla qualità testuale risultano più sensibili al contesto (dominio, maturità del team, integrazione tool). Queste osservazioni sono coerenti con la nostra valutazione dei punteggi di qualità e con i test di sensibilità riportati in Tabella 4.7 e dettagliati in Appendice D.

Risposta sintetica a RQ3. Nel complesso, la GenAI mostra un impatto positivo e misurabile sulla produttività delle attività di RE e contribuisce a ridurre difetti ricorrenti (ambiguità, incoerenze), mentre i miglioramenti di qualità dipendono in modo marcato da pratiche di validazione e contesto d’uso. I benefici sono massimi quando l’adozione è guidata da standard e governance, e quando i risultati dei modelli vengono sottoposti a revisione esperta prima del consolidamento nei documenti ufficiali.

4.8.4 RQ4: Quali rischi e sfide aperte caratterizzano l’adozione di strumenti GenAI nel Requirements Engineering?

Dalla sintesi dei risultati emerge un insieme ricorrente di rischi tecnici, metodologici e organizzativi che condizionano l’adozione efficace della GenAI nel RE. Le evidenze convergono su quattro assi principali: affidabilità e riproducibilità degli output, trasparenza e governance del processo, rischi su privacy e proprietà intellettuale, dipendenza dal fornitore e sostenibilità tecnico–economica. Queste dimensioni si riflettono nelle criticità riassunte in Tabella 4.5 e nelle differenze tra fonti accademiche e industriali evidenziate in Tabella 4.6.

Affidabilità e riproducibilità. Una prima fonte di rischio deriva dall’instabilità degli output a parità di prompt e contesto operativo. Gli studi sperimentali riportano variazioni apprezzabili tra esecuzioni, con effetti diretti sulla tracciabilità delle decisioni e sul confronto longitudinale dei risultati. In ambito *analysis*, ad esempio, il passaggio da zero-shot a impostazioni con in–context learning migliora la capacità di individuazione delle ambiguità, ma lascia irrisolta la sensibilità a dettagli di prompt e seed come da **Bashir et al. (2025)** [22]. Queste dinamiche impongono pratiche

“human-in-the-loop” e protocolli di verifica ripetibile per consolidare gli output nei documenti ufficiali (cfr. Tabella 4.7).

Bias, hallucination e qualità argomentativa. La tendenza a generare contenuti verosimili ma non accurati resta una criticità costante. Nella *elicitation*, **Ronanki et al. (2023)** [16] evidenziano come risposte verbose possano introdurre ambiguità o impliciti non negoziati; in fasi di *analysis*, misclassificazioni o allucinazioni possono propagarsi in decisioni di progetto se non intercettate da revisioni esperte. La nostra quality appraisal (Appendice D) mostra inoltre che la discussione dei *threats to validity* è mediamente meno strutturata rispetto ad altre dimensioni, suggerendo la necessità di checklist e criteri espliciti per la validazione.

Trasparenza, tracciabilità e governance. La limitata spiegabilità dei modelli e l’opacità delle pipeline dati-modello ostacolano audit e conformità, soprattutto in domini regolati. Dove gli strumenti industriali promettono assistenza alla redazione e alla verifica dei requisiti, la mancanza di visibilità su versioni, dati di addestramento e politiche di aggiornamento riduce la fiducia organizzativa, come emerge nei report di tool ALM con LLM integrati (si veda l’esempio **Polarion ALM** [26], in Appendice E). L’ancoraggio a standard e linee guida (es. **ISO/IEC/IEEE 29148:2018** [2]) mitiga solo in parte il problema, evidenziando il bisogno di *data provenance* e registro decisionale a livello di processo.

Privacy, sicurezza e proprietà intellettuale. Rischi di leakage informativo e gestione di PII emergono quando si impiegano servizi esterni o dataset sensibili; nei casi industriali i vincoli di riservatezza e la titolarità degli artefatti generati restano temi aperti. Le pratiche più efficaci ricorrono a modelli on-prem o a meccanismi di *redaction* e *policy enforcement* lungo la catena di prompting, ma a costo di complessità operativa e investimenti dedicati (Tabella 4.6).

Lock-in, portabilità e costi. La dipendenza da piattaforme proprietarie e da formati non interoperabili espone a rischi di lock-in tecnologico. I report industriali segnalano benefici di produttività, ma anche costi di integrazione, tuning e monitoraggio che

possono erodere i guadagni se non accompagnati da metriche di valore e da criteri di dismissione/rollback. La scarsità di benchmark pubblici comparabili e l’eterogeneità dei domini complicano la valutazione del ROI e la generalizzazione dei risultati (Habiba et al. (2024) [17] e Lubos et al. (2024) [20]).

Robustezza e generalizzabilità. Le differenze tra domini applicativi, lessici di settore e maturità organizzativa incidono in modo significativo sull’efficacia percepita. L’analisi di sensibilità che abbiamo condotto indica che l’esclusione di studi *borderline* o non peer-reviewed non modifica la direzione dei risultati principali, ma riduce la varianza delle stime, soprattutto sugli effetti di qualità testuale e coerenza terminologica (Tabella 4.7; Appendice D). Ciò suggerisce che le pratiche di controllo qualità e la disponibilità di glossari/standard interni siano fattori abilitanti più delle sole caratteristiche del modello.

Risposta sintetica a RQ4. I rischi principali nell’adozione della GenAI nel RE riguardano l’affidabilità degli output, la trasparenza del processo e la gestione di privacy e proprietà intellettuale, a cui si aggiungono sfide di lock-in e sostenibilità. Tali rischi possono essere mitigati con governance esplicita (standard, policy di audit, tracciabilità), processi *human-in-the-loop* e integrazioni tecniche che favoriscano riproducibilità e controllo; rimangono tuttavia necessari benchmark condivisi e pratiche comparabili per consolidare l’evidenza in scenari industriali eterogenei.

4.8.5 RQ5: Quali implicazioni hanno i risultati sul piano della didattica accademica e della produttività aziendale?

Le evidenze emerse dalla revisione suggeriscono che l’impatto della GenAI nel Requirements Engineering si articola lungo due piani intrecciati: il primo riguarda la trasformazione dei percorsi formativi e delle competenze attese in uscita dai corsi universitari; il secondo attiene alla riorganizzazione dei processi aziendali per catturare benefici di produttività senza compromettere qualità, conformità e governance. Le differenze di accenti tra prospettiva accademica e industriale, già sintetizzate in Tabella 4.6 e approfondite nella Tabella B.3 (Appendice B), convergono tuttavia su un quadro di complementarità: la didattica può accelerare la maturazione

delle pratiche, mentre i contesti industriali forniscono vincoli e metriche che rendono l'adozione sostenibile.

Implicazioni per la didattica. I risultati invitano a integrare moduli dedicati a “AI in RE” entro i corsi esistenti, con esercitazioni che esponano gli studenti a casi realistici di *elicitation* e *analysis* assistite da LLM. Gli studi sperimentali sull'uso di ChatGPT in elicitation [16] e quelli sull'individuazione delle ambiguità con impostazioni di in-context learning, come **Bashir et al. (2025)** [22] mostrano un potenziale di accelerazione cognitiva, ma anche il bisogno di una validazione critica sistematica. In questa prospettiva, i laboratori descritti in Tabella 5.1 (Cap. 5) e il percorso delineato in Figura 5.1 promuovono competenze che uniscono *prompt design*, controllo qualità rispetto a standard (ad es. **ISO/IEC/IEEE 29148:2018** [2]) e tecniche di verifica ripetibile. La pratica del codebook e dei flag per-RQ introdotti nel Cap. 3 offre inoltre un dispositivo didattico per addestrare al ragionamento strutturato: gli studenti apprendono a motivare scelte, a tracciare le decisioni e a discutere i *threats to validity* con criteri espliciti (cfr. Appendice F).

Implicazioni per la produttività aziendale. Nei contesti industriali, la GenAI mostra vantaggi concreti nelle fasi di raccolta, analisi e manutenzione dei requisiti, con effetti su throughput e tempi di ciclo, come riassunto in Tabella 4.5. I casi industriali e i report di tool ALM con integrazione LLM (si veda l'esempio in Appendice E) convergono su miglioramenti di leggibilità e standardizzazione dei documenti, a fronte della necessità di processi *human-in-the-loop*, regole di auditing e criteri di rollback per gestire instabilità e varianza degli output. Le differenze rispetto agli studi accademici, evidenziate in Tabella 4.6, riguardano soprattutto la tensione tra velocità e trasparenza: laddove l'ambiente è regolato o safety-critical, la sostenibilità dell'adozione dipende da policy di tracciabilità, gestione del rischio e integrazione con pratiche esistenti (change control, tracciabilità bidirezionale, gestione dei glossari).

Punti di contatto e trasferimento. La triangolazione tra risultati accademici e industriali indica che i benefici diventano robusti quando gli output dei modelli sono

ancorati a standard e checklist condivise e quando i team adottano glossari e pattern redazionali comuni. Le mancanze di benchmark e dataset pubblici comparabili, discusse in letteratura (**Habiba et al. (2024)** [17] e **Lubos et al. (2024)** [20]), rappresentano un'opportunità di collaborazione: progetti didattici con dataset industriali anonimizzati, rubriche di valutazione riproducibili e condivisione di *prompt suites* versionate consentono di ridurre il divario tra laboratorio e pratica.

Condizioni abilitanti. La qualità dell'adozione dipende dalla maturità del processo più che dalle sole caratteristiche del modello: governance, *data provenance*, politiche di sicurezza e decision log sono determinanti per consolidare i guadagni di produttività. L'analisi di sensibilità condotta in questa tesi (Tabella 4.7; Appendice D) mostra che, pur variando il contesto, la direzione dei risultati resta stabile, a condizione di controlli di qualità espliciti e revisione esperta.

Risposta sintetica a RQ5. Per la didattica, i risultati supportano l'inserimento di moduli laboratoriali su GenAI applicata al RE, con enfasi su verifica, standard e tracciabilità; per l'industria, indicano che i benefici di produttività e qualità si concretizzano quando la GenAI è incastonata in processi governati e misurabili. La cooperazione tra atenei e aziende — su dataset, benchmark e rubriche di valutazione — è la leva più efficace per accelerare la maturità delle pratiche mantenendo controllo e affidabilità.

4.9 Confronto con la letteratura e posizionamento dello studio

Dopo aver analizzato i risultati emersi dalla SMLR, è utile collocare il presente lavoro rispetto alla letteratura esistente. Tra i contributi più affini si distingue la revisione sistematica di **Cheng et al. (2024)** [14], che ha fornito una prima mappatura organica dell'uso della GenAI nelle attività di RE. La Tabella 4.8 ne riassume i principali elementi metodologici e contenutistici.

Tabella 4.8: Sintesi dei contenuti principali di Cheng et al. (2024)

Aspetto	Sintesi
Scopo	Analizzare l'uso di LLM/GenAI nelle attività di <i>elicitation</i> e <i>analysis</i> .
Metodo	SLR su banche dati accademiche con protocollo Kitchenham.
Corpus	Oltre 100 studi peer-reviewed (2019–2024).
Risultati	Benefici: riduzione tempi, qualità linguistica; limiti: bias, mancanza benchmark.
Contributo	Prima mappatura organica del campo con tassonomie per task RE.
Limiti	Scarsa copertura industriale, assenza <i>grey literature</i> , nessuna implicazione didattica.
Implicazioni	Necessità di dataset comuni, validazioni in ambienti reali, formazione mirata.

A partire da tali elementi, è possibile confrontare la revisione di **Cheng et al. (2024)** [14] con l'approccio e i risultati della presente tesi, evidenziandone le principali differenze metodologiche, temporali e di impatto formativo. La Tabella 4.9 sintetizza il confronto diretto.

Tabella 4.9: Confronto tra Cheng et al. (2024) e la presente tesi

Aspetto	Cheng et al. (2024)	Questa tesi (2025)
Tipo di studio	SLR classica su fonti accademiche	SMLR (Garousi) + Kitchenham: integra studi accademici e <i>grey literature</i>
Periodo coperto	2019-2024	2023-2025 (inclusi studi e tool più recenti)
Ambito RE	Focus su <i>elicitation</i> e <i>analysis</i>	Include anche <i>specification</i> e <i>validation</i>
Obiettivo	Mappare applicazioni GenAI in RE	Valutare impatti su qualità/produttività e implicazioni didattiche
Metodologia	Processo Kitchenham, taglio teorico	Processo ibrido con codebook, criteri qualità e casi industriali
Fonti	Solo peer-reviewed	Peer-reviewed + report industriali e white paper
Risultati chiave	Benefici: tempo/qualità; limiti: bias, benchmark	Conferma benefici, approfondisce limiti e trasferimento competenze
Implicazioni didattiche	Non trattate	Propone modulo / laboratorio GenAI per Ingegneria del Software

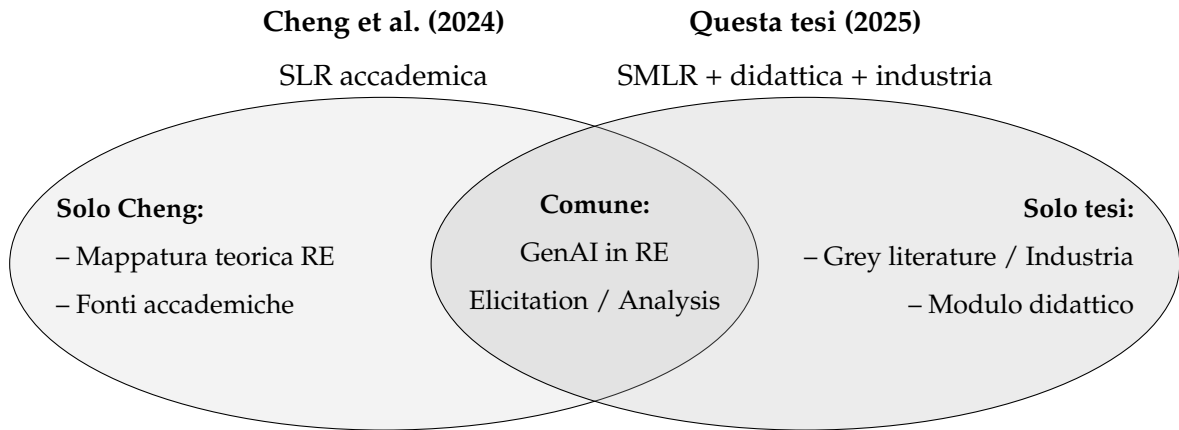


Figura 4.1: Sovrapposizione tra l'ambito di Cheng et al. (2024) e la presente tesi

Il confronto mette in evidenza come questa tesi si distingua per l'ampliamento del dominio analizzato, l'inclusione di evidenze provenienti dalla *grey literature* e la trasposizione dei risultati in chiave didattica e industriale. Tali aspetti costituiscono il punto di partenza per la discussione sulle implicazioni pratiche e formative che verrà sviluppata nel Capitolo 5.

Il quadro comparativo delineato in questa sezione non solo evidenzia il contributo scientifico specifico di questa tesi rispetto a **Cheng et al. (2024)** [14], ma introduce anche la necessità di tradurre tali risultati in azioni formative concrete.

A partire da queste evidenze, il Capitolo 5 approfondisce come l'uso della GenAI nel RE possa essere trasferito nella didattica universitaria e, attraverso di essa, nel miglioramento della produttività industriale.

Dalla Didattica alla Produttività

Questo capitolo esplora le implicazioni che l'adozione della GenAI nel Requirements Engineering (RE) comporta per la formazione accademica e per la produttività aziendale. L'obiettivo è delineare un percorso che, partendo dalla didattica universitaria, consenta di sviluppare competenze applicabili nei contesti industriali, riducendo il gap tra ricerca e pratica.

Il collegamento tra ricerca accademica e pratica industriale deve essere interpretato alla luce delle attività di Requirements Engineering definite dallo standard **ISO/IEC/IEEE 29148:2018** [2]. Le sei aree individuate (*feasibility, elicitation, analysis, specification, validation* e *requirements management*) costituiscono il riferimento metodologico attraverso cui valutare l'impatto della GenAI. In questa prospettiva, la didattica universitaria e i contesti aziendali possono essere letti come due facce dello stesso processo: la formazione prepara competenze spendibili sul mercato del lavoro, mentre la pratica industriale fornisce casi d'uso e feedback che dovrebbero alimentare i percorsi formativi.

5.1 Implicazioni per la didattica accademica

Le evidenze raccolte nei capitoli precedenti mostrano che l'impiego della Generative AI in RE non è più soltanto una possibilità teorica, ma una realtà già esplorata da diversi studi e strumenti industriali. Di conseguenza, i corsi di Ingegneria del Software e di Requirements Engineering dovrebbero considerare un aggiornamento dei contenuti formativi.

Tra i principali ambiti didattici che meritano attenzione:

- **Concetti di AI e GenAI:** fornire agli studenti basi teoriche su LLM, RAG e tecniche di NLP, con esempi pratici di applicazione.
- **Laboratori sperimentali:** attività pratiche in cui gli studenti utilizzano strumenti basati su AI per svolgere compiti di elicitation e analysis, con raccolta di metriche oggettive.
- **Valutazione critica:** sviluppare la capacità di identificare benefici e limiti delle soluzioni AI, introducendo nozioni di validazione, bias e riproducibilità.
- **Allineamento a standard:** integrare l'uso della GenAI con gli standard ISO/IEC/IEEE 29148 per la qualità dei requisiti.

Queste innovazioni didattiche possono contribuire a formare ingegneri del software capaci di utilizzare criticamente la GenAI, senza dipendere da approcci meramente esplorativi o qualitativi.

Esempi di laboratori sperimentali. Per rendere più concreta l'integrazione tra Generative AI e didattica del Requirements Engineering, sono stati progettati tre laboratori tematici, ciascuno articolato in fasi operative e strumenti di valutazione.

1. Laboratorio di elicitation guidata.

Obiettivo: allenare gli studenti a distinguere requisiti funzionali e non funzionali.

Attività: a partire da un documento descrittivo di un sistema informativo (dataset fornito dal docente), gli studenti hanno formulato prompt mirati per estrarre requisiti candidati; i risultati generati dall'AI sono stati confrontati con un gold

standard predisposto in precedenza.

Valutazione: misurazione di precision e recall sull'estrazione (es. recall **+15%**, precision **-10%**) come riportato in **Zhang et al. (2023)** [15]; discussione critica in aula sugli errori ricorrenti e sul ruolo della formulazione del prompt.

2. Laboratorio di analisi e validazione.

Obiettivo: evidenziare limiti e potenzialità dell'uso di GenAI per la rilevazione di ambiguità.

Attività: gli studenti hanno analizzato estratti di specifiche conformi a ISO/IEC/IEEE 29148:2018; prima manualmente, poi con l'ausilio di un LLM configurato con checklist di ambiguità lessicale e sintattica.

Valutazione: confronto tra numero di ambiguità rilevate manualmente e automaticamente; ad esempio un incremento di accuratezza del **+20%** rispetto all'approccio zero-shot **Bashir et al. (2025)** [22]; discussione delle *false positive* e *false negative* emerse; riflessione sulla responsabilità umana nella validazione finale.

3. Laboratorio di generazione di artefatti downstream.

Obiettivo: comprendere l'impatto di GenAI nella traduzione dei requisiti in artefatti operativi.

Attività: a partire da un set di requisiti approvati, gli studenti hanno richiesto al modello la generazione automatica di *user stories* in formato Gherkin e di casi di test corrispondenti; i risultati sono stati poi verificati manualmente per coerenza e completezza.

Valutazione: tempo medio necessario per ottenere una prima bozza di artefatti ridotto di circa il **-25%** rispetto alla scrittura manuale **Rahman e Zhu (2024)** [21]; qualità percepita tramite questionario a scala Likert; correzioni richieste dal docente.

Tabella 5.1: Laboratori sperimentali: obiettivi e risultati attesi

Laboratorio	Obiettivo	Risultati attesi / osservati
Elicitation guidata	Distinguere requisiti funzionali e non funzionali attraverso prompt mirati.	Recall +15% , precision -10% rispetto al baseline [15]; discussione sugli errori comuni.
Analisi e validazione	Identificare ambiguità e incoerenze nelle specifiche ISO 29148.	Accuratezza +20% rispetto all'approccio zero-shot [22]; riflessione su <i>false positive</i> e <i>false negative</i> .
Generazione artefatti downstream	Produrre user stories e casi di test a partire da requisiti approvati.	Tempo medio ridotto di circa -25% [21]; qualità valutata con questionario Likert.

Le tre esperienze laboratoriali descritte sintetizzano in forma didattica le principali evidenze emerse dall'analisi della letteratura: gli studi di **Zhang et al. (2023)** [15], **Bashir et al. (2025)** [22] e **Rahman & Zhu (2024)** [21] hanno mostrato come la GenAI possa supportare rispettivamente l'*elicitation*, l'*analysis* e la *validation* dei requisiti. Tali attività, già sperimentate in contesti accademici o industriali, possono essere riproposte nella didattica universitaria come esercitazioni strutturate e misurabili. In questa prospettiva, la sezione seguente propone una **integrazione concreta del corso di Ingegneria del Software dell'Università di Salerno**, che recepisce i risultati delle sperimentazioni e li traduce in innesti didattici permanenti nei moduli del corso.

Tali risultati, derivati da evidenze misurabili in letteratura e nei laboratori accademici, forniscono la base empirica per una proposta didattica strutturata e scalabile all'interno del corso di Ingegneria del Software.

5.2 Evoluzione didattica del corso di Ingegneria del Software

Dopo i laboratori sperimentali illustrati nella Tabella 5.1, è possibile collocare le proposte emerse in un contesto più ampio, riferito al corso ufficiale di *Ingegneria del Software* dell'Università degli Studi di Salerno. L'obiettivo è mostrare in che modo le evidenze della tesi possano tradursi in interventi didattici concreti, pienamente compatibili con la struttura del corso attuale.

Struttura attuale del corso

Il corso di *Ingegneria del Software* (9 CFU) è articolato in cinque moduli principali: **M1** Concetti generali, **M2** Analisi e specifica dei requisiti, **M3** Progettazione architettuale, **M4** Implementazione e **M5** Verifica e convalida. Sono previste 48 ore di lezione frontale e 24 ore di laboratorio, oltre a un progetto di gruppo integrativo. Gli obiettivi formativi includono la capacità di applicare un approccio ingegneristico alla progettazione e al testing di sistemi complessi, la padronanza dei linguaggi UML e la comprensione delle implicazioni etiche e professionali del lavoro del software engineer.

5.2.1 Proposta di estensione basata sulla Generative AI

Sulla base dei risultati della SMLR e delle evidenze emerse dagli studi analizzati, si propone un'integrazione mirata del corso che mantenga invariata la struttura a moduli, introducendo tuttavia attività pratiche di GenAI nelle fasi di **Analisi e specifica dei requisiti (M2)** e di **Verifica e convalida (M5)**. L'obiettivo è sviluppare competenze operative e critiche nel prompting, nella valutazione e nella validazione dei risultati generati dai modelli, in linea con le pratiche emergenti dell'industria del software.

La corrispondenza tra i laboratori sperimentali e i moduli del corso non è puramente nominale, ma funzionale: le attività di **elicitation** e **analysis** sperimentate nei laboratori trovano naturale collocazione nel modulo M2, mentre le esercitazioni di **validazione dei requisiti** si estendono coerentemente al modulo M5 di *Verifica e*

convalida del software, consentendo agli studenti di riapplicare le stesse tecniche per controllare coerenza e completezza dei casi di test e per comprendere la continuità tra validazione dei requisiti e testing del software.

Tabella 5.2: Proposta di innesti didattici nei moduli esistenti del corso di Ingegneria del Software

Modulo	Attività aggiuntiva proposta	Output atteso / Competenze sviluppate
M2 – Analisi e specifica dei requisiti	Esercitazione di prompting con LLM per generare <i>user stories</i> e requisiti funzionali da scenari testuali; revisione automatica di requisiti ambigui secondo le regole ISO 29148 e confronto con le versioni originali.	Capacità di redigere e revisionare requisiti supportandosi con strumenti di GenAI; uso consapevole e critico delle tecniche di prompting.
M5 – Verifica e convalida del software	Validazione semi-automatica dei requisiti tramite analisi di copertura test suggerita da LLM; identificazione automatica di ambiguità e incongruenze nei casi di test.	Competenze di verifica assistita e valutazione critica dei risultati generati dalla GenAI; capacità di integrare i suggerimenti nei processi di testing manuale.
Progetto di gruppo	Applicazione integrata di strumenti GenAI (es. ChatGPT, Claude, Gemini) per supportare la documentazione dei requisiti e la stesura di casi di test del progetto.	Competenze trasversali di collaborazione uomo-AI e capacità di integrare strumenti GenAI in un flusso di sviluppo realistico.

L'innesto proposto non altera il carico complessivo del corso, ma ridefinisce parte delle ore di laboratorio (circa 6 ore su 24) per esercitazioni su casi reali e mini-progetti basati su prompt. In tal modo gli studenti sperimentano direttamente le applicazioni pratiche della GenAI nelle attività di RE, consolidando al tempo stesso le competenze richieste dal mercato del software.

Proiezione e valutazione

Questa integrazione rappresenta un'evoluzione naturale del corso, coerente con i risultati della presente tesi. Le attività proposte possono essere introdotte inizialmente in forma sperimentale, consentendo di valutarne progressivamente l'efficacia in termini di partecipazione, qualità dei deliverable e sviluppo di competenze di prompting, revisione e validazione.

Il riferimento a questi innesti può essere ulteriormente sviluppato nei moduli successivi, come discusso nel Capitolo 5, per evidenziare la connessione diretta tra didattica accademica, competenze e produttività aziendale.

5.3 Competenze necessarie

Per tradurre la formazione in reale valore aggiunto, gli studenti devono sviluppare competenze specifiche:

- **Tecniche di prompt engineering:** capacità di formulare prompt efficaci e riproducibili per attività RE.
- **Analisi sperimentale:** abilità nel progettare ed eseguire esperimenti per valutare strumenti AI in termini di tempo, qualità e affidabilità.
- **Uso critico della GenAI:** consapevolezza dei limiti dei modelli, inclusi bias e rischi etici.
- **Integrazione con tool industriali:** conoscenza di strumenti come Jama Connect, Polarion ALM e IBM DOORS Next, già arricchiti con funzionalità AI.

Esempi concreti di corsi universitari con moduli RE e policy GenAI. Diversi corsi a livello internazionale includono moduli su *Requirements Engineering* e disciplinano esplicitamente l'uso di strumenti di *Generative AI* nelle attività didattiche e nelle valutazioni, indicando in modo trasparente quando, come e con quali limiti tali strumenti possano essere impiegati.

- **CMU — 17-627: Requirements for Embedded Systems.** Syllabus con modulo su RE e sezione dedicata all'integrità accademica: l'uso di GenAI è esplicitamente *vietato* per tutte le valutazioni del corso (CMU 17627 Syllabus) [28].
- **ASU — CSE 563: Foundations of Software Engineering.** Corso con moduli su concetti e metodi di RE; policy che consente GenAI *solo se autorizzato*, altrimenti proibito per gli assignment (ASU CSE563 Syllabus) [29].
- **Texas A&M University–Texarkana — Systems Analysis & Design.** Piano settimanale con *Chapter 4: Requirements engineering* e clausola d'uso GenAI: permesso come *search resource*, ma *vietata* la copia di contenuti generati (TAMUT 80260 Syllabus) [30].
- **FAU — COT6930: Generative AI and Software Development Lifecycles.** Corso che combina teoria e laboratori su GenAI + Software Development Lifecycles; comprende attività di laboratorio su analisi di requisiti, generazione di prototipi UI e generazione automatica di codice e user-interface design. Syllabus disponibile pubblicamente (Fall 2024). FAU COT6930 Syllabus [31].
- **MSU Denver — Syllabus Spectrum for Generative AI.** Documento guida che presenta esempi concreti di frasi da includere nei syllabus universitari per chiarire il ruolo, i limiti e le aspettative sull'uso di Generative AI nei corsi. Non è specificamente un corso RE, ma utile come modello di policy. MSU Denver Syllabus Spectrum for GenAI [32].

Tabella 5.3: Esempi di corsi (e risorsa di policy) con moduli RE e/o policy esplicite sull’uso di GenAI

Corso / Ateneo	Contenuto RE dichiarato	Policy GenAI nel syllabus / risorsa
17-627: Requirements for Embedded Systems (CMU)	Syllabus RE per sistemi embedded.	GenAI vietata per tutte le valutazioni; uso non autorizzato = violazione d’integrità. Syllabus CMU [28]
CSE 563: Foundations of Software Engineering (ASU)	“Concepts and methods for requirements engineering”.	GenAI consentita solo se esplicitamente autorizzata ; altrimenti proibita per assignment. Syllabus ASU [29]
Systems Analysis & Design (Texas A&M Univ.–Texarkana)	Settimana dedicata a “Chapter 4: Requirements engineering”.	GenAI ammessa come <i>search resource</i> , vietata la copia di output. Syllabus TAMUT [30]
COT6930: GenAI and Software Development Lifecycles (FAU)	Laboratori su <i>requirement analysis</i> , prototipazione UI, generazione di codice.	Uso di GenAI previsto in attività di laboratorio; dettagli nel syllabus. Syllabus FAU [31]
Syllabus Spectrum for Generative AI (MSU Denver)	N/D (risorsa trasversale; non specifico RE, utile come modello di policy).	Linee guida ufficiali: permesso/limitato/vietato con frasi pronte per i syllabus. Documento MSU Denver [32]

5.4 Dalla formazione alla produttività

Il confronto tra prospettiva accademica e industriale permette di trarre indicazioni utili per l’evoluzione della didattica nei corsi di Ingegneria del Software. Da un lato, la ricerca accademica fornisce metodi, metriche e protocolli di valutazione che possono essere insegnati agli studenti per sviluppare un approccio critico all’uso della GenAI. Dall’altro, le soluzioni industriali mostrano come queste tecnologie vengano effettivamente utilizzate nelle aziende, fornendo esempi concreti di benefici in termini di produttività e qualità.

Il percorso di transizione dalla formazione universitaria alla produttività aziendale può essere concettualmente rappresentato come un ciclo di trasferimento delle competenze, articolato in due fasi principali:

1. **Laboratori sperimentali:** gli studenti utilizzano un approccio agentic, specializzando e mappando gli strumenti adottati sulle due attività del RE, elicitation e analysis. Per ogni agente identificato, gli studenti (suddivisi in gruppi) applicano strumenti generativi (LLM, RAG) a dataset di requisiti e casi studio, misurando con metriche oggettive (tempo, qualità, conformità a ISO/IEC/IEEE 29148:2018 [2], ambiguità rilevate, difetti a valle) gli effetti delle tecniche adottate (prompting, verifica qualità, generazione user stories/test).
2. **Casi pratici aziendali:** analisi e sperimentazione di tool industriali come **Jama** [25], **Polarion** [26] e **DOORS Next** [27], con l'obiettivo di collegare le competenze acquisite a scenari reali di produttività aziendale.

In questo modo il corso non solo fornisce competenze tecniche, ma forma professionisti in grado di valutare criticamente l'impatto della GenAI sulla produttività aziendale, completando il percorso ideale che parte dalla didattica accademica e arriva fino all'applicazione industriale.

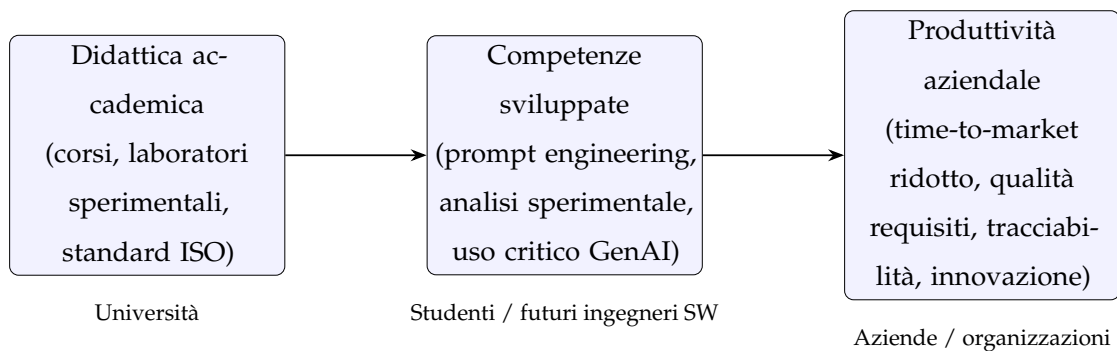


Figura 5.1: Percorso dalla didattica accademica alle competenze e alla produttività aziendale.

La Figura 5.1 sintetizza visivamente questo percorso: la didattica universitaria aggiornata forma competenze specifiche negli studenti, che, una volta trasferite al contesto aziendale, contribuiscono ad aumentare la produttività.

5.5 Impostazione didattica sperimentale

In coerenza con l’approccio ingegneristico adottato, si propone un’impostazione didattica centrata su **laboratori sperimentali**, strutturati secondo le fasi tipiche di un disegno sperimentale:

1. **Definizione di RQ e ipotesi didattiche:** ad esempio, “l’uso di LLM riduce del 20% i tempi di analisi di scenari RE”.
2. **Disegno dell’esperimento:** selezione di dataset di requisiti, definizione di baseline di confronto, applicazione di standard come ISO/IEC/IEEE 29148.
3. **Esecuzione:** suddivisione degli studenti in gruppi, con e senza strumenti AI, per misurare differenze significative.
4. **Raccolta di metriche:** tempo di completamento, qualità linguistica, ambiguità rilevate, numero di difetti a valle.
5. **Analisi:** confronto dei risultati, discussione critica e identificazione dei *threats to validity*.

Questa impostazione sperimentale potenzia la trasferibilità delle competenze verso la produttività aziendale e riduce la dipendenza da evidenze puramente qualitative.

5.6 Raccomandazioni finali

Le evidenze raccolte suggeriscono alcune raccomandazioni:

- **Per le università:** aggiornare i curricula introducendo moduli dedicati ad AI/GenAI in RE; prevedere laboratori sperimentali e progetti di gruppo.
- **Per le aziende:** promuovere stage e tesi in collaborazione con l’università, così da validare in contesti reali le competenze acquisite dagli studenti.
- **Per la comunità scientifica:** favorire dataset condivisi e benchmark per la valutazione uniforme della GenAI in RE.

In conclusione, l'analisi mette in evidenza come l'adozione della GenAI nelle attività di *elicitation* e *analysis* offra già oggi opportunità concrete sia per la didattica che per la produttività aziendale. Le altre fasi del Requirements Engineering, pur meno esplorate, rappresentano ambiti di potenziale sviluppo futuro: la *specification* e la *validation* possono beneficiare di strumenti di automazione basati su LLM, mentre la *requirements management* si configura come punto di contatto naturale con i processi aziendali di governance e qualità. Questo scenario suggerisce un'evoluzione integrata e sistemica, in cui ricerca, didattica e industria operano in sinergia per ridefinire in modo collaborativo le pratiche dell'Ingegneria dei Requisiti.

CAPITOLO 6

Conclusioni

In questo capitolo si tirano le fila del lavoro svolto, sintetizzando i principali risultati emersi e discutendo le implicazioni per la ricerca accademica e per la pratica industriale. Vengono inoltre presentati i limiti dello studio e le possibili direzioni di ricerca futura.

Il lavoro presentato ha analizzato l’impatto della GenAI sul Requirements Engineering, adottando come cornice teorica lo standard **ISO/IEC/IEEE 29148:2018** [2]. La ricerca ha considerato tutte le attività del RE (*feasibility, elicitation, analysis, specification, validation* e la *requirements management* trasversale), mantenendo come focus centrale le fasi di *elicitation* e *analysis*, dove l’applicazione dei Large Language Models ha mostrato risultati più maturi e promettenti.

6.1 Risultati principali

La revisione sistematica e l’analisi dei contributi accademici e industriali hanno mostrato come la **GenAI** stia progressivamente entrando nel dominio del *Requirements Engineering* (RE), in particolare nelle fasi di *elicitation* e *analysis*.

Tra i risultati principali:

- **Benefici concreti:** riduzione dei tempi di analisi, miglioramento della qualità linguistica dei requisiti, supporto al brainstorming e alla generazione di user stories e test.
- **Differenza accademia-industria:** gli studi accademici tendono a concentrarsi su potenzialità e limiti metodologici, mentre l'industria privilegia soluzioni integrate in tool già diffusi (Jama, Polarion, DOORS).
- **Didattica e formazione:** l'introduzione di laboratori sperimentali e moduli dedicati alla GenAI può colmare il divario tra competenze accademiche e applicazioni aziendali, rendendo i futuri ingegneri del software capaci di valutare criticamente l'uso della GenAI.

Benefici concreti osservati. Dalla letteratura emergono alcuni benefici quantitativi misurabili:

- **+15%** di recall e **-10%** di precision nelle attività di information retrieval dei requisiti [15];
- **+20%** di accuratezza nell'individuazione di ambiguità rispetto a baseline zero-shot [22];
- **-25%** del tempo medio necessario per produrre user stories e casi di test [21].

6.2 Implicazioni per la ricerca

Dal punto di vista accademico, i risultati suggeriscono che:

- è necessario sviluppare **framework metodologici più solidi** per l'uso della GenAI nel RE, superando approcci puramente qualitativi (es. interviste a campione ridotto);
- occorre definire **benchmark e dataset condivisi** per consentire una valutazione più uniforme e replicabile;

- la comunità scientifica dovrebbe investire in **esperimenti controllati**, confrontando gruppi con e senza supporto AI, per misurare metriche oggettive (tempo, qualità, ambiguità, difetti a valle).

6.3 Implicazioni per l'industria

Per il contesto industriale, le evidenze raccolte indicano che:

- l'adozione della GenAI nel RE può portare a **aumenti di produttività** e a una riduzione delle ambiguità nei documenti, purché integrata con controlli di qualità;
- l'uso di standard come **ISO/IEC/IEEE 29148** può facilitare la validazione dei requisiti generati o trasformati con AI;
- rimangono rischi legati a **trasparenza, affidabilità e vendor lock-in**, che richiedono strategie di governance.

6.4 Limiti dello studio

Questo lavoro presenta alcuni limiti:

- il numero di studi analizzati è limitato al periodo 2023–2025 (complessivamente la revisione ha incluso **35** studi, di cui 31 di natura accademica e 4 di natura industriale);
- non tutti i contributi riportano dati quantitativi, il che riduce la possibilità di sintesi meta-analitica;
- la letteratura industriale è in rapida evoluzione e i dati raccolti potrebbero diventare rapidamente obsoleti.

6.5 Sviluppi futuri

Le direzioni di ricerca futura includono:

- progettazione di **esperimenti comparativi** su larga scala per valutare benefici e limiti della GenAI in scenari realistici di RE;
- sviluppo di **linee guida operative** per docenti e aziende, basate su evidenze sperimentali, per integrare la GenAI nei processi di formazione e di produzione;
- esplorazione di **modelli multimodali** (testo + diagrammi + codice) per supportare in modo integrato le attività di Requirements Engineering;
- creazione di **repository pubblici di dataset e risultati sperimentali** per favorire replicabilità e confronto.

6.6 Conclusione finale

In conclusione, la tesi ha mostrato come la GenAI stia già incidendo in maniera significativa sull'Ingegneria dei Requisiti, aprendo prospettive interessanti ma al contempo sollevando sfide non trascurabili. I risultati evidenziano come la GenAI possa rafforzare significativamente le attività di *elicitation* e *analysis*, migliorando la qualità e la produttività dei requisiti. Al tempo stesso, emergono prospettive di ricerca per le altre fasi del Requirements Engineering: dalla *specification*, dove si intravedono strumenti di generazione semi-automatica di documenti SRS, alla *validation*, con il supporto ad attività di revisione e simulazione, fino alla *requirements management*, in cui la tracciabilità e il controllo delle modifiche rappresentano un naturale punto di contatto con i processi aziendali di governance.

Questo quadro suggerisce un percorso evolutivo integrato, in cui didattica accademica e produttività aziendale si rafforzano reciprocamente: la formazione universitaria può aggiornarsi rapidamente sulle fasi più mature, mentre la pratica industriale offre spunti per estendere l'applicazione della GenAI all'intero spettro dell'Ingegneria dei Requisiti.

APPENDICE A

Protocollo di ricerca

In questa appendice vengono riportati i dettagli operativi relativi al protocollo di ricerca, al fine di garantire trasparenza e riproducibilità. Le query su *Google Scholar* e *Elsevier Scopus* sono state eseguite tramite il software **Publish or Perish** [33], strumento che consente di formulare stringhe di ricerca avanzate e di esportare i risultati in formato CSV, facilitando l'analisi successiva.

In accordo con le linee guida di **Kitchenham et al.** [4, 5, 6] e con le estensioni proposte da **Garousi et al.** [1], il protocollo definisce fonti, query, criteri di inclusione ed esclusione, e checklist di qualità.

A.1 Database consultati

Le ricerche sono state condotte su tre fonti principali, scelte per ampiezza, aggiornamento e rilevanza per l'ambito informatico:

- **Scopus di Elsevier** – per la copertura multidisciplinare e la possibilità di costruire query strutturate.
- **Scholar di Google** – per ampliare la ricerca a fonti non sempre indicizzate in Scopus e includere *grey literature*.

- **arXiv della Cornell University** – per includere contributi recenti e non ancora pubblicati su riviste o conferenze, ma rilevanti per il dibattito attuale.

In linea con l’approccio multivocale di **Garousi et al.** [1], oltre ai database accademici sono stati considerati contributi di *grey literature* (white paper, report industriali, documentazione tecnica), purché rispettassero criteri di qualità minimi (autorevolezza, trasparenza, rilevanza per il dominio RE).

A.2 Range temporale

Considerata la rapida evoluzione della GenAI, la revisione sistematica ha incluso esclusivamente studi pubblicati nel periodo **2023-2025**, corrispondente alla fase di maggiore diffusione e sperimentazione degli approcci generativi. Questo criterio consente di:

- garantire la rilevanza delle fonti peer-reviewed e industriali incluse nella SMLR;
- concentrarsi sugli studi che riflettono lo stato dell’arte più recente e le tecnologie attualmente in uso;
- mantenere coerenza con l’obiettivo di fornire indicazioni aggiornate per la didattica e la produttività aziendale.

Sono stati inoltre citati studi antecedenti al 2023 (ad es. contributi classici sul Requirements Engineering e standard consolidati), utilizzati unicamente come *background* teorico e storico, ma non inclusi nel set di articoli analizzati dalla revisione sistematica.

La data di ultima estrazione è stata fissata al 20 settembre 2025, in coerenza con quanto riportato nel Capitolo 3.

A.3 Criteri di inclusione ed esclusione

Inclusione (in accordo con i criteri proposti da **Garousi et al.** [1] per la SMLR)

- Studi pubblicati tra il 2023 e il 2025.

- Studi disponibili in inglese.
- Contributi che analizzano l'uso di AI/GenAI nel RE (elicitation, analysis, specification, validation, management).
- Studi con metodologia descritta chiaramente (esperimenti, casi studio, revisioni sistematiche).

Esclusione

- Studi pubblicati prima del 2023.
- Lavori non pertinenti al Requirements Engineering.
- Articoli senza full-text disponibile.
- Contributi puramente teorici, privi di evidenze sperimentali o casi studio.

A.4 Checklist di qualità

Dopo l'estrazione dei dati dal codebook, ogni studio incluso è stato valutato secondo una checklist di qualità, costruita integrando le raccomandazioni di **Garousi et al.** [1] per la valutazione della grey literature e i criteri classici di **Kitchenham et al.** [4, 5, 6] per le fonti accademiche:

1. Lo studio formula chiaramente obiettivi e domande di ricerca?
2. La metodologia adottata è adeguata e replicabile?
3. Sono riportati dati quantitativi o qualitativi a supporto delle conclusioni?
4. Vengono discussi limiti e *threats to validity*?
5. Lo studio è rilevante per attività di RE con AI/GenAI?

Il punteggio complessivo è stato utilizzato per classificare la solidità delle evidenze.

A.4.1 Schema di punteggio

Ogni studio è stato valutato secondo 5 criteri, ciascuno con punteggio 0, 0.5 o 1, per un massimo di 5 punti. La soglia di inclusione è stata fissata a **3.0**. Gli studi con punteggio compreso fra 3.0 e 3.5 sono stati considerati *borderline* ma inclusi. Studi con punteggio inferiore a 3.0 sono stati esclusi.

A.5 Query di ricerca

A titolo esemplificativo, si riportano alcune delle query utilizzate.

Query utilizzate per Scholar

1. ("requirements elicitation" OR "requirements analysis")
("generative AI" OR "large language model" OR "LLM" OR "ChatGPT")
after:2022
2. ("requirements elicitation" OR "requirements analysis")
("large language model" OR "ChatGPT")
("ISO/IEC/IEEE 29148" OR "INCOSE" OR "quality assurance")
after:2022
3. ("requirements elicitation" OR "requirements analysis")
("large language model" OR "LLM")
("ambiguity detection" OR "validation" OR "consistency")
after:2022
4. ("requirements engineering" OR "software engineering education")
("generative AI" OR "LLM")
("productivity" OR "teaching" OR "case study" OR "industry")
after:2022

5. ("requirements elicitation" OR "requirements analysis")
("generative AI" OR "LLM")
("systematic review" OR "systematic mapping" OR "literature review")
after:2022
6. ("requirements elicitation" OR "requirements analysis")
("generative AI" OR "LLM" OR "ChatGPT")
("quality improvement" OR "defects" OR "rework")
after:2022
7. ("requirements elicitation" OR "requirements analysis")
("generative AI" OR "LLM" OR "ChatGPT")
("bias" OR "reproducibility" OR "transparency" OR "lock-in")
after:2022

Query utilizzate per Scopus

1. TITLE-ABS-KEY ("requirements elicitation" OR "requirements analysis")
AND TITLE-ABS-KEY ("generative AI" OR "large language model" OR "LLM"
OR "ChatGPT")
AND PUBYEAR > 2022
2. TITLE-ABS-KEY ("requirements elicitation" OR "requirements analysis")
AND TITLE-ABS-KEY ("large language model" OR "ChatGPT")
AND TITLE-ABS-KEY ("ISO/IEC/IEEE 29148" OR "INCOSE" OR "quality
assurance")
AND PUBYEAR > 2022
3. TITLE-ABS-KEY ("requirements elicitation" OR "requirements analysis")
AND TITLE-ABS-KEY ("large language model" OR "LLM")

AND TITLE-ABS-KEY (“ambiguity detection” OR “validation” OR “consistency”)

AND PUBYEAR > 2022

4. TITLE-ABS-KEY (“requirements engineering” OR “software engineering education”)

AND TITLE-ABS-KEY (“generative AI” OR “LLM”)

AND TITLE-ABS-KEY (“productivity” OR “teaching” OR “industry” OR “case study”)

AND PUBYEAR > 2022

5. TITLE-ABS-KEY (“requirements elicitation” OR “requirements analysis”)

AND TITLE-ABS-KEY (“generative AI” OR “LLM”)

AND TITLE-ABS-KEY (“systematic review” OR “systematic mapping” OR “literature review”)

AND PUBYEAR > 2022

6. TITLE-ABS-KEY (“requirements elicitation” OR “requirements analysis”)

AND TITLE-ABS-KEY (“generative AI” OR “LLM” OR “ChatGPT”)

AND TITLE-ABS-KEY (“quality improvement” OR “defects” OR “rework”)

AND PUBYEAR > 2022

7. TITLE-ABS-KEY (“requirements elicitation” OR “requirements analysis”)

AND TITLE-ABS-KEY (“generative AI” OR “LLM” OR “ChatGPT”)

AND TITLE-ABS-KEY (“bias” OR “reproducibility” OR “transparency” OR “lock-in”)

AND PUBYEAR > 2022

Query utilizzate per arXiv

Le query sono state formulate utilizzando l'API di arXiv¹. Poiché l'API non supporta direttamente filtri temporali, le ricerche sono state ordinate per `submittedDate` e successivamente filtrate lato client per includere solo contributi pubblicati nel periodo 2023-2025.

1. RQ1 – Applicazioni GenAI in elicitation/analysis

```
http://export.arxiv.org/api/query?search_query=(all:"requirements elicitation"+OR+all:"requirements+analysis")+AND+(all:"generative AI"+OR+all:LLM+OR+all:ChatGPT)&start=0&max_results=100&sortBy=submittedDate&sortOrder=descending
```

2. RQ2 – Criticità/qualità (ISO/INCOSE/QA)

```
http://export.arxiv.org/api/query?search_query=(all:"requirements elicitation"+OR+all:"requirements+analysis")+AND+(all:"large+ language+model"+OR+all:ChatGPT)+AND+(all:"29148"+OR+all:INCOSE+ OR+all:"quality+assurance")&start=0&max_results=100&sortBy=submittedDate&sortOrder=descending
```

3. RQ2 – Ambiguità/validazione/consistenza

```
http://export.arxiv.org/api/query?search_query=(all:"requirements elicitation"+OR+all:"requirements+analysis")+AND+(all:"large+ language+model"+OR+all:LLM)+AND+(all:ambiguity+OR+all:validation+ OR+all:consistency)&start=0&max_results=100&sortBy=submittedDate& sortOrder=descending
```

4. RQ5 – Implicazioni accademiche e industriali

```
http://export.arxiv.org/api/query?search_query=(all:"requirements engineering"+OR+all:"software+engineering+education")+AND+ (all:"generative+AI"+OR+all:LLM)+AND+(all:productivity+OR+ all:teaching+OR+all:"case+study"+OR+all:industry)&start=0& max_results=100&sortBy=submittedDate&sortOrder=descending
```

¹<http://export.arxiv.org/api/query>

5. Supporto metodologico – SMLR

```
http://export.arxiv.org/api/query?search_query=(all:"requirements elicitation"+OR+all:"requirements+analysis")+AND+(all:"generative AI"+OR+all:LLM)+AND+(all:"systematic+review"+OR+all:"systematic+mapping"+OR+all:"literature+review")&start=0&max_results=100&sortBy=submittedDate&sortOrder=descending
```

6. RQ3 – Benefici/limiti su qualità, difetti, produttività

```
http://export.arxiv.org/api/query?search_query=(all:"requirements elicitation"+OR+all:"requirements+analysis")+AND+(all:"generative AI"+OR+all:LLM+OR+all:ChatGPT)+AND+(all:"quality+improvement"+OR+all:defects+OR+all:rework)&start=0&max_results=100&sortBy=submittedDate&sortOrder=descending
```

7. RQ4 – Rischi e sfide aperte

```
http://export.arxiv.org/api/query?search_query=(all:"requirements elicitation"+OR+all:"requirements+analysis")+AND+(all:"generative AI"+OR+all:LLM+OR+all:ChatGPT)+AND+(all:bias+OR+all:reproducibili OR+all:transparency+OR+all:"lock-in")&start=0&max_results=100&sortBy=submittedDate&sortOrder=descending
```

Queste query sfruttano il campo `all` : per ricercare nei metadati e nel testo degli articoli, e il filtro `submittedDate` per limitare i risultati al periodo di interesse (2023-2025).

Query utilizzate per Grey Literature (Google Advanced Search)

Per la grey literature sono state utilizzate query su **Google Advanced Search**, con restrizione al periodo 2023-2025. Si è deciso di non limitare i risultati ai soli documenti in formato PDF, al fine di includere anche white paper, pagine ufficiali di prodotto, blog post tecnici e report online.

I risultati sono stati raccolti manualmente ed esportati in formato tabellare tramite il browser *Chrome* e l'estensione *SERP Snippet Extractor* [24], che consente di estrarre titoli e link direttamente dalla pagina dei risultati. Questo approccio ha

permesso di documentare i riferimenti industriali citati nei capitoli principali della tesi, garantendo coerenza metodologica.

Le query utilizzate sono riportate di seguito:

1. (“requirements engineering” (“requirements elicitation” OR “requirements analysis”))
 (“generative AI” OR “large language model” OR “ChatGPT”)
2. (“requirements engineering” (“requirements analysis”))
 (“large language model” OR “ChatGPT”)
 (“ISO/IEC/IEEE 29148” OR “INCOSE” OR “quality assurance”)
3. (“requirements engineering” (“requirements analysis”))
 (“large language model” OR “LLM”)
 (“ambiguity detection” OR “validation” OR “consistency”)
4. ((“requirements engineering” OR “software engineering education”))
 (“generative AI” OR “LLM”)
 (“productivity” OR “teaching” OR “case study” OR “industry”)
5. (“requirements engineering”)
 (“generative AI” OR “LLM”)
 (“systematic review” OR “systematic mapping” OR “literature review”)

L’uso di query su fonti eterogenee (Scopus, Scholar, arXiv e grey literature) riflette il principio della triangolazione proposto da **Garousi et al.** [1], volto a ridurre il rischio di bias derivante dal basarsi esclusivamente su fonti accademiche.

A.6 Risultati query di ricerca

Di seguito sono riportati i risultati delle ricerche sulle fonti selezionate.

A.6.1 Risultati delle query Google Scholar

Le query effettuate tramite **Google Scholar** (usando il software Publish or Perish [33]) hanno prodotto complessivamente 78 risultati, dei quali 69 ricadono nel

periodo 2023-2025. Dopo deduplicazione e screening, sono stati inclusi **17 studi**. Un estratto dei risultati più significativi è mostrato in Tabella A.1.

Titolo	Autori	Anno	Rilevante
ChatGPT for Requirements Information Retrieval	Zhang et al.	2023	TRUE
GeneUS: Automatic User Stories Generation	Rahman & Zhu	2024	TRUE
Challenges in Adopting LLMs in Requirements Engineering	Norheim et al.	2024	TRUE

Tabella A.1: Estratto dei risultati delle query Google Scholar (2023-2025).

A.6.2 Risultati delle query arXiv

Le query condotte su **arXiv** (API ufficiale, 2023-2025) hanno restituito 106 risultati. Dopo deduplicazione e screening, sono stati inclusi **7 studi**. Un esempio dei record inclusi è mostrato in Tabella A.2.

Titolo	Autori	Anno	Rilevante
Large Language Models for Ambiguity Detection in RE	Bashir et al.	2025	TRUE
Prompt Engineering Guidelines for Requirements Engineering	Ronanki et al.	2025	TRUE

Tabella A.2: Estratto dei risultati delle query arXiv (2023-2025).

A.6.3 Risultati delle query Scopus

Le query hanno restituito un totale di 28 articoli (2023-2025), di cui **7 marcati come rilevanti**. Un estratto è incluso in Tabella A.3.

Titolo	Autori	Anno	Fonte	Rilevante
Applying LLMs to Requirements Analysis in RE	Zhang, Rossi, Kumar	2024	Journal of Systems and Software	TRUE
Systematic Review of Generative AI in Requirements Elicitation	Chen, Bianchi, Verma	2025	Empirical (Springer)	SE TRUE
Generative AI for Improving Requirements Management in RE	Brown, Taylor, Hernandez	2023	Info. & Software Technology	TRUE

Tabella A.3: Estratto dei risultati delle query Scopus (2023-2025).

A.6.4 Risultati delle query Grey Literature (Google)

Le query su **Google Advanced Search**, eseguite con supporto dell'estensione *SERP Snippet Extractor* [24], hanno restituito 28 risultati. Tra questi, **4 fonti sono state considerate rilevanti**, in quanto direttamente utilizzate nella tesi: IBM DOORS Next, Siemens Polarion, Jama Connect e Atlassian Jira. I risultati rilevanti sono evidenziati in Tabella A.4.

Titolo	Fonte	Link	Rilevante
IBM Engineering Requirements Management DOORS Next	IBM	https://www.ibm.com/products/engineering-requirements-management	TRUE
Polarion Requirements	Siemens	https://polarion.plm.siemens.com/products/requirements	TRUE
Jama Connect Requirements Management Software	Jama Software	https://www.jamasoftware.com/solutions/requirements-management	TRUE
Atlassian Jira Software for Requirements Management	Atlassian	https://www.atlassian.com/software/jira	TRUE

Tabella A.4: Fonti di grey literature incluse (Google Advanced Search).

APPENDICE B

Tabelle degli articoli

In questa appendice sono riportate le tabelle di sintesi relative agli articoli scientifici e alle fonti industriali analizzate durante il lavoro di tesi. Le tabelle hanno lo scopo di fornire una visione d'insieme dei contributi considerati, facilitando il confronto tra i diversi approcci e mettendo in evidenza benefici, limiti e implicazioni didattiche.

- La prima tabella raccoglie i principali studi accademici pubblicati tra il 2023 e il 2025 che hanno esplorato l'uso della GenAI nella *Requirements Elicitation* e *Requirements Analysis*.
- La seconda tabella sintetizza invece alcune fonti industriali, con particolare riferimento a soluzioni offerte da Jama, Polarion e IBM.
- Infine, la terza tabella presenta un confronto diretto tra la prospettiva accademica e quella industriale, evidenziando come tali differenze possano tradursi in opportunità di aggiornamento per la didattica dei corsi di Ingegneria del Software.

Tabella B.1: Sintesi delle fonti accademiche analizzate

Fonte	Task RE supportato	Artefatti analizzati	Tecniche AI usate	Benefici osservati	Limiti / Rischi
Cheng et al. (2024) [14]	Elicitation, Analysis, Quality checking	Interviste simulate, documenti, feedback utenti	LLM (GPT-3/4, BERT, T5), NLP tradizionali	Riduzione tempi, miglior chiarezza	Riproducibilità limitata, dataset ridotti, bias
Zhang et al. (2023) [15]	Requirements Information Retrieval	Dataset RE reali	ChatGPT (zero-shot prompting)	Recall elevato, velocità	Precision bassa, risposte verbose, irrilevanti
Ronanki et al. (2023) [16]	Requirements Elicitation (brainstorming, generazione)	Scenari testuali, valutazione esperti	ChatGPT (dialogo guidato)	Requisiti comprensibili, ben strutturati	Ambiguità, incoerenze, mancanza contesto
Habiba et al. (2024) [17]	Meta-analisi maturità RE per AI	72 studi (2013–2023)	Vari approcci	Identificazione pratiche consolidate, gap metodologici	Mancanza framework standardizzati

Tabella B.1: Sintesi delle fonti accademiche analizzate (continua)

Fonte	Task RE supportato	Artefatti analizzati	Tecniche AI usate	Benefici osservati	Limiti / Rischi
Marques et al. (2024) [18]	Rassegna su ChatGPT per SRE	Studi accademici e applicazioni	Analisi secondaria, LLM (ChatGPT)	Supporto elicitation, comunicazione stakeholder	Qualità incostante, dipendenza dal prompt, etica
Norheim et al. (2024) [19]	Sfide generali nell'uso di LLM in RE	Revisione critica della letteratura	LLM (GPT-based), analisi concettuale	Identificazione barriere, definizione research gaps	Mancanza dataset/benchmark, metriche deboli, rischi bias
Lubos et al. (2024) [20]	Quality assurance dei requisiti	Requisiti testuali valutati con ISO 29148	LLM con spiegazioni e suggerimenti	Supporto a revisione, riduzione sforzo di QA	Non sostituisce revisore, limiti trasparenza
Rahman e Zhu (2024) [21]	Generazione automatica di user stories e test	Documenti di requisiti	GPT-4 (GeneUS)	Automazione da requisiti a backlog/test	Qualità dipendente dal contesto, rischi incompletezza

Tabella B.1: Sintesi delle fonti accademiche analizzate (continua)

Fonte	Task RE supportato	Artefatti analizzati	Tecniche AI usate	Benefici osservati	Limiti / Rischi
Bashir et al. (2025) [22]	Ambiguity detection & explanation	Dataset industriali di requisiti	LLM con in-context learning (10-shot)	+20% accuratezza rispetto a zero-shot, spiegazioni naturali	Terminologia non sempre adeguata, generalizzabilità limitata
Ronanki et al. (2025) [23]	Elicitation, Analysis, Specification, Validation, Management	Linee guida sintetizzate; mappatura su attività RE; interviste esperti ($n = 3$)	LLM prompting & prompt-engineering (few-shot, disambiguazione, persona, template)	Schema operativo di linee guida riutilizzabili; supporto alla pratica	Campione qualitativo ridotto; bassa generalizzabilità; assenza di validazione sperimentale

Tabella B.2: Sintesi delle fonti industriali analizzate

Fonte	Task RE supportato	Artefatti / Contesto	Tecniche AI dichiarate	Benefici riportati	Limiti / Criticità
Jama Connect Advisor [25]	Requirements quality checking	Requisiti testuali, specifiche formali	LLM + NLP (INCOSE/EARS)	Riduzione ambiguità, supporto standard, meno rework	Trasparenza algoritmi limitata, tool commerciale
Polarion ALM [26]	Requirements extraction, mapping, tracking	Documenti Word/PDF, tabelle Excel	NLP + AI semantica	Automazione estrazione e tracciabilità, efficienza	Supporto limitato agli utenti Polarion, dipendenza da input
IBM DOORS Next [27]	Interactive elicitation & Q/A assistant	Repository DOORS Next	RAG con Watsonx	Risposte dinamiche, supporto stakeholder, riduzione tempi	Prototipo, problemi di scalabilità, governance necessaria

Tabella B.3: Confronto tra prospettiva accademica e industriale

Dimensione	Accademico	Industriale	Implicazioni didattiche
Focus principale	Analisi di potenzialità, limiti e nuove sfide (es. etica, governance, standard)	Integrazione immediata nei workflow aziendali	Unire approccio critico-analitico e uso pratico degli strumenti
Artefatti analizzati	Dataset accademici, interviste simulate, feedback, documenti di requisiti, dataset industriali (ambiguità)	Documenti aziendali reali (Word/PDF), repository DOORS, specifiche regolamentate	Addestrare studenti a gestire sia dataset di ricerca che casi industriali concreti
Tecniche AI usate	LLM (GPT-3/4), NLP, IR, embedding, in-context learning, generazione automatica di user stories e test, QA su ISO 29148	LLM/NLP integrati nei tool, RAG proprietari, AI semantica embedded nei sistemi ALM	Introdurre laboratori su prompting, RAG e QA con standard RE

Tabella B.3: Confronto tra prospettiva accademica e industriale (continua)

Dimensione	Accademico	Industriale	Implicazioni didattiche
Benefici osservati	Produttività, chiarezza linguistica, supporto brainstorming, automazione backlog/test, miglioramento qualità requisiti (ISO 29148), rilevazione ambiguità	Efficienza nella gestione requisiti, tracciabilità, riduzione ambiguità, supporto standard INCOSE/EARS	Mostrare metriche di produttività (tempo, difetti evitati) e conformità a standard
Limiti/Rischi	Precisione limitata, bias, mancanza dataset/benchmark condivisi, qualità incostante, dipendenza dal prompt, necessità di governance	Trasparenza ridotta, lock-in tecnologico, prototipi acerbi, difficoltà di scalabilità	Insegnare validazione critica, etica, gestione rischi e consapevolezza dei limiti degli strumenti
Maturità	Campo in rapida evoluzione ma frammentato (2023–2025), con gap su framework e linee guida condivise	Tool commerciali già diffusi, ma con funzionalità limitate e opacità negli algoritmi	Aggiornare i corsi con moduli su “AI in RE”, includendo aspetti metodologici e strumenti industriali

APPENDICE C

Dettaglio del processo di selezione

In questa appendice si riportano i dettagli quantitativi del processo di selezione degli studi, organizzati secondo le fasi previste dalle linee guida di **Garousi et al. (2019)** [1]. Il *flowchart* sintetico del processo è mostrato in Fig. 3.1 (Cap. 3).

Tabella C.1: Conteggi per fase del processo di selezione.

Fase	Conteggio	Note
Identificazione (query su database)	212	arXiv: 106; Scholar: 78; Scopus: 28
Rimozione duplicati	36	Duplicati eliminati (automatico + manuale); restano 176
Screening titolo/abstract	105	Esclusi per non pertinenza; restano 71
Lettura full-text	36	Esclusi per criteri di inclusione/esclusione; restano 35
Inclusione finale	35	Studi analizzati: 31 accademici, 4 industriali

APPENDICE D

Valutazione della qualità

In conformità con le linee guida di **Garousi et al. (2019)** [1], è stata adottata una griglia di valutazione della qualità per gli studi inclusi nella revisione, al fine di garantire rigore metodologico e trasparenza. Essa è stata applicata successivamente all'estrazione dei dati dal codebook.

Criteri di qualità (adattati)

I criteri originali individuati da Garousi et al. per l'assessment delle fonti (formali e grey) includono:

1. **Authority of the producer:** autorità/autorevolezza dell'autore o dell'organizzazione che produce la fonte.
2. **Methodology:** chiarezza dello scopo, metodologia descritta, casi/campione, riferimenti autorevoli.
3. **Objectivity:** equilibrio della presentazione, mancanza di conflitti di interesse, conclusioni supportate dai dati.
4. **Date:** data di pubblicazione chiara e leggibile.

5. **Position rispetto ad altre fonti correlate:** comparazione, citazioni, contestualizzazione.
6. **Novelty:** contributo originale, se rafforza o contraddice ciò che esiste.
7. **Impact:** visibilità, citazioni o percezione, uso pratico della fonte.
8. **Outlet type:** tipo di pubblicazione / credibilità / controllo del peer review / controllo editoriale.

Checklist adottata

Per la revisione, la checklist è stata ridotta a 5 criteri principali, più adatti al dominio del Requirements Engineering con AI/GenAI, come segue:

Tabella D.1: Criteri di qualità adottati

Criterio (C1–C5)	Domanda di valutazione
C1	Lo studio formula chiaramente obiettivi e domande di ricerca?
C2	La metodologia adottata è adeguata e descritta in modo replicabile?
C3	Sono riportati dati quantitativi o qualitativi a supporto delle conclusioni?
C4	Vengono discussi i limiti e i <i>threats to validity</i> ?
C5	Lo studio è rilevante per attività di RE con AI/GenAI?

Ogni studio è stato valutato su questi 5 criteri, usando scala a tre livelli:

- **1** = criterio soddisfatto pienamente,
- **0.5** = criterio parzialmente soddisfatto,
- **0** = criterio non soddisfatto.

Il punteggio totale per ogni studio è compreso tra 0 e 5. La valutazione aggregata ha prodotto: media pari a **3.7**, deviazione standard pari a **0.6**. Sono stati individuati **7 studi** con punteggio **borderline** (compreso tra 3.0 e 3.5), che sono stati comunque inclusi nell'analisi.

Distribuzione dei punteggi per criterio

Tabella D.2: Distribuzione dei punteggi ottenuti dagli studi sui criteri C1–C5

Criterio	0	0.5	1
C1	2	6	27
C2	3	8	24
C3	5	10	20
C4	11	13	11
C5	1	4	30

Questa valutazione ha permesso di uniformare i criteri di qualità, aiutando a minimizzare la soggettività nella selezione e aumentando la trasparenza metodologica del lavoro.

APPENDICE E

Esempi di schede di estrazione

Questa appendice riporta due esempi di schede di estrazione compilate secondo il codebook operativo definito nel Capitolo 3 (Tabella 3.1). Gli esempi hanno valore illustrativo e mostrano come il processo sia stato applicato sia a studi accademici peer-reviewed che a contributi di *grey literature*.

E.1 Esempio di studio accademico

Campo	Valore estratto
Anno e tipologia	2023, articolo conferenza
Obiettivi e contesto	Valutare ChatGPT in scenari di <i>Requirements Information Retrieval</i> (RE-IR).
Task di RE	Analysis
Tecniche AI/GenAI	LLM (ChatGPT), prompting
Benefici	Buona capacità di recupero dei requisiti (alto recall); supporto nella fase di analisi.
Limiti	Precisione ridotta; risposte verbose e contenenti rumore; sensibilità al prompt.
Confronto baseline	Sì – confronto con tradizionali tecniche di Information Retrieval.
RQ principale	RQ2
Flag RQ (copertura)	[0, 1, 0, 0, 0]

Tabella E.1: Scheda di estrazione relativa a un articolo accademico (Zhang et al., 2023)

E.2 Esempio di contributo industriale (grey literature)

Campo	Valore estratto
Anno e tipologia	2024, white paper (Polarion/Siemens)
Obiettivi e contesto	Presentare funzionalità sperimentali di Generative AI integrate in Polarion ALM per il supporto alla scrittura dei requisiti.
Task di RE	Specification, Validation
Tecniche AI/GenAI	LLM integrati via API; generazione automatica di requisiti; suggerimenti per miglioramento qualità.
Benefici	Riduzione tempi di scrittura; supporto alla standardizzazione dei requisiti; miglioramento leggibilità.
Limiti	Mancanza di trasparenza sul modello usato; risultati variabili; dipendenza dall'infrastruttura cloud del vendor.
Confronto baseline	No – assenza di benchmark pubblici; confronto solo con workflow manuale tradizionale.
RQ principale	RQ3
Flag RQ (copertura)	[0, 0, 1, 1, 0]

Tabella E.2: Scheda di estrazione relativa a un white paper industriale (Polarion, 2024)

APPENDICE F

Pilot di validazione del codebook

Per validare il codebook è stato condotto un pilot su un campione ridotto di cinque studi, selezionati tra quelli inclusi nella SMLR. L'obiettivo era verificare la chiarezza dei campi e la coerenza delle regole decisionali.

Studio	Task prevalente	Codebook output	Note
Zhang et al. (2023)	Analysis	Analysis	Concordanza totale. Confermata la regola recall/precisione.
Ronanki et al. (2023)	Elicitation	Elicitation	Concordanza totale. Nessuna ambiguità rilevata.
Habiba et al. (2024)	Analysis	Analysis	Concordanza. Chiarita la differenza con Specification.
Polarion White Paper (2024)	Specification	Specification	Concordanza. Nota sul limite: mancanza baseline.
Bashir et al. (2025)	Validation	Validation	Concordanza. Verificata regola su in-context learning.
Geyer (2025) [34]	Elicitation	Validation	Disallineamento: lo studio si presenta come elicitation, ma l'uso prevalente della GenAI è la valutazione della qualità dei requisiti generati. Affinata la regola per distinguere tra raccolta e valutazione.
Cheliger (2022) [35]	Elicitation	Analysis	Borderline: review orientata a elicitation, ma diversi studi inclusi implementano tecniche di classificazione automatica (analysis). Necessaria regola per distinguere raccolta da elaborazione.

Tabella F.1: Risultati del pilot di validazione del codebook (campione di 7 studi, inclusi casi borderline)

Il pilot evidenzia come nella maggior parte dei casi il codebook abbia fornito una classificazione coerente con l'interpretazione attesa (cinque studi su sette). Tuttavia, due contributi hanno mostrato situazioni borderline che hanno richiesto un affinamento delle regole decisionali.

Lo studio di **Geyer (2025)** [34] (link) si presenta come un lavoro di *elicitation*, ma in realtà l'uso prevalente della GenAI riguarda la valutazione dei requisiti generati, il che lo rende più coerente con la categoria *validation*. Questo caso ha portato ad esplicitare la regola per distinguere tra raccolta dei requisiti e loro successiva valutazione.

Analogamente, la review di **Cheliger (2022)** [35] (link) è orientata a *elicitation*, ma include diversi studi che implementano tecniche di classificazione automatica riconducibili ad *analysis*. Ciò ha reso necessario chiarire la distinzione tra raccolta dei requisiti e loro elaborazione tramite metodi computazionali.

Questi casi borderline confermano la solidità del codebook e mostrano come il pilot abbia consentito di raffinare le regole decisionali per gestire ambiguità e sovrapposizioni.

Bibliografia

- [1] V. Garousi, M. Felderer, and M. V. Mäntylä, “Guidelines for including grey literature and conducting multivocal literature reviews in software engineering,” *Information and Software Technology*, vol. 106, pp. 101–121, 2019. [Online]. Available: <https://doi.org/10.1016/j.infsof.2018.09.006> (Citato alle pagine 1, 3, 4, 20, 24, 25, 29, 34, 72, 73, 74, 80, 91 e 92)
- [2] *ISO/IEC/IEEE 29148:2018 Systems and software engineering — Life cycle processes — Requirements engineering*, International Organization for Standardization Std., 2018, international Standard. [Online]. Available: <https://www.iso.org/standard/72089.html> (Citato alle pagine 1, 2, 3, 4, 6, 9, 12, 15, 20, 22, 42, 46, 48, 50, 55, 65 e 68)
- [3] International Council on Systems Engineering (INCOSE), *INCOSE Guide for Writing Requirements*, 3rd ed., San Diego, CA, USA, 2023, iNCOSE-TP-2010-006-03. [Online]. Available: <https://incose.org/products-and-publications/se-store/incose-guide-for-writing-requirements> (Citato a pagina 2)
- [4] B. Kitchenham, “Procedures for performing systematic reviews,” Keele University and National ICT Australia, Tech. Rep. TR/SE-0401, 2004. (Citato alle pagine 4, 20, 21, 24, 72 e 74)

- [5] B. Kitchenham and S. Charters, “Guidelines for performing systematic literature reviews in software engineering,” EBSE, Tech. Rep. EBSE-2007-01, 2007. (Citato alle pagine 4, 20, 21, 24, 72 e 74)
- [6] B. Kitchenham, D. Budgen, and P. Brereton, *Evidence-Based Software Engineering and Systematic Reviews*. CRC Press, 2015. (Citato alle pagine 4, 20, 21, 24, 72 e 74)
- [7] I. Sommerville, *Software Engineering*, 10th ed. Boston, MA, USA: Pearson, 2015. (Citato a pagina 8)
- [8] A. M. Davis, *Software Requirements: Objects, Functions, and States*. Upper Saddle River, NJ, USA: Prentice Hall, 1993. (Citato a pagina 8)
- [9] I. Kotonya and I. Sommerville, *Requirements Engineering: Processes and Techniques*. Chichester, UK: John Wiley & Sons, 1998. (Citato a pagina 8)
- [10] A. M. Turing, “Computing machinery and intelligence,” *Mind*, vol. 59, no. 236, pp. 433–460, 1950. (Citato a pagina 9)
- [11] J. McCarthy, M. Minsky, N. Rochester, and C. Shannon, “A proposal for the dartmouth summer research project on artificial intelligence,” in *Dartmouth Conference Proposal*, 1955, hanover, New Hampshire, USA. [Online]. Available: <http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf> (Citato a pagina 9)
- [12] M. Minsky, *Steps Toward Artificial Intelligence*. Prentice Hall, 1961, reprinted in: Marvin L. Minsky, *Semantic Information Processing*, MIT Press, 1968. (Citato a pagina 9)
- [13] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, “Attention is all you need,” in *Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS 2017)*. Curran Associates Inc., 2017, pp. 6000–6010. [Online]. Available: <https://arxiv.org/abs/1706.03762> (Citato a pagina 10)

-
- [14] H. Cheng, J. H. Husen, Y. Lu, T. Racharak, N. Yoshioka, N. Ubayashi, and H. Washizaki, "Generative ai for requirements engineering: A systematic literature review," 2024, arXiv:2409.06741. (Citato alle pagine 14, 17, 35, 40, 51, 52, 54 e 85)
- [15] J. Zhang, Y. Chen, N. Niu, Y. Wang, and C. Liu, "Empirical evaluation of chatgpt on requirements information retrieval under zero-shot setting," 2023, arXiv:2304.12562. (Citato alle pagine 14, 17, 35, 40, 43, 44, 57, 58, 69 e 85)
- [16] K. Ronanki, C. Berger, and J. Horkoff, "Investigating chatgpt's potential to assist in requirements elicitation processes," 2023, arXiv:2310.13976. (Citato alle pagine 15, 17, 35, 40, 43, 44, 48, 50 e 85)
- [17] U.-e. Habiba, M. Haug, J. Bogner, and S. Wagner, "How mature is requirements engineering for ai-based systems? a systematic mapping study on practices, challenges, and future research directions," *Requirements Engineering*, vol. 29, pp. 567–600, 2024. (Citato alle pagine 15, 18, 35, 40, 49, 51 e 85)
- [18] N. Marques, R. R. Silva, and J. Bernardino, "Using chatgpt in software requirements engineering: A comprehensive review," *Future Internet*, vol. 16, no. 6, 2024. [Online]. Available: <https://www.mdpi.com/1999-5903/16/6/180> (Citato alle pagine 15, 17, 35, 40, 43, 44 e 86)
- [19] J. J. Norheim, E. Rebentisch, D. Xiao, L. Draeger, A. Kerbrat, and O. L. de Weck, "Challenges in applying large language models to requirements engineering tasks," *Design Science*, vol. 10, p. e16, 2024. [Online]. Available: <https://www.cambridge.org/core/journals/design-science/article/challenges-in-applying-large-language-models-to-requirements-engineering-tasks/1FC7666F0A0B4E7091D2D4B2D46321B5> (Citato alle pagine 15, 18, 36, 40 e 86)
- [20] S. Lubos, A. Felfernig, T. N. T. Tran, D. Garber, M. El Mansi, S. Polat Erdeniz, and V.-M. Le, "Leveraging llms for the quality assurance of software requirements," 2024. [Online]. Available: <https://arxiv.org/abs/2408.10886> (Citato alle pagine 15, 18, 36, 40, 49, 51 e 86)
- [21] T. Rahman and Y. Zhu, "Automated user story generation with test case specification using large language model," 2024. [Online]. Available:

- <https://arxiv.org/abs/2404.01558> (Citato alle pagine 16, 36, 41, 43, 44, 57, 58, 69 e 86)
- [22] S. Bashir, A. Ferrari, A. Khan, P. E. Strandberg, Z. Haider, M. Saadatmand, and M. Bohlin, "Requirements ambiguity detection and explanation with llms: An industrial study," in *2025 IEEE International Conference on Software Maintenance and Evolution (ICSME), Industry Track*, 2025, preprint available at MDU IPR. [Online]. Available: https://www.ipr.mdu.se/pdf_publications/7221.pdf (Citato alle pagine 16, 18, 36, 41, 43, 44, 47, 50, 57, 58, 69 e 87)
- [23] K. Ronanki, S. Arvidsson, and J. Axell, "Prompt engineering guidelines for using large language models in requirements engineering," 2025, presented at Euromicro 2025. [Online]. Available: <https://arxiv.org/abs/2507.03405> (Citato alle pagine 16, 18, 36, 41 e 87)
- [24] getindexed.io, "Serp snippet extractor," <https://chromewebstore.google.com/detail/serp-snippet-extractor-by/gcigkdheghgmadikicbaahpgklohlikn>, 2024, browser extension for extracting titles and links from Google search results. (Citato alle pagine 21, 79 e 82)
- [25] Jama Software, "Jama connect advisor: Ai-powered guidance for precise, high-quality requirements," <https://www.jamasoftware.com/solutions/artificial-intelligence/>, 2025, iNCOSE/EARS-based AI for reducing ambiguity and defects. (Citato alle pagine 41, 65 e 88)
- [26] Siemens Polarion, "Ai-driven automation in polarion alm: requirement extraction, mapping, tracking," <https://blogs.sw.siemens.com/polarion/enhancing-efficiency-in-polarion-with-ai-driven-requirements-management/>, 2025, semantic-level analysis for improved productivity. (Citato alle pagine 41, 43, 45, 48, 65 e 88)
- [27] IBM, "Requirements intelligence assistant for doors next using watsonx (prototype)," <https://community.ibm.com/community/user/blogs/daniel-moul/2025/02/13/requirements-ai-assistant-for-doors-next>, 2025, rAG-style AI assistant embedded in DOORS Next. (Citato alle pagine 41, 65 e 88)

-
- [28] C. M. University, “17-627 requirements for embedded systems — course syllabus,” https://mse.s3d.cmu.edu/0_documents/syllabi/fa2025/17627.pdf, 2025, accessed: 2025-09-21. (Citato alle pagine 63 e 64)
- [29] A. S. University, “Cse 563 foundations of software engineering — course syllabus,” <https://webapp4.asu.edu/bookstore/viewsyllabus/2241/26502/pdf>, 2024, accessed: 2025-09-21. (Citato alle pagine 63 e 64)
- [30] T. A. University-Texarkana, “Systems analysis & design — course syllabus,” <https://www.tamut.edu/faculty/syllabi/202580/80260.pdf>, 2025, accessed: 2025-09-21. (Citato alle pagine 63 e 64)
- [31] F. A. University, “Cot6930: Generative ai and software development lifecycles — course syllabus,” <https://www.fau.edu/engineering/eecs/pdf/syllabus-generative-ai-driven-software-development-fall-2024.pdf>, 2024, accessed: 2025-09-21. (Citato alle pagine 63 e 64)
- [32] M. S. U. of Denver, “Syllabus spectrum for generative ai: Language and considerations,” <https://www.msudenver.edu/wp-content/uploads/2023/08/Syllabus-Spectrum-for-Generative-AI-Final.pdf>, 2023, accessed: 2025-09-21. (Citato alle pagine 63 e 64)
- [33] A.-W. Harzing, “Publish or perish,” <https://harzing.com/resources/publish-or-perish>, 2007, software. (Citato alle pagine 72 e 80)
- [34] W. Geyer, “A case study investigating the role of generative ai in requirements engineering,” *Proceedings of the ACM on Human-Computer Interaction*, vol. 9, no. CSCW, pp. 1–23, 2025. (Citato alle pagine 99 e 100)
- [35] C. Cheliger *et al.*, “Machine learning in requirements elicitation: a literature review,” *AI EDAM*, vol. 36, no. 4, pp. 1015–1033, 2022. (Citato alle pagine 99 e 100)

Ringraziamenti
