UNIVERSITÀ POLITECNICA DELLE MARCHE
FACOLTÀ DI ECONOMIA "GIORGIO FUÀ"

CORSO DI LAUREA MAGISTRALE IN: SCIENZE ECONOMICHE E FINANZIARIE

INTELLIGENZA ARTIFICIALE E STRATEGIE DI INVESTIMENTO

Il contributo dell'IA alla generazione di proposte di allocazione di portafoglio

ARTIFICIAL INTELLIGENCE AND INVESTMENT STRATEGIES

AI's Contribution to Generating Portfolio Allocation Proposals

Studente:
MATTEO SEMPRINI

Relatrice:
PROF.SSA CATERINA LUCARELLI

ANNO ACCADEMICO 2025-2026

SOMMARIO

INTRODUZIONE 4

CAPITOLO 1 INTELLIGENZA ARTIFICIALE E FINANZA 6

1.1 Intelligenza artificiale e settore finanziario 7

1.2 Diffusione e utilizzo dell'AI nei mercati finanziari 10

1.3 Obiettivo e domanda di ricerca 15

CAPITOLO 2 INTELLIGENZA ARTIFICIALE E STRATEGIE DI INVESTIMENTO 16

2.1 Asset allocation e benchmark di investimento 17

2.2 Strategie quantitative e Machine Learning 19

2.3 GenAI e portfolio selection 22

2.4 Limiti metodologici: overfitting e bias nei backtest 23

2.5 Interpretabilità e Explainable AI 27

CAPITOLO 3 METODOLOGIA 30

3.1 Dataset e universo investibile 31

3.2 Costruzione dei portafogli benchmark 34

3.3 Generazione di strategie tramite intelligenza artificiale 36

3.4 Metodologia di backtesting e metriche di valutazione 39

CAPITOLO 4 ANALISI EMPIRICA 42

4.1 Implementazione delle strategie 42
4.2 Risultati dei portafogli benchmark 45
4.3 Risultati delle strategie generate dall'intelligenza artificiale 48
4.4 Confronto tra strategie 52
CAPITOLO 5 DISCUSSIONE 56
5.1 Interpretazione dei risultati 56
5.2 Limiti del metodo 65
5.3 Implicazioni per l'asset management 67
5.4 Verifica di riproducibilità 69
CONCLUSIONI 73
BIBLIOGRAFIA 76
SITOGRAFIA 81
APPENDICE A - OUTPUT INTEGRALI DEI MODELLI 82
A.1 ChatGPT (OpenAI) 82
A.2 Claude (Anthropic) 85
A.3 Gemini (Google) 89
A.4 - Rigenerazione (prompt in italiano) 93
A.4.1 ChatGPT (OpenAI) 93
A.4.2 Claude (Anthropic) 96
A.4.3 Gemini (Google) 100
A.5 - Rigenerazione (prompt in inglese) 103
A.5.1 ChatGPT (OpenAI) 103
A.5.2 Claude (Anthropic) 107

A.5.3 Gemini (Google) 110
APPENDICE B - CODICE 114
B.0 - Istruzioni di replica 114
B.1 - carica_dati.py 115
B.2 - backtest.py 119
B.3 - backtest_benchmark.py 127
B.4 - backtest_portafogli_ai.py 129
B.5 - grafici.py 134
B.6 - backtest_rigenerazione.py 144
APPENDICE C - LOG DELLE CORREZIONI SUI DATI 149
APPENDICE D - VERIFICA DI ROBUSTEZZA SUI BENCHMARK DI CONFRONTO 151
D.1 - Costruzione dei benchmark alternativi 152
D.2 - Risultati 153
D.3 - Limiti della verifica 155
RINGRAZIAMENTI 156

INTRODUZIONE

Negli ultimi anni l'intelligenza artificiale generativa ha assunto sempre più un ruolo rilevante nei mercati finanziari, offrendo strumenti in grado di elaborare grandi quantità di dati, in tempi relativamente brevi, e di produrre risposte articolate a partire da istruzioni fornite in un linguaggio naturale. Nonostante ciò, l'utilizzo dell'AI nei processi decisionali di investimento rimane tuttora limitato: le rilevazioni più recenti mostrano che le applicazioni più vicine alle decisioni di investimento rimangono una parte residuale rispetto ai casi d'uso censiti nel settore finanziario, a fronte di un maggior utilizzo in attività di supporto operativo.

È proprio da questo divario – tra il potenziale attribuito all'AI e il suo utilizzo ancora circoscritto nei processi di allocazione del capitale – che nasce il presente lavoro.

L'obiettivo della tesi è verificare, attraverso un'analisi sperimentale, se ed in che misura l'intelligenza artificiale generativa possa essere di supporto alle decisioni di allocazione del capitale. In particolare si impiegano tre modelli linguistici di ampia diffusione commerciale (ChatGPT, Claude, Gemini) al fine di verificare se questi siano in grado di generare proposte di allocazione strategica in grado di ottenere rendimenti aggiustati per il rischio superiori, o comparabili, rispetto a quelli di benchmark tradizionali costruiti sullo stesso universo investibile. Accanto a questa domanda principale, il lavoro affronta due questioni più specifiche: quali limiti metodologici - in particolare overfitting e bias nei backtest – riguardano le strategie generate dai modelli di intelligenza

artificiale e quanto tali strategie risultino interpretabili e spiegabili, una volta generate.

Per rispondere a queste domande è stato adottato un disegno sperimentale che prevede la costruzione di nove portafogli – tre modelli, tre profili di rischio ciascuno – a partire da un prompt congelato e condiviso, il loro confronto con quattro benchmark tradizionali, costruiti sullo stesso universo investibile e la valutazione tramite backtest su un decennio di dati storici (2016-2025).

Il lavoro è organizzato come segue. Il Capitolo 1 introduce il contesto e la domanda di ricerca. Il Capitolo 2 ripercorre la letteratura di riferimento, dalla teoria classica dell'asset allocation fino ai contributi più recenti sull'impiego dei modelli di intelligenza artificiale generativa a supporto del settore finanziario. Il Capitolo 3 descrive la metodologia sperimentale. Il Capitolo 4 presenta l'implementazione delle strategie e fornisce i risultati dei backtest. Il Capitolo 5 discute criticamente tali risultati, ne riconosce i limiti e ne verifica, in parte, la riproducibilità. Le Conclusioni, infine, raccolgono le risposte alla domanda di ricerca posta in apertura.

Capitolo 1

INTELLIGENZA ARTIFICIALE E FINANZA

Negli ultimi anni il settore finanziario è stato protagonista di una profonda evoluzione digitale, sostenuta dalla crescente disponibilità di dati, dal costante incremento della capacità di calcolo e dal rapido sviluppo di tecnologie sempre più sofisticate. Questo processo ha contribuito a rafforzare il ruolo dell'analisi quantitativa all'interno dei processi decisionali degli intermediari finanziari, rendendo possibile l'elaborazione di grandi quantità di informazioni e offrendo un supporto sempre più rilevante ad attività quali la gestione del rischio, la valutazione degli investimenti e l'allocazione del capitale.

L'utilizzo dell'intelligenza artificiale nei mercati finanziari, in realtà, non costituisce una novità. Da diversi anni, infatti, tecniche di machine learning e altri algoritmi per l'analisi dei dati trovano applicazione in numerosi ambiti della finanza. Negli ultimi tempi, però, la diffusione dei Large Language Models (LLMs), lo sviluppo dell'intelligenza artificiale generativa (Generative AI o GenAI) e il continuo progresso delle capacità computazionali hanno rappresentato un importante punto di svolta. Queste innovazioni hanno ampliato in misura significativa le possibilità offerte dall'intelligenza artificiale, alimentando un crescente interesse non solo tra gli operatori di mercato, ma anche da parte delle autorità di vigilanza e della comunità scientifica. In particolare, l'intelligenza artificiale generativa può essere intesa come una categoria di sistemi di intelligenza artificiale in grado di produrre nuovi

contenuti - come testi, immagini, audio o video - a partire dai dati su cui sono stati addestrati e in risposta a istruzioni fornite dall'utente, generalmente sotto forma di prompt[1]. Questa innovazione ha reso concretamente realizzabili, anche su larga scala, applicazioni che fino a pochi anni fa risultavano difficili da implementare, favorendo un'accelerazione dei processi di adozione e incentivando nuovi investimenti nel settore finanziario.

In questo quadro, l'intelligenza artificiale rappresenta oggi una delle innovazioni tecnologiche con il maggiore potenziale di incidere sul funzionamento dei mercati finanziari. Alle opportunità legate all'aumento dell'efficienza operativa, alla capacità di elaborare grandi volumi di dati e al miglioramento del supporto ai processi decisionali si affiancano, però, nuove criticità, che riguardano la qualità dei dati, la trasparenza dei modelli, la gestione dei rischi operativi e l'evoluzione del quadro regolamentare.

Il presente capitolo si propone quindi di delineare il contesto teorico nel quale si inserisce il lavoro di ricerca, esaminando il rapporto tra intelligenza artificiale e settore finanziario, la diffusione delle principali applicazioni nei mercati finanziari e le ragioni che rendono particolarmente rilevante l'analisi dell'impiego dell'intelligenza artificiale nella costruzione delle strategie di investimento.

1.1 Intelligenza artificiale e settore finanziario

La crescente digitalizzazione dei mercati finanziari ha determinato un forte aumento della quantità di dati prodotti e disponibili, rendendo necessari

[1] Philippe Lorenz et al., *Initial Policy Considerations for Generative Artificial Intelligence*, no. 1, OECD Artificial Intelligence Papers (OECD Publishing, 2023), https://doi.org/10.1787/fae2d1e6-en.

strumenti in grado di elaborare informazioni complesse in tempi rapidi e di supportare processi decisionali caratterizzati da forte incertezza.

L'utilizzo di strumenti computazionali, ovviamente, precede l'utilizzo dell'intelligenza artificiale all'interno dei processi decisionali degli investitori. Già, in precedenza, si faceva ricorso a modelli quantitativi, tecniche statistiche e algoritmi di machine learning per supportare l'attività di investimento attraverso la valutazione del rischio e l'automazione delle negoziazioni: un fenomeno che le istituzioni internazionali monitorano e documentano da tempo[2]. Il recente sviluppo dei Large Language Models (LLMs) e dell'intelligenza artificiale generativa ha ampliato le applicazioni dell'AI nei mercati finanziari, rendendo possibile l'automazione di processi complessi, l'analisi di dati non strutturati e il supporto alla generazione di nuove proposte di investimento[3].

Più nello specifico, la caratteristica distintiva dell'intelligenza artificiale generativa è la capacità di produrre contenuti nuovi a partire da istruzioni formulate in linguaggio naturale, utilizzando le regolarità apprese dai dati di addestramento. Rispetto ai modelli tradizionali, orientati prevalentemente alla classificazione o alla previsione, questi sistemi possono generare testi, sintetizzare documenti, rielaborare informazioni non strutturate e supportare l'utente nella formulazione di analisi o proposte di investimento. L'interesse nei confronti dell'intelligenza artificiale è riconducibile alle caratteristiche stesse dell'attività finanziaria. I mercati producono quotidianamente enormi quantità di dati eterogenei, provenienti da prezzi di mercato, bilanci societari, indicatori macroeconomici, notizie finanziarie, comunicazioni delle imprese e numerose

[2] Financial Stability Board, *Artificial Intelligence and Machine Learning in Financial Services: Market Developments and Financial Stability Implications* (Financial Stability Board, 2017).

[3] International Monetary Fund, "Advances in Artificial Intelligence: Implications for Capital Market Activities," in *Global Financial Stability Report, October 2024: Steadying the Course: Uncertainty, Artificial Intelligence, and Financial Stability* (International Monetary Fund, 2024), https://www.imf.org/-/media/files/publications/gfsr/2024/october/english/textrevised.pdf.

altre fonti informative. La capacità di acquisire, elaborare e interpretare tali informazioni in tempi estremamente ridotti costituisce un elemento fondamentale per il conseguimento di vantaggi competitivi. In questo scenario, gli strumenti tradizionali di analisi risultano spesso insufficienti a cogliere la complessità delle relazioni esistenti tra le variabili economico-finanziarie, mentre le tecniche di intelligenza artificiale consentono di individuare pattern complessi, relazioni non lineari e informazioni difficilmente rilevabili mediante approcci convenzionali[4].

La crescente integrazione tra intelligenza artificiale, data science e machine learning ha progressivamente trasformato il modo in cui vengono progettati prodotti e servizi finanziari. Tali tecnologie stanno modificando non soltanto gli strumenti operativi impiegati dagli intermediari, ma anche gli obiettivi, i modelli organizzativi e i paradigmi dell'intero ecosistema economico-finanziario, contribuendo allo sviluppo di una finanza sempre più intelligente, personalizzata e basata sui dati[5]. In tale prospettiva, l'intelligenza artificiale non rappresenta esclusivamente un insieme di tecniche computazionali, ma costituisce un fattore abilitante capace di ridefinire numerosi processi caratteristici del settore finanziario. A tali applicazioni si sono recentemente affiancati strumenti basati sull'intelligenza artificiale generativa, utilizzati per la produzione automatica di documenti, la sintesi di informazioni, il supporto agli operatori attraverso assistenti virtuali e l'elaborazione di grandi quantità di dati non strutturati. Da quanto detto finora emerge un ruolo sempre più centrale dell'intelligenza artificiale nel settore finanziario, anche se resta da capire quanto questa

[4] Söhnke M. Bartram et al., *Artificial Intelligence in Asset Management*, Literature Review (CFA Institute Research Foundation, 2020), https://doi.org/10.2139/ssrn.3692805.

[5] Longbing Cao, "AI in Finance: Challenges, Techniques, and Opportunities," *ACM Computing Surveys* 55, no. 3 (2022): 1–38, https://doi.org/10.1145/3502289.

diffusione sia, in concreto, già avvenuta. È quello che si prova a fare nel paragrafo successivo, sulla base dei dati raccolti dal report ESMA.

1.2 Diffusione e utilizzo dell'AI nei mercati finanziari

A livello globale, le evidenze disponibili mostrano un'ampia diffusione dell'intelligenza artificiale, ma ancora disomogenea nella capacità di generare impatti organizzativi ed economici misurabili. Il McKinsey Global Survey 2025[6] rileva che l'88% degli intervistati dichiara un utilizzo regolare di strumenti di AI in almeno una funzione aziendale, ma solo circa un terzo delle organizzazioni afferma di avere iniziato a scalarne l'impiego a livello aziendale. Questo suggerisce che l'adozione dell'AI sia ormai diffusa, ma che il passaggio dalla sperimentazione all'integrazione stabile nei processi decisionali rimanga ancora incompleto.

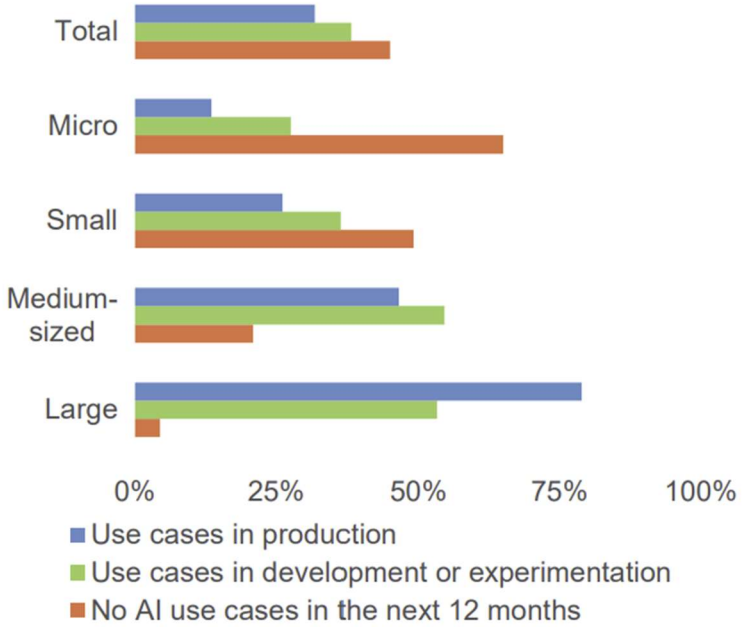
Nel settore finanziario europeo, un quadro più specifico è offerto dai dati raccolti da ESMA[7], che nell'estate del 2025 ha condotto un'indagine su 728 operatori finanziari di 19 paesi dell'Unione Europea - tra gestori di investimento, imprese di investimento, enti creditizi, infrastrutture di mercato e agenzie di rating[8].

[6] Alex Singla et al., *The State of AI in 2025: Agents, Innovation, and Transformation* (McKinsey & Company, 2025), https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai#/.

[7] European Securities and Markets Authority, *AI Adoption and Trends in Securities Markets: EU Evidence*, ESMA TRV Risk Analysis – Financial Innovation nos. ESMA50-481369926–30599 (European Securities and Markets Authority, 2026).

[8] La partecipazione all'indagine è stata su base volontaria ed i tassi di partecipazione sono stati eterogenei tra paesi e settori. I risultati non sono pertanto necessariamente rappresentativi della popolazione complessiva delle imprese domiciliate in UE (ESMA, 2026).

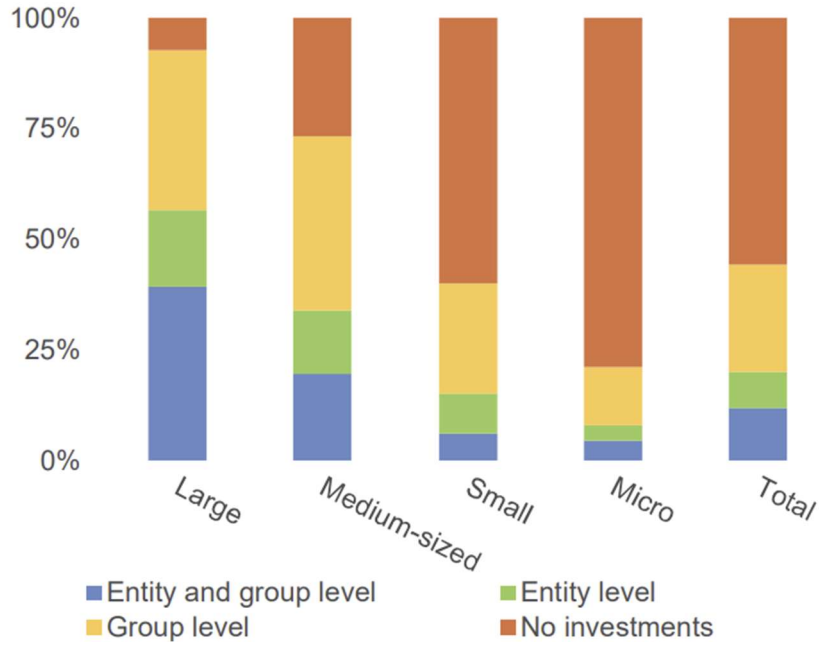
Figura 1.1: Adozione dell'intelligenza artificiale per dimensione d'impresa



Fonte: ESMA (2026), *AI adoption and trends in securities markets: EU evidence*, Chart 1, dati NCAs ed ESMA.

Il quadro che emerge è quello di un'adozione ancora parziale e piuttosto disomogenea. Quasi la metà delle imprese intervistate (45%) dichiara di non avere alcun caso d'uso dell'AI, né in produzione né in fase di sperimentazione - ma il dato cambia molto a seconda della dimensione dell'impresa. Tra le grandi imprese, ormai quasi tutte (96%) hanno già adottato l'AI o stanno pianificando di farlo; tra le micro-imprese, invece, il 65% non ha alcun progetto né attivo né pianificato per i prossimi dodici mesi.

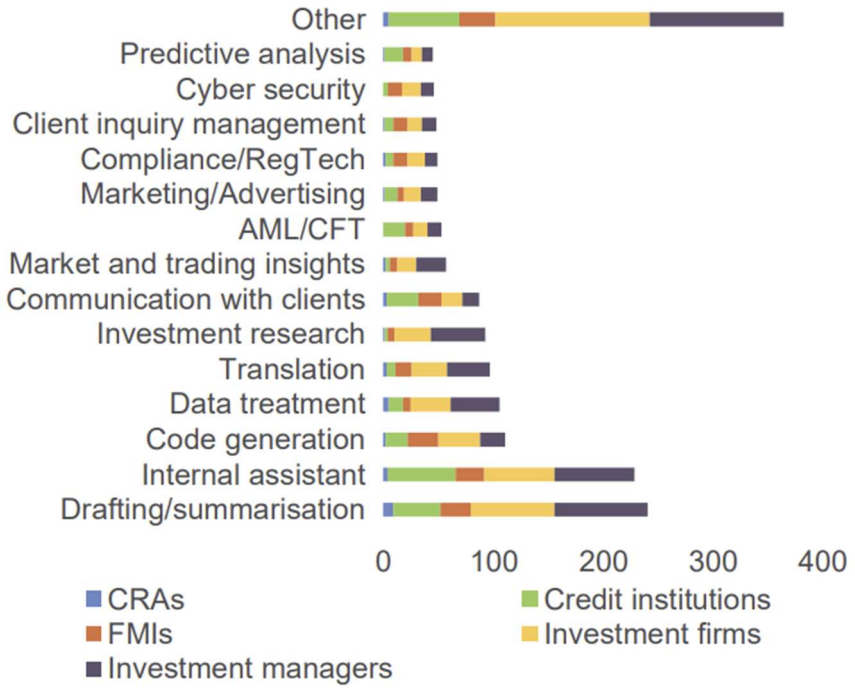
Figura 1.2: Investimenti in tecnologie AI per dimensione d'impresa nell'Unione europea nel 2024



Fonte: ESMA (2026), *AI adoption and trends in securities markets: EU evidence*, Chart 2, dati NCAs ed ESMA.

Lo stesso divario si ritrova negli investimenti: nel 2024 ha investito in AI solo il 44% del campione, ma si passa dal 93% delle grandi imprese ad appena il 21% delle micro-imprese. Le prospettive restano comunque positive: il 70% delle imprese intervistate prevede di aumentare gli investimenti in AI tra il 2025 e il 2027.

Figura 1.3: Casi d'uso dell'intelligenza artificiale nei mercati finanziari europei



Fonte: ESMA (2026), *AI adoption and trends in securities markets: EU evidence*, Chart 4, dati NCAs ed ESMA.

Guardando ai casi d'uso concreti (Figura 1.3), l'immagine che emerge è quella di un'AI usata soprattutto per attività di supporto interno, a bassa autonomia decisionale[9]. Sugli 847 casi d'uso censiti nell'indagine, le voci più ricorrenti sono la redazione e sintesi di documenti (29%) e gli assistenti virtuali interni (27%), seguite da generazione di codice e trattamento dati. Le applicazioni più vicine alle decisioni di investimento vere e proprie - gestione del rischio di portafoglio, ottimizzazione, trading algoritmico - restano invece

[9] Per "bassa autonomia decisionale" ESMA (2026) intende i casi d'uso in cui è richiesta un'approvazione umana o in cui il sistema di AI fornisce esclusivamente suggerimenti, senza operare autonomamente.

una nicchia, poco più del 6% del totale[10]. Non stupisce che la GenAI sia la tecnologia dominante, indicata nel 71% dei casi d'uso, mentre il natural language processing compare nel 28% dei casi. È proprio in questo spazio ancora poco esplorato - l'AI a supporto diretto delle decisioni di investimento - che si inserisce l'oggetto di questa tesi.

Ai benefici percepiti corrispondono però rischi altrettanto concreti. Il vantaggio più citato dalle imprese è la capacità di elaborare grandi volumi di dati (75%), seguito dal miglioramento dei processi interni (67%) e dalla riduzione dei costi (63%). Sul fronte opposto, la maggior parte delle imprese (82%) segnala almeno un rischio legato ai dati - soprattutto protezione (59%) e qualità (55%) - mentre più della metà dei rispondenti (51%) teme le cosiddette allucinazioni dei modelli, e circa un'impresa su tre indica cybersicurezza e dipendenza da fornitori terzi tra le proprie preoccupazioni.

C'è poi la questione dell'infrastruttura su cui girano queste soluzioni. Quasi due terzi delle imprese (62%) si affidano esclusivamente al cloud commerciale, spesso di un solo fornitore, mentre le infrastrutture private restano una prerogativa delle grandi imprese (64%, contro appena il 18% delle micro-imprese). Il mercato dei fornitori è inoltre molto concentrato: Microsoft è il principale provider per quasi metà del campione, seguito da OpenAI e AWS, e solo l'8% dei fornitori indicati dalle imprese ha sede nell'Unione Europea - un dettaglio che dice qualcosa sulla dipendenza del settore finanziario europeo da attori extra-UE.

Nel complesso, il quadro che emerge è quello di un'adozione ancora cauta e concentrata soprattutto nelle imprese più grandi, ma con aspettative diffusamente ottimistiche sui benefici che l'AI potrà portare nei prossimi anni.

[10] Calcolo effettuato sommando le categorie "gestione del rischio di portafoglio", "ottimizzazione di portafoglio", "trading algoritmico", "trading ad alta frequenza", e "robo-advising" sugli 847 casi d'uso complessivi censiti da ESMA (2026).

1.3 Obiettivo e domanda di ricerca

Come si è visto nel paragrafo precedente, l'AI nei mercati finanziari europei viene oggi usata soprattutto per attività di supporto operativo, mentre le applicazioni più vicine alle decisioni di investimento restano una nicchia, poco più del 6% dei casi d'uso censiti da ESMA. È proprio da questo divario - tra il potenziale spesso attribuito all'intelligenza artificiale e il suo utilizzo ancora limitato nei processi decisionali di investimento - che nasce il presente lavoro di ricerca.

L'obiettivo di questa tesi è verificare, con un'analisi sperimentale, che contributo può dare l'intelligenza artificiale - in particolare i modelli generativi basati su Large Language Models - alla generazione di proposte di allocazione di portafoglio, confrontando le performance delle strategie costruite con il supporto dell'AI rispetto a strategie più tradizionali, usate come benchmark.

La domanda di ricerca a cui questo lavoro cerca di rispondere è, quindi: l'intelligenza artificiale generativa è in grado di generare proposte di allocazione di portafoglio capaci di ottenere performance comparabili, o superiori, rispetto a strategie tradizionali? A questa domanda principale se ne affiancano altre due, più specifiche: quali sono i limiti metodologici - overfitting e bias nei backtest, in particolare - a cui vanno incontro le strategie generate con l'AI? E quanto sono, in pratica, interpretabili e spiegabili?

Per rispondere a queste domande si è scelto un approccio sperimentale, descritto nel dettaglio nel Capitolo 3, che mette a confronto portafogli costruiti con strategie tradizionali e portafogli costruiti a partire da proposte di allocazione di portafoglio generate con l'intelligenza artificiale.

Capitolo 2

INTELLIGENZA ARTIFICIALE E STRATEGIE DI INVESTIMENTO

Il capitolo precedente ha delineato il contesto generale in cui si inserisce l'intelligenza artificiale nel settore finanziario, mostrando come il suo utilizzo nei processi decisionali di investimento rimanga ancora limitato rispetto al potenziale attribuito a queste tecnologie. Il presente capitolo costruisce il quadro teorico necessario per interpretare l'analisi sperimentale sviluppata nei capitoli successivi.

L'obiettivo del presente capitolo non è descrivere in modo esaustivo tutte le applicazioni dell'intelligenza artificiale in finanza, ma individuare i contributi scientifici utili ad inquadrare l'esperimento empirico.

Per questo motivo, il capitolo parte dalla letteratura sull'asset allocation e sui benchmark di investimento, poiché il lavoro non riguarda la selezione di singoli titoli, ma la costruzione di portafogli multi-asset. Successivamente, vengono richiamate le strategie quantitative tradizionali e le applicazioni del machine learning in finanza, distinguendo i modelli predittivi addestrati sui dati dall'approccio adottato in questa tesi, in cui l'AI generativa viene utilizzata come generatore di proposte di allocazione.

Il paragrafo 2.3 restringe quindi il campo alla letteratura più recente sull'impiego dei modelli linguistici nella selezione e costruzione di portafogli.

Gli ultimi due paragrafi affrontano i principali problemi metodologici dell'analisi: il rischio di overfitting e data snooping nei backtest, il possibile data leakage legato all'utilizzo di modelli linguistici ed il tema dell'interpretabilità dei risultati.

2.1 Asset allocation e benchmark di investimento

La costruzione di un portafoglio nasce dal problema di allocare il capitale tra attività finanziarie caratterizzate da rendimenti attesi, rischi e correlazioni differenti: la qualità della strategia non dipende quindi solo dalle singole attività considerate isolatamente, ma dal modo in cui esse vengono combinate. Il punto di partenza della moderna teoria di portafoglio è il contributo di Markowitz, che formalizza la scelta di portafoglio come un problema di equilibrio tra rendimento atteso e rischio e mostra che il rischio di un investimento non può essere valutato considerando i singoli asset isolatamente, ma va misurato a livello di portafoglio, tenendo conto delle correlazioni tra le attività[11].

Questa impostazione è direttamente rilevante per il presente lavoro. L'intelligenza artificiale non viene impiegata per selezionare singoli titoli, ma per assegnare pesi a un insieme chiuso di indici rappresentativi di asset class differenti. La qualità dell'output generato dal modello coincide quindi con la coerenza dell'allocazione: un portafoglio aggressivo dovrebbe assumere più rischio, uno conservativo dovrebbe privilegiare stabilità e protezione, mentre un portafoglio bilanciato dovrebbe collocarsi in posizione intermedia.

Sull'importanza dell'allocazione la letteratura empirica offre un riferimento classico: analizzando un campione di grandi fondi pensione statunitensi, la

[11] Harry Markowitz, "Portfolio Selection," *The Journal of Finance* 7, no. 1 (1952): 77–91, https://doi.org/10.1111/j.1540-6261.1952.tb01525.x.

politica di allocazione strategica risulta spiegare in media oltre il 90% della variabilità dei rendimenti complessivi nel tempo[12]. Ibbotson e Kaplan (2000) riprendono successivamente questo risultato, chiarendone la portata: l'asset allocation spiega circa il 90% della variabilità dei rendimenti di un fondo nel tempo, circa il 40% delle differenze di rendimento tra fondi diversi e quasi la totalità del livello dei rendimenti[13].

La lettura congiunta di questi contributi è utile per la tesi in due direzioni. Da un lato, conferma che l'allocazione tra asset class è l'oggetto rilevante su cui interrogare l'intelligenza artificiale; dall'altro, ricorda che confrontare strategie diverse richiede benchmark adeguati. Una performance elevata può, infatti, derivare semplicemente da una maggiore esposizione ad asset rischiosi, e non da una reale capacità nella costruzione del portafoglio.

È questa la ragione per cui, nell'analisi empirica, ciascun portafoglio generato dall'intelligenza artificiale sarà confrontato con benchmark differenti, selezionati in base al livello di rischio del portafoglio esaminato, oltre che con un riferimento comune.

Un benchmark particolarmente importante è l'equal-weight, o strategia 1/N, che assegna lo stesso peso a ciascun asset dell'universo investibile. Pur trattandosi di una regola apparentemente semplice, la sua importanza empirica è significativa: diversi modelli di ottimizzazione di portafoglio, infatti, non riescono a superare in modo sistematico, fuori campione, la semplice allocazione equiponderata in termini di performance aggiustata per il rischio[14]. La ragione

[12] Gary P. Brinson et al., "Determinants of Portfolio Performance," *Financial Analysts Journal* 42, no. 4 (1986): 39–44, https://doi.org/10.2469/faj.v42.n4.39.

[13] Roger G. Ibbotson and Paul D. Kaplan, "Does Asset Allocation Policy Explain 40, 90, or 100 Percent of Performance?," *Financial Analysts Journal* 56, no. 1 (2000): 26–33, https://doi.org/10.2469/faj.v56.n1.2327.

[14] Victor DeMiguel et al., "Optimal Versus Naive Diversification: How Inefficient Is the 1/*N* Portfolio Strategy?," *Review of Financial Studies* 22, no. 5 (2009): 1915–53, https://doi.org/10.1093/rfs/hhm075.

principale è che gli errori nella stima dei rendimenti attesi e delle covarianze possono ridurre, o persino annullare, i vantaggi teorici dell'ottimizzazione.

Da questa letteratura deriva una prima scelta metodologica dell'analisi empirica: i portafogli generati dall'intelligenza artificiale saranno confrontati sia con benchmark tradizionali coerenti per profilo di rischio, sia con una regola comune e trasparente come l'equal-weight sull'universo investibile. In questo modo sarà possibile valutare se l'eventuale contributo dell'AI vada oltre quanto ottenibile attraverso una semplice regola di diversificazione o attraverso allocazioni tradizionali coerenti con il profilo di rischio.

2.2 Strategie quantitative e Machine Learning

Per strategia quantitativa di investimento si intende una strategia in cui le decisioni di portafoglio derivano da regole sistematiche, definite ex ante e applicate in modo replicabile, anziché dal solo giudizio discrezionale del gestore[15]. In questa prospettiva, l'allocazione dipende da criteri formalizzati: selezione dell'universo investibile, definizione dei vincoli, regole di ponderazione, controllo del rischio e modalità di ribilanciamento. Un riferimento centrale di questa tradizione è il factor investing, secondo cui i rendimenti attesi degli asset possono essere ricondotti all'esposizione verso alcune caratteristiche osservabili, dette fattori, associate a premi al rischio persistenti[16].

La letteratura fattoriale nasce dall'idea che il solo rischio di mercato non sia sufficiente a spiegare le differenze di rendimento tra titoli. Fama e French mostrano, ad esempio, che due variabili facilmente osservabili - la dimensione

[15] Richard C. Grinold and Ronald N. Kahn, *Active Portfolio Management: A Quantitative Approach for Producing Superior Returns and Controlling Risk*, 2nd ed. (McGraw-Hill, 2000).

[16] Andrew Ang, *Asset Management: A Systematic Approach to Factor Investing* (Oxford University Press, 2014).

dell'impresa e il rapporto tra valore contabile e valore di mercato - contribuiscono a spiegare i rendimenti azionari meglio del solo beta di mercato[17]. A queste regolarità si aggiunge il momentum, cioè la tendenza dei titoli che hanno ottenuto le migliori performance nei mesi precedenti a continuare a sovraperformare anche nel periodo successivo[18]. Il modello a quattro fattori di Carhart sintetizza questi contributi includendo mercato, size, value e momentum nella spiegazione del rendimento in eccesso dei portafogli[19].

Per il presente lavoro, il punto rilevante non è la specifica forma econometrica di questi modelli, ma la loro implicazione metodologica. Un rendimento superiore non costituisce, di per sé, evidenza di abilità o di superiorità della strategia: può riflettere semplicemente una maggiore esposizione a fattori o asset class già remunerati dal mercato.

Negli ultimi anni l'evoluzione delle strategie quantitative si è intrecciata con lo sviluppo del machine learning. Con questa espressione si indica un insieme di tecniche statistiche in cui la relazione tra variabili non viene interamente specificata a priori dal ricercatore, ma appresa dai dati con l'obiettivo di migliorare la capacità predittiva fuori campione[20]. Rispetto ai modelli lineari tradizionali, questi metodi consentono di gestire un numero elevato di predittori, forme funzionali più flessibili e interazioni non lineari tra variabili. In finanza, ciò riguarda soprattutto metodi in grado di gestire un elevato numero di variabili

[17] Eugene F. Fama and Kenneth R. French, "The Cross-Section of Expected Stock Returns," *The Journal of Finance* 47, no. 2 (1992): 427–65, https://doi.org/10.1111/j.1540-6261.1992.tb04398.x.

[18] Narasimhan Jegadeesh and Sheridan Titman, "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency," *The Journal of Finance* 48, no. 1 (1993): 65–91, https://doi.org/10.1111/j.1540-6261.1993.tb04702.x.

[19] Mark M. Carhart, "On Persistence in Mutual Fund Performance," *The Journal of Finance* 52, no. 1 (1997): 57–82, https://doi.org/10.1111/j.1540-6261.1997.tb03808.x.

[20] Sendhil Mullainathan and Jann Spiess, "Machine Learning: An Applied Econometric Approach," *Journal of Economic Perspectives* 31, no. 2 (2017): 87–106, https://doi.org/10.1257/jep.31.2.87.

e relazioni non lineari - dalle regressioni con vincoli sui coefficienti ai modelli ad albero, fino alle reti neurali - utilizzati per affrontare relazioni instabili, dati rumorosi e strutture difficili da descrivere con modelli parametrici semplici[21].

Una verifica particolarmente rilevante in ambito di asset pricing è quella di Gu et al. (2020)[22], che confrontano diverse tecniche di machine learning nella previsione dei rendimenti azionari statunitensi su un ampio campione di titoli e variabili predittive. I risultati mostrano che i metodi capaci di cogliere non linearità e interazioni, in particolare alberi e reti neurali, ottengono prestazioni superiori rispetto ai modelli lineari tradizionali. Questa letteratura, ormai ampia, è stata successivamente sistematizzata da Kelly e Xiu (2023)[23], che ne ricostruiscono le principali applicazioni nell'ambito del financial machine learning. Per questa tesi, tuttavia, l'aspetto più importante non è soltanto la maggiore accuratezza predittiva, ma la disciplina metodologica che tali approcci richiedono: separazione tra dati di stima e dati di test, valutazione fuori campione e attenzione al rischio di adattare il modello al rumore storico.

La tesi si colloca quindi all'incrocio tra strategie quantitative e intelligenza artificiale generativa. Dalle prime eredita l'idea di portafogli costruiti secondo regole esplicite, replicabili e valutabili; dalla letteratura sul machine learning eredita invece la necessità di una verifica fuori campione e la consapevolezza dei principali rischi metodologici che saranno approfonditi nei paragrafi successivi.

Questa impostazione prepara direttamente l'analisi empirica dei capitoli successivi. I portafogli generati dall'intelligenza artificiale saranno statici, long-

[21] Trevor Hastie et al., *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. (Springer, 2009).

[22] Shihao Gu et al., "Empirical Asset Pricing via Machine Learning," *The Review of Financial Studies* 33, no. 5 (2020): 2223–73, https://doi.org/10.1093/rfs/hhaa009.

[23] Bryan Kelly and Dacheng Xiu, "Financial Machine Learning," *Foundations and Trends in Finance* 13, nos. 3–4 (2023): 205–363, https://doi.org/10.1561/500000064.

only, privi di leva finanziaria e costruiti su un universo investibile prestabilito. La loro validità sarà valutata attraverso metriche standard di performance e rischio e mediante il confronto con benchmark coerenti.

2.3 GenAI e portfolio selection

La diffusione dei Large Language Models ha aperto una linea di ricerca recente, distinta dal machine learning predittivo tradizionale: il modello non viene necessariamente impiegato per stimare una variabile finanziaria, ma come strumento di generazione, selezione o valutazione di idee di investimento.

Il risultato di maggior rilievo di questo filone mostra che GPT-4 è in grado di prevedere la direzione della reazione dei prezzi azionari alle notizie societarie meglio dei metodi tradizionali di analisi del sentiment, pur senza alcun addestramento specifico in ambito finanziario[24]. Lo studio è rilevante per questa tesi anche sul piano del metodo: per escludere che i risultati riflettano informazioni già incorporate nell'addestramento, il modello viene valutato esclusivamente su notizie successive al suo knowledge cutoff - una strategia di identificazione che sarà ripresa nel paragrafo successivo.

Sul terreno più vicino a questa tesi, quello della costruzione di portafogli, l'analisi dell'impiego di ChatGPT nella selezione dell'universo di investimento evidenzia sia il potenziale dei modelli generativi come supporto decisionale, sia la necessità di verificare e validare accuratamente gli output prodotti: l'affidamento diretto alle raccomandazioni del modello è problematico, anche per il rischio di hallucinations, cioè di contenuti presentati come fattuali ma che

[24] Alejandro Lopez-Lira and Yuehua Tang, "Can ChatGPT Forecast Stock Price Movements? Return Predictability and Large Language Models," *Journal of Financial Economics*, forthcoming, https://doi.org/10.2139/ssrn.4412788.

sono falsi, inventati o non supportati dalle fonti[25]. Altri studi documentano che i portafogli suggeriti dal modello presentano benefici di diversificazione non banali[26] e una discreta capacità nella valutazione di singoli titoli[27]. Nel complesso, la letteratura converge su un punto: i modelli linguistici possono essere strumenti utili di supporto al processo di investimento, ma i loro output richiedono una valutazione quantitativa successiva e indipendente.

È esattamente la posizione adottata in questo lavoro. L'intelligenza artificiale generativa non viene trattata come un decisore autonomo né come un ottimizzatore di portafoglio: viene trattata come un generatore di proposte di allocazione di portafoglio. Il suo compito è formulare, una sola volta e sotto vincoli espliciti, tre portafogli multi-asset coerenti con tre diversi profili di rischio - aggressivo, bilanciato e conservativo - la cui validità è verificata solo successivamente, attraverso un backtest condotto con metriche standard e il confronto con benchmark coerenti. Rispetto alla letteratura appena richiamata, la tesi sposta quindi l'attenzione dalla previsione puntuale e dalla selezione di titoli alla generazione di allocazioni strategiche: un compito in cui l'output del modello è, per costruzione, esplicito e verificabile.

2.4 Limiti metodologici: overfitting e bias nei backtest

La valutazione empirica delle strategie di investimento presenta un rischio metodologico ricorrente: che una strategia appaia efficace sui dati storici pur non possedendo alcuna reale capacità predittiva. Il fenomeno è noto come overfitting

[25] Oleksandr Romanko et al., "ChatGPT-Based Investment Portfolio Selection," *Operations Research Forum* 4, no. 4 (2023): 91, https://doi.org/10.1007/s43069-023-00277-6.

[26] Hyungjin Ko and Jaewook Lee, "Can ChatGPT Improve Investment Decisions? From a Portfolio Management Perspective," *Finance Research Letters* 64 (June 2024): 105433, https://doi.org/10.1016/j.frl.2024.105433.

[27] Matthias Pelster and Joel Val, "Can ChatGPT Assist in Picking Stocks?," *Finance Research Letters* 59 (January 2024): 104786, https://doi.org/10.1016/j.frl.2023.104786.

e si verifica quando una regola di investimento si adatta eccessivamente alle caratteristiche specifiche - comprese quelle puramente casuali - del campione su cui è stata costruita[28].

Nei backtest finanziari il problema assume una forma particolare, nota come data snooping. Con questo termine la letteratura indica la distorsione che emerge quando lo stesso insieme di dati viene utilizzato per formulare, testare e selezionare le proposte di investimento: se molte regole vengono provate sulla medesima serie storica, alcune risulteranno profittevoli per puro effetto del caso. Il fenomeno opera anche involontariamente - ad esempio quando la costruzione stessa di un test è influenzata da regolarità individuate ispezionando gli stessi dati - e questo è sufficiente a gonfiare la significatività statistica apparente dei risultati[29].

L'evidenza empirica più nota è stata condotta su regole di trading tecnico: valutando congiuntamente l'intero universo delle regole esaminate dalla letteratura, la performance delle regole apparentemente migliori si ridimensiona drasticamente una volta tenuto conto del numero complessivo di regole provate[30].

Lo stesso fenomeno è stato formalizzato per i backtest: con un numero sufficiente di configurazioni testate, ottenere una strategia con uno Sharpe ratio elevato sui dati storici è pressoché garantito, indipendentemente da qualsiasi valore predittivo.

---

[28] David H. Bailey et al., "Pseudo-Mathematics and Financial Charlatanism: The Effects of Backtest Overfitting on Out-of-Sample Performance," *Notices of the American Mathematical Society* 61, no. 5 (2014): 458–71, https://doi.org/10.1090/noti1105.

[29] Andrew W. Lo and A. Craig MacKinlay, "Data-Snooping Biases in Tests of Financial Asset Pricing Models," *Review of Financial Studies* 3, no. 3 (1990): 431–67, https://doi.org/10.1093/rfs/3.3.431.

[30] Ryan Sullivan et al., "Data-Snooping, Technical Trading Rule Performance, and the Bootstrap," *The Journal of Finance* 54, no. 5 (1999): 1647–91, https://doi.org/10.1111/0022-1082.00163.

Il problema investe la stessa ricerca accademica: il riesame di oltre trecento fattori di rendimento pubblicati sulle principali riviste suggerisce che una parte rilevante dei risultati presenti in letteratura sia frutto di selezione tra molteplici test, al punto da richiedere un innalzamento sostanziale delle soglie statistiche per dichiarare significativa una nuova scoperta[31].

Applicata all'intelligenza artificiale generativa, questa letteratura individua con precisione i comportamenti da evitare. Se il prompt viene riformulato dopo aver osservato i risultati, o se vengono generate molte strategie e viene riportata soltanto la migliore, l'esperimento ricade in una forma di *"cherry-picking"* del tutto analoga al data snooping: cambia lo strumento, non la distorsione. Per questo il protocollo empirico adottato prevede tre presìdi, mutuati direttamente da questa letteratura: il prompt viene definito e congelato prima di calcolare qualsiasi backtest; la generazione dei portafogli avviene una sola volta; tutti gli output prodotti vengono riportati e valutati, senza selezione a posteriori.

Una seconda famiglia di problemi riguarda l'uso improprio di informazioni future. Questo fenomeno generale è noto come data leakage: si ha leakage quando nel processo di costruzione o valutazione di un modello entra informazione che non sarebbe stata disponibile al momento della previsione[32]. Nei backtest finanziari la manifestazione classica del leakage prende il nome di *"look-ahead bias"*: distorsione che si forma quando la simulazione storica impiega dati che, alla data della decisione simulata, non erano ancora noti al mercato[33].

---

[31] Campbell R. Harvey et al., "… and the Cross-Section of Expected Returns," *Review of Financial Studies* 29, no. 1 (2016): 5–68, https://doi.org/10.1093/rfs/hhv059.

[32] Shachar Kaufman et al., "Leakage in Data Mining: Formulation, Detection, and Avoidance," *ACM Transactions on Knowledge Discovery from Data* 6, no. 4 (2012): 1–21, https://doi.org/10.1145/2382577.2382579.

[33] Rolf W. Banz and William J. Breen, "Sample-Dependent Results Using Accounting and Market Data: Some Evidence," *The Journal of Finance* 41, no. 4 (1986): 779–93, https://doi.org/10.1111/j.1540-6261.1986.tb04548.x.

Applicati ai modelli linguistici, leakage e look-ahead bias assumono una forma più insidiosa, in quanto la contaminazione non avviene più sul dataset del backtest, ma sulla conoscenza incorporata nel modello stesso. Poiché questi modelli apprendono in fase di addestramento gli eventi di mercato fino al proprio knowledge cutoff, imporre al modello una data informativa passata non garantisce un vero e proprio test out-of-sample: il modello potrebbe comunque implicitamente attingere a conoscenze successive a quella data. Il rischio non è ipotetico: è stato documentato empiricamente che le previsioni sui rendimenti generate da modelli quali ChatGPT a partire da notizie precedenti al cutoff beneficiano di un look-ahead bias - anche solo attraverso il riconoscimento dell'identità delle società coinvolte - e che la performance si ridimensiona una volta rimosso questo vantaggio informativo[34].

Per gestire questo rischio l'analisi empirica adotta un vincolo informativo imposto tramite il prompt: al modello viene richiesto esplicitamente di non utilizzare informazioni, dati o conoscenze successive a una data prefissata, e i portafogli generati sono valutati sul periodo seguente. Questo disegno beneficia di una finestra di valutazione lunga, ma va qualificato per ciò che è - un test pseudo out-of-sample - per due ragioni distinte: in primo luogo, il rispetto del vincolo è affidato all'istruzione, e non a una reale segregazione dei dati forniti al modello; in secondo luogo, l'intero periodo di valutazione è anteriore al knowledge cutoff effettivo del modello impiegato, che ha dunque già incorporato in fase di addestramento l'andamento dei mercati su cui i portafogli vengono giudicati. Un test autenticamente fuori campione richiederebbe di

[34] Paul Glasserman and Caden Lin, "Assessing Look-Ahead Bias in Stock Return Predictions Generated by GPT Sentiment Analysis," *The Journal of Financial Data Science* 6, no. 1 (2024): 25–42, https://doi.org/10.3905/jfds.2023.1.143.

valutare i portafogli sul periodo successivo al knowledge cutoff del modello, secondo la strategia di identificazione adottata in letteratura[35].

Si tratta di una verifica che i vincoli temporali del presente lavoro non consentono di condurre su una finestra sufficientemente lunga. In assenza di questa verifica, la difesa metodologica sul fronte del leakage si affida a tre elementi: la natura dell'esercizio richiesto al modello, che genera allocazioni statiche fondate su principi generali di costruzione di portafoglio e non previsioni puntuali di rendimento; la trasparenza del protocollo, con prompt congelato e generazione unica degli output; e la dichiarazione esplicita di questo limite, di cui l'interpretazione dei risultati dovrà tenere conto.

2.5 Interpretabilità e Explainable AI

L'interpretabilità, ossia la capacità di un sistema di spiegare le proprie decisioni in termini comprensibili a un essere umano[36], è un requisito centrale nell'applicazione dell'intelligenza artificiale alla finanza. Le decisioni di investimento non possono essere valutate soltanto in base alla performance realizzata: devono poter essere comprese, motivate e comunicate, tanto più quando riguardano l'allocazione di capitale altrui.

La letteratura sull'Explainable AI nasce dall'esigenza di rendere comprensibili modelli complessi e opachi. L'Explainable AI è infatti l'insieme di metodi e tecniche che rendono comprensibili a un essere umano le decisioni e il funzionamento interno di un sistema di intelligenza artificiale. I metodi di riferimento sono le spiegazioni post-hoc: l'approssimazione locale del

[35] Lopez-Lira and Tang, "Can ChatGPT Forecast Stock Price Movements?"

[36] Finale Doshi-Velez and Been Kim, "Towards A Rigorous Science of Interpretable Machine Learning," arXiv:1702.08608, preprint, arXiv, March 2, 2017, https://doi.org/10.48550/arXiv.1702.08608.

comportamento del modello attorno a una singola previsione mediante un modello lineare semplice (LIME)[37] e l'attribuzione a ciascuna variabile di un contributo alla previsione fondato sui valori di Shapley della teoria dei giochi (SHAP)[38]. Questi strumenti sono preziosi quando l'oggetto da spiegare è la previsione di un modello black-box. La stessa letteratura, tuttavia, ha maturato una posizione critica sul ricorso sistematico alle spiegazioni a posteriori: per le decisioni ad alto impatto, spiegare ex post un modello opaco è una soluzione fragile - la spiegazione è a sua volta un'approssimazione, che può risultare infedele - ed è preferibile, quando possibile, impiegare modelli i cui output siano interpretabili per costruzione[39].

È la prospettiva adottata in questo lavoro. Il modello linguistico resta opaco nel processo interno con cui formula le allocazioni; ma l'oggetto dell'analisi empirica non è quel processo, bensì i suoi output: tre portafogli multi-asset con pesi espliciti su indici predefiniti, cioè output interpretabili per costruzione. L'interpretabilità non viene quindi ricercata con strumenti post-hoc, ma attraverso la lettura economica delle esposizioni generate: un portafoglio aggressivo dovrebbe presentare un peso prevalente verso asset rischiosi, come l'azionario e il credito high yield; un portafoglio conservativo dovrebbe privilegiare strumenti difensivi, come Treasury a breve scadenza e obbligazionario investment grade; un portafoglio bilanciato dovrebbe collocarsi in posizione intermedia. L'analisi dei pesi diventa così parte integrante della

[37] Marco Tulio Ribeiro et al., "'Why Should I Trust You?': Explaining the Predictions of Any Classifier," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, August 13, 2016, 1135–44, https://doi.org/10.1145/2939672.2939778.

[38] Scott Lundberg and Su-In Lee, "A Unified Approach to Interpreting Model Predictions," *Advances in Neural Information Processing Systems 30 (NIPS 2017)* (Long Beach, CA), 2017, 4765–74, https://doi.org/10.48550/arXiv.1705.07874.

[39] Cynthia Rudin, "Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead," *Nature Machine Intelligence* 1, no. 5 (2019): 206–15, https://doi.org/10.1038/s42256-019-0048-x.

valutazione: consente di verificare non solo se le allocazioni siano competitive in termini di performance, ma anche se siano coerenti con il profilo di rischio richiesto.

Un discorso a parte riguarda le motivazioni testuali che il modello fornisce a corredo di ciascun portafoglio. Queste motivazioni costituiscono materiale qualitativo utile a supportare la lettura economica delle allocazioni, ma non rappresentano una descrizione fedele del processo generativo: un modello di intelligenza artificiale può produrre giustificazioni plausibili, che però non riflettono i meccanismi effettivamente sottostanti all'output, configurandosi come razionalizzazioni ex-post. Per questa ragione la valutazione dell'interpretabilità si fonda sui pesi osservabili, mentre le motivazioni sono oggetto di analisi critica, non assunte come spiegazione del processo.

Capitolo 3

METODOLOGIA

Questo capitolo cerca di tradurre in un protocollo operativo la domanda di ricerca posta nel Capitolo 1 e le cautele metodologiche discusse nel Capitolo 2. L'esperimento chiede ai 3 modelli di intelligenza artificiale generativa - ChatGPT (OpenAI), Gemini (Google) e Claude (Anthropic) - di costruire, a partire da uno stesso prompt, tre portafogli statici multiasset corrispondenti ad altrettanti profili di rischio: aggressivo, bilanciato e conservativo. I 9 portafogli risultanti verranno poi sottoposti a backtest sul decennio 2016-2025 e confrontati, per profilo, tramite benchmark tradizionali costruiti sullo stesso universo.

L'ordine di esposizione cerca di seguire l'ordine in cui le decisioni sono state prese e congelate: prima il dataset e l'universo investibile (paragrafo 3.1), poi i benchmark di confronto (paragrafo 3.2), successivamente la generazione delle strategie tramite i vari modelli di AI (paragrafo 3.3) ed infine le regole del backtest e le metriche di valutazione (paragrafo 3.4).

Questa sequenza non è solo esplicativa, ma sostanziale al progetto, in quanto ogni elemento del disegno - dati, benchmark, prompt e metriche di valutazione - è fissato prima di osservare qualunque risultato, così da escludere aggiustamenti a posteriori.

3.1 Dataset e universo investibile

L'esperimento si svolge su un universo investibile chiuso, composto da otto indici rappresentativi di diverse asset class: tre comparti azionari (Stati Uniti, mercati sviluppati ex-USA, mercati emergenti), quattro comparti obbligazionari (titoli di Stato americani a lunga e a breve scadenza, obbligazioni investment grade e high yield) e l'oro. La scelta di un universo chiuso e prestabilito risponde a un'esigenza precisa del disegno sperimentale: i modelli generativi devono operare tutti sullo stesso insieme di alternative, senza possibilità di introdurre strumenti esterni, così che le differenze tra le allocazioni siano dovute esclusivamente a differenze nelle scelte di peso degli asset e non alla selezione di strumenti.

Le sette serie riferite ai comparti azionari e obbligazionari sono espresse in versione total return, mentre per l'oro viene utilizzato il prezzo spot in dollari per oncia. Tutte le serie sono denominate in dollari statunitensi. L'utilizzo di indici total return - che incorpora il reinvestimento di tutte le cedole e dividendi - rappresenta una condizione essenziale per un corretto confronto tra varie asset-class; la differenza tra un indice di prezzo ed il corrispettivo total return è nell'ordine di svariati punti percentuali per i comparti azionari su un orizzonte temporale di 10 anni e rappresenta la quasi totalità del rendimento per i comparti obbligazionari. La tabella 3.1 riporta l'anagrafica completa delle serie utilizzate.

Tabella 3.1: Universo investibile - Anagrafica delle serie

Asset class	Indice	Codice (RIC)	Prima data
Azionario USA	S&P 500 Total Return Index	.SPXTR	03/01/2006
Azionario sviluppato ex-USA	MSCI EAFE Net Total Return Index	.dMIEA00000NUS	02/01/2006

(Tabella 3.1 segue in pagina successiva)

Tabella 3.1 (prosegue dalla pagina precedente)

Azionario emergente	MSCI Emerging Markets Net Total Return Index	.dMIEF00000NUS	02/01/2006
Treasury lungo	FTSE US 10-Year Treasury Benchmark Total Return Index	.FTBNCH_OFF10LCLT	01/01/2009
Treasury breve	FTSE US 2-Year Treasury Benchmark Total Return Index	.FTBNCH_OFF2LCLT	01/01/2009
Investment grade "Broad"	FTSE US Broad Investment-Grade Bond Total Return Index	.FTUSBIGUSDT	03/01/2006
Corporate high yield	FTSE US High-Yield Market Total Return Index	.FTHYMUSDT	03/01/2006
Oro	Gold Spot (USD/oncia)	XAU=	03/01/2006

Fonte: LSEG Workspace (Refinitiv Eikon)

Oltre a questi dati, il tasso privo di rischio, utilizzato nel calcolo delle misure aggiustate per il rischio, è il rendimento del Treasury Bill statunitense a tre mesi sul mercato secondario (serie DTB3), tratto dal database FRED della Federal Reserve Bank of St. Louis. Si tratta dell'unica serie esterna alla piattaforma LSEG: il risk-free non è un componente dell'universo investibile ma un parametro di valutazione, e la fonte istituzionale ne garantisce la piena tracciabilità.

La composizione dell'universo è l'esito di una verifica sistematica di disponibilità e qualità condotta prima del congelamento del disegno. Per il comparto obbligazionario si è scelta la famiglia FTSE in quanto essa è risultata accessibile, completa delle quattro asset class obbligazionarie e con un'adeguata profondità storica.

Il periodo di valutazione va dal 1° gennaio 2016 al 31 dicembre 2025, con dati a frequenza giornaliera: 2514 osservazioni per serie sul calendario di negoziazione statunitense. La frequenza giornaliera garantisce stime più accurate della volatilità e del massimo drawdown, rispetto ad una frequenza

mensile, dove quest'ultimo risulta sistematicamente sottostimato per effetto di compensazioni dei movimenti infra-mese.

Il trattamento dei dati grezzi segue tre regole fondamentali. Primo: i file esportati tramite le piattaforme non vengono mai modificati manualmente; si procede alla trasformazione di questi ultimi esclusivamente tramite codice, riportato integralmente in Appendice B, in modo da ridurre le probabilità di errore. Secondo: le otto serie vengono allineate sul calendario di negoziazione statunitense, assumendo come riferimento le date dell'indice S&P500; le date in cui una serie non presenta alcun valore vengono colmate con l'ultimo valore disponibile, per un massimo di cinque giorni consecutivi. L'intervento ha riguardato 32 date per le tre serie FTSE dei Treasury e del credito investment grade e 2 date per l'high yield, su 2.514 totali. Terzo: ogni correzione puntuale viene documentata in un registro allegato in Appendice C. L'unico intervento di questo tipo ha riguardato l'osservazione del 18 gennaio 2022 per le tre serie FTSE Treasury 10 anni, Treasury 2 anni e investment grade, che riportavano ribassi compresi tra il 3 e l'11 per cento con pieno recupero il giorno successivo: un movimento incompatibile rispetto alla duration di quegli indici - il rendimento decennale variò di circa 9 punti base quel giorno, a fronte dei circa 122 che sarebbero stati necessari a giustificare la variazione registrata - e riconducibile ad un errore di pubblicazione[40]. Il valore è stato quindi rimosso ed aggiustato con interpolazione lineare.

[40] Tra il 14 gennaio 2022, ultima rilevazione precedente disponibile, e il 18 gennaio 2022, i rendimenti Treasury constant maturity aumentarono di 7 punti base sulla scadenza biennale, di 10 punti base sulle scadenze a cinque e sette anni e di 9 punti base sulla scadenza decennale. In base all'approssimazione di primo ordine fondata sulla duration modificata, secondo cui $\Delta P/P \approx -D_{mod}\Delta y$, un aumento dei rendimenti di 10 punti base determinerebbe, per un indice con duration modificata pari a circa 9, una perdita nell'ordine dello 0,9%. Per giustificare una diminuzione dell'11% sarebbe invece necessario, a parità di duration, un aumento dei rendimenti di circa 122 punti base. La stima costituisce un'approssimazione lineare e non considera l'effetto della convessità. Fonte dei rendimenti: U.S. Department of the Treasury, *Daily Treasury Par Yield Curve Rates*, archivio 2020–2023

3.2 Costruzione dei portafogli benchmark

La valutazione dei portafogli generati dai modelli di intelligenza artificiale generativa richiede termini di confronto che siano coerenti con i vari profili di rischio: confrontare portafogli conservativi con benchmark fortemente esposti all'azionario, o viceversa, renderebbe il giudizio privo di significato. Per questo motivo si prevede l'utilizzo di un benchmark dedicato per ciascuno dei tre profili di rischio richiesti ai modelli, più un riferimento comune. Tutti i benchmark sono costruiti esclusivamente tramite gli otto indici dell'universo investibile e sottoposti allo stesso identico backtest dei portafogli AI - stessi dati, stesso periodo e stesso ribilanciamento - in modo tale che ogni differenza di risultato sia attribuibile all'allocazione e non alle condizioni del confronto.

I tre benchmark condividono la stessa architettura: una componente azionaria ed una obbligazionaria. La componente azionaria è ripartita per il 60 per cento sull'indice S&P500, per il 25 per cento sull'indice MSCI EAFE e per il 15 per cento sull'indice MSCI Emerging Markets: proporzioni che approssimano i pesi del mercato azionario mondiale, così che il blocco azionario rappresenti il portafoglio "neutrale" di un investitore globale. La componente obbligazionaria è ripartita per il 40 per cento sul Treasury decennale, per il 30 per cento sul Treasury a due anni e per il 30 per cento sull'investment grade: composizione che replica i tratti essenziali di un indice obbligazionario aggregato statunitense. Si nota che l'high yield e l'oro rimangono esclusi dalla creazione dei benchmark, ma questi rimangono pienamente disponibili ai modelli generativi: se l'intelligenza artificiale decide di utilizzarli il confronto con benchmark che non li contengono rappresenta una parte del risultato.

Le quote azionarie dei tre benchmark - 80, 60 e 20 per cento - corrispondono ai tre profili di rischio richiesti ai modelli e trovano riscontro nella prassi dell'industria del risparmio gestito: un'allocazione dell'80 per cento in azioni

corrisponde ad un profilo *"Growth"*, una del 60 per cento nel profilo *"Moderate Growth"* ed una del 20 per cento definisce il profilo *"Conservative"*[41].

Il quarto riferimento è il portafoglio equal-weight, che assegna un peso identico del 12,5 per cento a ciascuno degli otto indici. La sua funzione è differente rispetto a quella dei benchmark di profilo: rappresenta la diversificazione *"naïve"* che costituisce un termine di paragone sorprendentemente difficile da battere per qualunque strategia di allocazione ottimizzata[42]. Il portafoglio equal-weight è inoltre l'unico benchmark a includere l'high yield e l'oro e per questo motivo funge da controprova.

Tabella 3.2: Composizione dei portafogli benchmark (pesi percentuali)

Benchmark	S&P500	EAFE	EM	T10Y	T2Y	IG	HY	Oro
80/20 (aggressivo)	48%	20%	12%	8%	6%	6%	-	-
60/40 (bilanciato)	36%	15%	9%	16%	12%	12%	-	-
20/80 (conservativo)	12%	5%	3%	32%	24%	24%	-	-
Equal-weight (1/N)	12,5%	12,5%	12,5%	12,5%	12,5%	12,5%	12,5%	12,5%

Fonte: elaborazione propria.

Tutti i portafogli di confronto, come quelli generati dai modelli di intelligenza artificiale, sono statici e vengono ribilanciati con frequenza trimestrale: il primo giorno di negoziazione di gennaio, aprile, luglio e ottobre vengono riportati esattamente ai pesi iniziali, mentre tra un ribilanciamento e l'altro vengono lasciati liberi di variare con i prezzi di mercato. La composizione riportata in Tabella 3.2 è stata definita e congelata prima della finalizzazione del prompt e di qualunque generazione da parte dei modelli, coerentemente con i presidi discussi antecedentemente.

[41] Charles Schwab, "A Guide to Risk Profiles," Schwab Brokerage, May 29, 2026, https://www.schwab.com/learn/story/guide-to-risk-profiles.

[42] DeMiguel et al., "Optimal Versus Naive Diversification."

3.3 Generazione di strategie tramite intelligenza artificiale

Il paragrafo precedente ha definito e congelato la composizione dei benchmark che verranno utilizzati per confrontare i portafogli generati tramite intelligenza artificiale. Lo stesso principio guida la costruzione del prompt, definito prima di qualunque generazione ed utilizzato in modo analogo per tutti i differenti modelli di intelligenza artificiale.

Il rischio di data leakage, già discusso nel paragrafo 2.4 come limite dell'esperimento, viene contenuto dichiarando ai modelli una data informativa simulata - il 31 dicembre 2015, pari allo stesso cutoff che delimita il periodo di backtest - e vietando riferimenti ad eventi o andamenti di mercato successivi a tale data. Il rischio di overfitting sui dati storici viene invece contenuto a priori, semplicemente non fornendo ai modelli la serie storica degli otto indici precedente al momento del cutoff, ma si fornisce esclusivamente una breve descrizione qualitativa di ciascun indice: non fornendo ai modelli dati sui quali poter calcolare correlazioni e rendimenti, i modelli saranno vincolati ad utilizzare strategie generali di asset allocation.

I vincoli quantitativi del prompt - universo chiuso agli otto indici, pesi interi con somma pari a 100, assenza di leva e di vendite allo scoperto, portafoglio statico con ribilanciamento trimestrale - sono gli stessi vincoli che verranno poi utilizzati per il backtest descritto nel paragrafo successivo: rappresenta una corrispondenza necessaria, in quanto quello che viene richiesto al modello deve poi coincidere con ciò che viene effettivamente misurato. Il formato dell'output - tabella dei pesi, motivazione economica, punto di forza e rischio principale - è vincolato al fine di rendere le risposte dei tre modelli facilmente confrontabili tra di loro e al fine di fornire materiale già strutturato per l'analisi quantitativa.

Tabella 3.3: PROMPT dei modelli

PROMPT (identico per i tre modelli)
Sei un asset allocator istituzionale. Il tuo compito è costruire tre portafogli strategici multi-asset di lungo periodo, utilizzando esclusivamente gli 8 indici elencati di seguito. Data informativa simulata: 31/12/2015. Assumi di operare a tale data. Utilizza esclusivamente informazioni e conoscenze finanziarie che sarebbero state disponibili entro il 31 dicembre 2015. Non fare riferimento, esplicito o implicito, a eventi, crisi, politiche monetarie, rendimenti o andamenti di mercato successivi a tale data. **Universo investibile:** 1. **S&P 500 Total Return** - Indice rappresentativo delle principali società statunitensi a grande capitalizzazione; componente core per la crescita del capitale nel lungo periodo. 2. **MSCI EAFE Net Total Return** - Azionario dei mercati sviluppati esclusi Stati Uniti e Canada; utilizzato principalmente per la diversificazione geografica. 3. **MSCI Emerging Markets Net Total Return** - Azionario dei mercati emergenti; caratterizzato da maggiore potenziale di crescita ma anche da rischio economico, politico e valutario più elevato. 4. **US Treasury 10Y Total Return** - Titoli di Stato statunitensi a medio-lunga scadenza; componente difensiva con esposizione al rischio di tasso d'interesse. 5. **US Treasury 2Y Total Return** - Titoli di Stato statunitensi a breve scadenza; orientati alla stabilità e al contenimento della volatilità. 6. **US Investment Grade Total Return** - Obbligazionario statunitense investment grade in forma aggregata, comprendente titoli di Stato e obbligazioni societarie di elevato merito creditizio; fonte di reddito con rischio di credito moderato. 7. **US High Yield Corporate Total Return** - Obbligazioni societarie speculative; maggiore rendimento potenziale accompagnato da un rischio di credito più elevato. 8. **Gold** - Oro; asset comunemente utilizzato per finalità di diversificazione e come possibile protezione in alcune fasi di inflazione o instabilità finanziaria.

(Tabella 3.3 segue in pagina successiva)

Tabella 3.3 (prosegue dalla pagina precedente)

Non sono fornite serie storiche né statistiche quantitative sugli indici. Costruisci i portafogli utilizzando esclusivamente le informazioni riportate sopra e le conoscenze finanziarie disponibili alla data informativa simulata.

Assumi che l'andamento futuro dei mercati non possa essere previsto con affidabilità. Le allocazioni devono quindi riflettere scelte strategiche di lungo periodo, senza ricorrere a market timing o a previsioni su specifici scenari futuri.

Portafogli da costruire:

1. AGGRESSIVO

Obiettivo: massimizzare la crescita del capitale nel lungo periodo, accettando elevata volatilità e drawdown.

2. BILANCIATO

Obiettivo: ricercare un equilibrio tra crescita del capitale e controllo del rischio.

3. CONSERVATIVO

Obiettivo: privilegiare la preservazione del capitale e il contenimento della volatilità e dei drawdown.

Vincoli (validi per ciascun portafoglio):

- utilizza esclusivamente gli 8 indici indicati;
- nessuno strumento esterno;
- pesi percentuali interi compresi tra 0 e 100;
- somma dei pesi esattamente pari a 100;
- portafoglio long-only, nessuna vendita allo scoperto;
- nessuna leva finanziaria;
- portafoglio statico per l'intero orizzonte di analisi;
- ribilanciamento trimestrale ai pesi indicati;
- nessun market timing;
- nessuna regola dinamica o discrezionale.

Formato dell'output

Per ciascun portafoglio:

1) presenta una tabella con due colonne: indice; peso percentuale;
2) fornisci una motivazione economica di massimo 150 parole, spiegando perché l'allocazione è coerente con il profilo di rischio richiesto;
3) indica separatamente: principale punto di forza dell'allocazione; principale rischio dell'allocazione.

Non aggiungere raccomandazioni di investimento, disclaimer, avvertenze o contenuti diversi da quelli richiesti.

Fonte: elaborazione propria.

La generazione è avvenuta una sola volta. Per Claude e Gemini è stata avviata una nuova conversazione, senza messaggi precedenti, mentre per ChatGPT è stato utilizzato un nuovo account, privo di cronologia e memorie salvate. Questa scelta è stata adottata per ridurre il rischio che gli output fossero influenzati dal contesto di conversazioni precedenti o da forme di personalizzazione dell'account. In tutti e tre i casi il prompt è stato incollato come primo e unico messaggio. Per ciascun modello è stata eseguita una sola generazione, senza rigenerazioni né richieste di modifiche; questo al fine di evitare il cherry-picking delle strategie migliori.

Prima del backtest, ciascuno dei nove portafogli è stato sottoposto a verifica dei criteri di validità - pesi interi e non negativi con somma pari a 100 - implementata nello script riportato in appendice B.4: tutti i portafogli generati hanno superato questo test e vengono tutti riportati e valutati senza esclusioni. Gli output generati dai tre modelli di intelligenza artificiale vengono riportati integralmente in Appendice A.

3.4 Metodologia di backtesting e metriche di valutazione

I vincoli imposti allo script descritto in questo paragrafo sono gli stessi richiesti ai modelli di intelligenza artificiale, discussi nel paragrafo precedente: universo chiuso agli otto indici, portafoglio statico e ribilanciamento trimestrale ai pesi target. Questa corrispondenza è ciò che rende il confronto significativo: ciò che viene chiesto al modello è esattamente ciò che viene implementato successivamente tramite lo script. Il motore, implementato in Python (Appendice B.2), riporta i portafogli, generati dall'intelligenza artificiale, ai propri pesi target il primo giorno di negoziazione di ciascun trimestre solare, mentre tra un trimestre e l'altro i pesi dei vari asset sono lasciati liberi di fluttuare

in base alle condizioni di mercato: sull'intero periodo di valutazione, 2016-2025, questo produce quaranta date di ribilanciamento dei portafogli.

Prima di essere applicato ai dati reali, il motore del backtest è stato sottoposto a sette test di validazione su serie sintetiche costruite manualmente con proprietà note (Appendice B.2): tra queste la verifica che una crescita composta del 5% annuo restituisca un CAGR[43] del 5,00%, che un ribasso costruito esattamente al 30% dal picco fornisca tale output, che un portafoglio investito al 100% in un indice ne replichi esattamente i rendimenti, che i pesi di un portafoglio multi-asset tornino effettivamente ai pesi target alle date di ribilanciamento, che lo Sharpe ratio[44] sia pari a 0 quando il rendimento del portafoglio è pari al tasso risk-free, e che il VaR e l'Expected Shortfall[45] storici coincidano con valori attesi su distribuzioni create ad hoc. Tutti i test sono stati superati prima dell'applicazione ai dati reali: un confronto elementare e concreto sulla riproducibilità dei risultati, nello spirito di trasparenza metodologica adottato per questa tesi.

Per ciascun portafoglio vengono calcolate sette misure sui rendimenti giornalieri del periodo 2016-2025. Il CAGR dipende esclusivamente dal valore iniziale e finale del portafoglio ed è quindi indipendente dalla frequenza di calcolo; la volatilità annualizzata corrisponde alla deviazione standard dei rendimenti giornalieri riportata su base annua; l'indice di Sharpe misura il rendimento in eccesso rispetto al tasso privo di rischio rapportato alla volatilità; il maximum drawdown è la massima perdita dal picco al minimo successivo sulla serie giornaliera - la cui accuratezza, già discussa precedentemente, è uno

[43] Acronimo di Compound Annual Growth Rate, il tasso di crescita annuo composto. Si veda ad es. Zvi Bodie et al., *Investments*, 13th ed. (McGraw-Hill Education, 2021).

[44] William F. Sharpe, "Mutual Fund Performance," *The Journal of Business* 39, no. S1 (1966): 119, https://doi.org/10.1086/294846.

[45] Philippe Artzner et al., "Coherent Measures of Risk," *Mathematical Finance* 9, no. 3 (1999): 203–28, https://doi.org/10.1111/1467-9965.00068.

dei motivi della scelta della frequenza giornaliera rispetto a quella mensile; l'indice di Calmar[46] rapporta il CAGR al maximum drawdown in valore assoluto. A queste misure standard si affiancano anche misure statistiche più elaborate, quali il Value at Risk e l'Expected Shortfall storici al 95%, calcolati sulla distribuzione empirica dei rendimenti giornalieri, senza assumere alcuna ipotesi sulla distribuzione; entrambi riportati su base giornaliera.

Si specifica che il tasso privo di rischio è pari al rendimento medio del periodo 2016-2025 della serie DTB3 (Appendice B.1), utilizzato come costante.

Le stesse metriche, calcolate tramite lo stesso script, sono state applicate sia ai quattro benchmark, sia ai portafogli generati dai tre modelli di intelligenza artificiale (Appendice B.3 e B.4).

[46] Terry W. Young, "Calmar Ratio: A Smoother Tool," *Futures Magazine*, 1991.

CAPITOLO 4
ANALISI EMPIRICA

Questo capitolo presenta i risultati del backtest descritti nel capitolo precedente. Il paragrafo 4.1 richiama l'implementazione delle strategie fornite dai modelli di intelligenza artificiale, mostrandone anche la loro composizione. Il paragrafo 4.2 presenta i risultati dei quattro benchmark, il 4.3 quelli dei portafogli AI ed infine il 4.4 mostra un confronto complessivo tra i due gruppi, che rappresenta la domanda di ricerca ridotta nei suoi termini empirici.

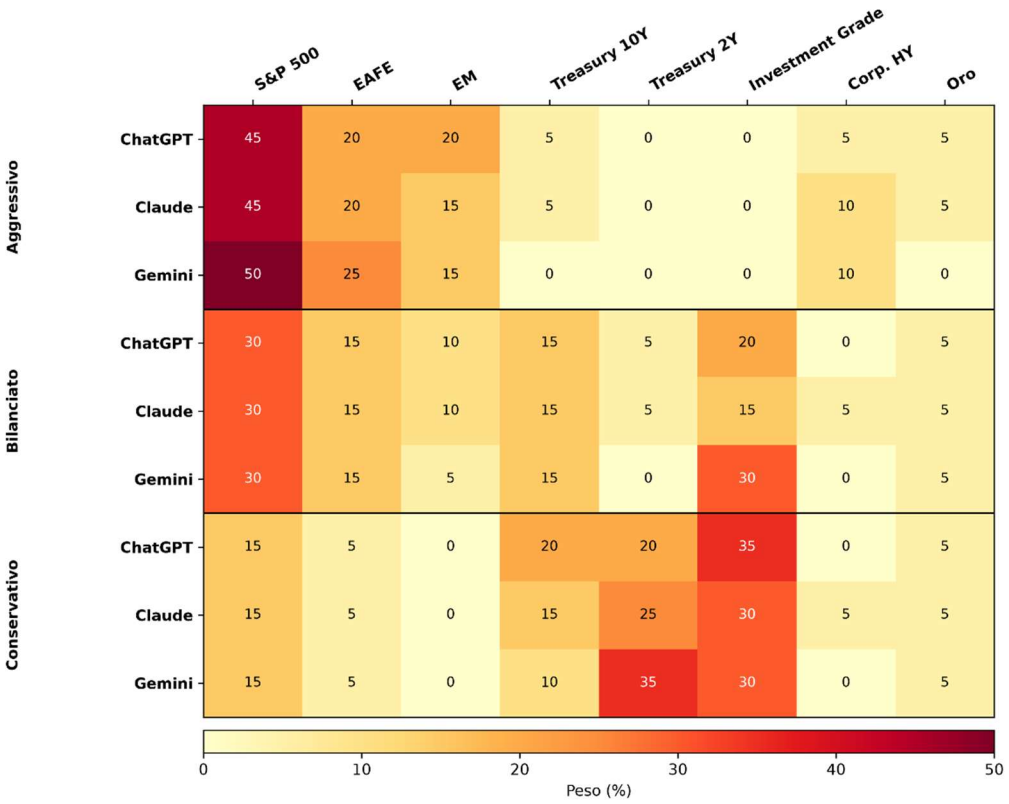
Anticipando ciò che emergerà dai dati: i portafogli generati dai tre modelli di intelligenza artificiale superano, quasi sistematicamente, i rispettivi benchmark tradizionali di riferimento, costruiti sullo stesso profilo di rischio; ma nessuno dei portafogli riesce a eguagliare le prestazioni corrette per il rischio del portafoglio equally-weighted, che resta il termine di paragone più esigente dell'intero esperimento.

4.1 Implementazione delle strategie

Ciascuno dei tredici portafogli – i nove generati dai modelli di intelligenza artificiale e i quattro utilizzati come benchmark – è stato sottoposto allo stesso backtest, sugli stessi dati e con la stessa tipologia di ribilanciamento trimestrale, come descritto nel paragrafo 3.4. La figura 4.1 riporta la composizione dei nove

portafogli generati dai tre modelli – ChatGPT, Claude e Gemini – raggruppati per profili di rischio.

Figura 4.1: Pesi assegnati dai nove portafogli AI, raggruppati per profilo di rischio[47].



Fonte: elaborazione propria.

Il tratto più facilmente individuabile della Figura 4.1 è la forte similarità tra i modelli a parità di profilo di rischio, ma vale la pena precisare dove questa similarità risulta essere più forte e dove lascia spazio a scelte simili, ma diverse. Nel profilo aggressivo tutti i portafogli allocano tra il 45 e il 50 per cento sull'indice S&P500 e tra il 20 e il 25 per cento sull'azionario dei mercati

[47] Per ciascuna riga (portafoglio), la somma dei pesi assegnati alle otto asset class è pari al 100 per cento

sviluppati; l'eccezione principale riguarda Gemini, il quale alloca la quota maggiore all'azionario (90 per cento), contro l'80-85 di ChatGPT e Claude, non allocando alcun peso in Treasury a 10 anni e all'oro; risultando così il portafoglio più aggressivo tra i tre selezionati.

Sul profilo bilanciato la similarità è ancora più marcata, al punto da essere quasi totale: S&P500 (30 per cento), EAFE (15 per cento), Treasury decennale (15 per cento) e oro (5 per cento) sono identici in tutti e tre modelli. Il punto principale di divergenza tra i modelli bilanciati riguarda la componente obbligazionaria a breve termine e creditizia: ChatGPT e Claude ripartiscono il rimanente tra Treasury a due anni e investment grade, mentre Gemini omette completamente i Treasury a due anni e gli High Yield e alloca tutto il restante nella componente investment grade, dimostrandosi ancora una volta l'outlier principale tra i tre modelli.

Sul profilo conservativo la stessa logica si ripete in forma ancora più marcata: l'intero blocco azionario - S&P500 al 15 per cento, EAFE al 5 per cento, emergenti allo zero per cento - e oro al 5 per cento sono identici per tutti e tre i portafogli, così come la quota investment grade è molto simile, compresa tra il 30 e il 35 per cento. La differenza si concentra per intero nella ripartizione tra Treasury a due e a dieci anni: ChatGPT mantiene un equilibrio paritario (20 per cento per entrambi), Claude si sposta leggermente sul breve termine (25 per cento per i Treasury a due anni contro il 15 per cento su quelli a dieci anni), Gemini lo fa in modo più netto (35 per cento per i Treasury a due anni e il 10 per cento sui Treasury a dieci anni).

Le allocazioni non sono quindi mai identiche, ma la struttura sottostante rimane la stessa: nessuno dei tre modelli propone, per lo stesso profilo di rischio, una struttura di portafoglio radicalmente diversa dagli altri.

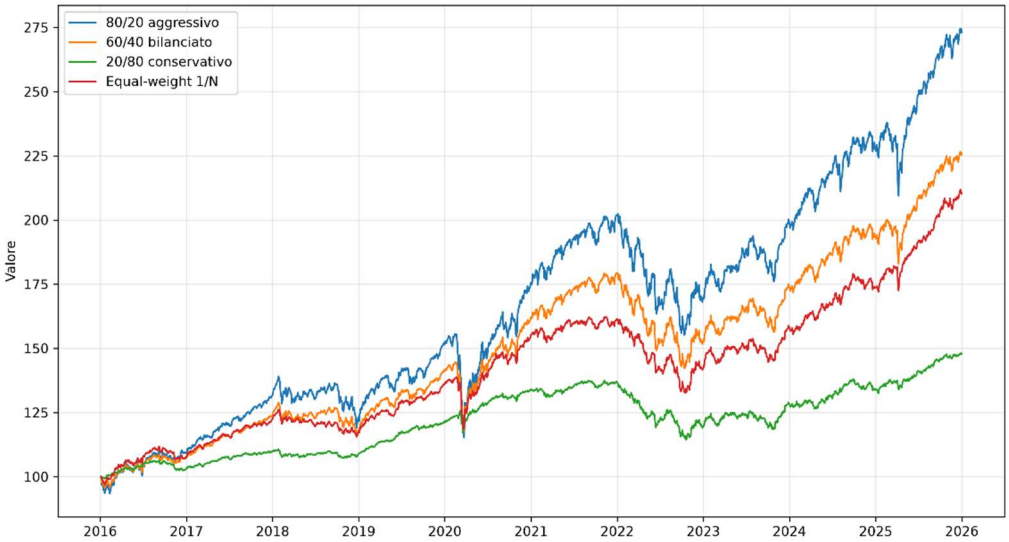
Un'eccezione va segnalata: i portafogli di Gemini sono, tra i tre, quelli che si differenziano maggiormente dagli altri e lo fanno in due direzioni opposte: più esposto sull'azionario nel profilo aggressivo, come appena visto, e più prudente nel profilo conservativo, dove destina la maggiore allocazione (35 per cento) sui Treasury a breve termine contro il 20-25 per cento di ChatGPT e Claude, a fronte di un blocco azionario identico tra i tre modelli. Le implicazioni di queste scelte saranno discusse con maggiore approfondimento nei paragrafi successivi.

4.2 Risultati dei portafogli benchmark

I risultati dei benchmark mostrano un ordinamento coerente con il livello di rischio assegnato ex ante. Passando dal portafoglio conservativo al bilanciato ed infine all'aggressivo, aumentano sia il rendimento sia le principali misure di rischio. Il benchmark 20/80 realizza un CAGR del 3,99 per cento con una volatilità annualizzata del 4,47 per cento; il 60/40 raggiunge rispettivamente l'8,49 e l'8,57 per cento; mentre l'80/20 presenta il rendimento più elevato, pari al 10,60 per cento, a fronte di una volatilità annua pari all'11,43 per cento.

Lo stesso ordinamento emerge, coerentemente, per le perdite massime osservate nel periodo di riferimento. Il maximum drawdown aumenta, in valore assoluto, dal 16,70 per cento del portafoglio conservativo al 20,71 per cento del bilanciato, fino al 25,94 per cento dell'aggressivo. I risultati indicano, quindi, che la maggiore esposizione azionaria si è tradotta in maggiori rendimenti, a fronte di un aumento del rischio complessivo. Di conseguenza, i tre benchmark hanno prodotto ex post una graduazione coerente rispetto ai vari profili di rischio, per i quali erano stati costruiti.

Figura 4.2: Andamento del valore dei portafogli benchmark (2016-2025, base 100)



Fonte: elaborazione propria.

La Figura 4.2 mostra l'evoluzione del valore cumulato dei quattro benchmark nel periodo 2016-2025, assumendo un valore iniziale convenzionalmente pari a 100. Il benchmark 80/20 registra l'incremento cumulato più elevato, ma anche oscillazioni più pronunciate durante le principali fasi di correzione del periodo, in particolare nel crollo del 2020, in concomitanza con lo shock provocato dalla pandemia di Covid-19, e nel 2022, nel contesto dell'accelerazione dell'inflazione e del conseguente irrigidimento della politica monetaria da parte delle principali banche centrali, attuato anche attraverso l'innalzamento dei tassi di interesse.

Il portafoglio equal-weight presenta, invece, caratteristiche differenti rispetto agli altri benchmark. Il suo CAGR, pari al 7,73 per cento, è inferiore a quello dell'80/20 e del 60/40, ma superiore rispetto a quello del 20/80. Sotto il profilo del rischio, la volatilità del 6,27 per cento e il maximum drawdown del 18,25 per cento risultano inferiori rispetto ai benchmark aggressivo e bilanciato, ma superiori a quelli del benchmark conservativo. L'equal-weight si colloca quindi,

in termini di rischio, tra il profilo conservativo e quello bilanciato, pur ottenendo un rendimento annualizzato relativamente vicino a quello del profilo bilanciato.

Il risultato più rilevante per il portafoglio equal-weight emerge dalle misure corrette per il rischio. L'indice di Sharpe, infatti, raggiunge lo 0,877 - valore più elevato in assoluto - nel portafoglio equal-weight, contro valori compresi tra 0,417 e 0,751 per i tre benchmark di profilo; anche l'indice di Calmar, pari a 0,424, è il più elevato tra i quattro portafogli. Nel periodo analizzato, la strategia equiponderata è quella che ha presentato il miglior rapporto tra rendimento e rischio, pur applicando una regola di allocazione particolarmente semplice e priva di ottimizzazione.

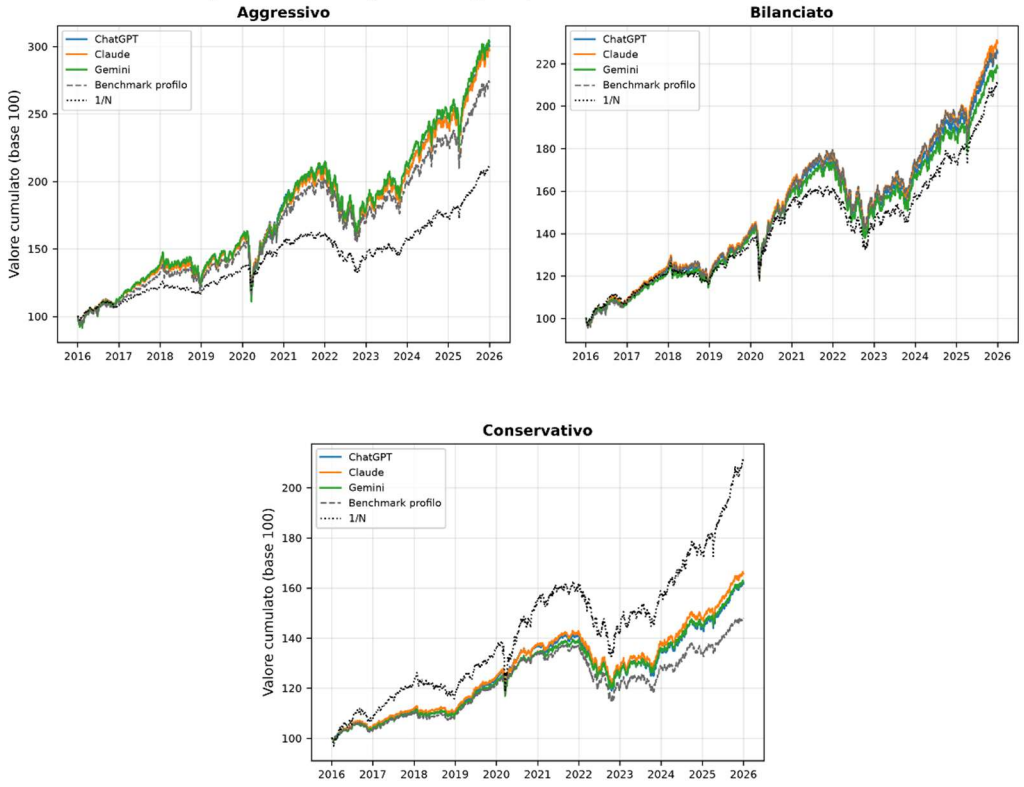
Una possibile spiegazione potrebbe risiedere nella maggiore ampiezza della diversificazione: il portafoglio equal-weight è infatti l'unico benchmark che distribuisce capitale su tutte le otto asset class presentate, includendo anche oro e obbligazioni high yield. Il vantaggio osservato potrebbe riflettere l'effetto di una maggiore diversificazione e di una differente distribuzione dei pesi; i risultati del backtest non consentono di isolare il contributo delle singole asset-class. Il risultato appare comunque coerente con la letteratura sulla diversificazione naïve presentata precedentemente nel paragrafo 2.1.

Le misure di rischio di coda completano questa lettura: l'equal-weight presenta un Value at Risk giornaliero al 95 per cento dello 0,58 per cento e un Expected Shortfall giornaliero al 95 per cento dello 0,90 per cento, corrispondente alla perdita media registrata nelle osservazioni che superano la soglia del VaR; valori che risultano, ancora una volta, inferiori rispetto ai benchmark 80/20 e 60/40, ma superiori rispetto al 20/80. Di conseguenza, anche sotto questo profilo di rischio, il portafoglio equiponderato occupa una posizione intermedia in termini di rischio assoluto, pur risultando il più efficiente quando rischio e rendimento vengono valutati congiuntamente.

4.3 Risultati delle strategie generate dall'intelligenza artificiale

I nove portafogli generati dall'intelligenza artificiale sono stati sottoposti allo stesso motore di backtest e alle stesse metriche già utilizzate per i benchmark. Prima di entrare nel dettaglio delle singole misure, la Figura 4.3 offre una prima lettura d'insieme sulla crescita del capitale nei tre diversi profili di rischio.

Figura 4.3: Valore cumulato dei portafogli AI, confrontati con benchmark di profilo e l'equal weight (2016-2025, base 100)



Fonte: elaborazione propria.

La Figura 4.3 mostra la crescita del capitale per i tre profili di rischio. Nel profilo aggressivo le tre curve dei modelli restano quasi sempre sia sopra il benchmark di profilo, sia sopra il portafoglio equal-weight per l'intero periodo di osservazione. Nel profilo bilanciato, invece, il quadro appare più sfumato:

solo Claude si mantiene stabilmente sopra il benchmark 60/40, mentre ChatGPT e Gemini vi restano sotto per gran parte del decennio, pur chiudendo il periodo su valori relativamente vicini – un esito coerente con le differenze contenute osservate nei rispettivi CAGR – a differenza di un vantaggio più netto presente sul profilo aggressivo. Nel profilo conservativo i tre modelli si mantengono stabilmente al di sopra del benchmark di riferimento, ma al di sotto dell'equal-weight, pur non essendo un benchmark di quel profilo; una lettura coerente con il CAGR più basso osservato per questo profilo (4,93-5,19 per cento contro il 7,73 per cento dell'equal-weight).

I nove portafogli generati dai modelli di intelligenza artificiale mostrano risultati coerenti con la struttura dei pesi osservata nella Figura 4.1. Nel profilo aggressivo, ChatGPT e Claude rendono rispettivamente l'11,65 e l'11,54 per cento annuo con Sharpe di 0,793 e 0,806; Gemini rende leggermente di più (11,76 per cento) – a fronte di una maggiore esposizione azionaria – ma con una volatilità sensibilmente maggiore (13,18 contro rispettivamente il 12,11 e l' 11,75 per cento degli altri due modelli) e il drawdown più profondo dell'intero esperimento (-32,05 per cento): il suo Sharpe Ratio, 0,747, è inoltre l'unico a non superare il valore del corrispettivo benchmark di profilo. Questo rappresenta la conferma quantitativa di quanto già discusso nel paragrafo 4.1: rinunciare a Treasury e oro riduce la diversificazione e produce maggiore volatilità e ribassi più accentuati, senza – in questo caso – un beneficio di rendimento sufficientemente elevato per compensare i maggiori rischi.

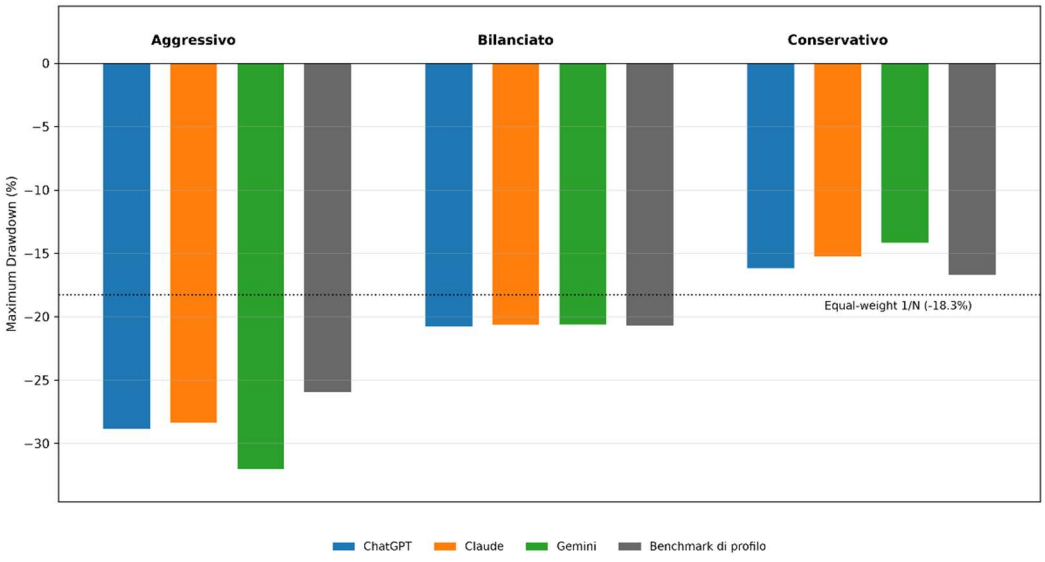
Nel profilo bilanciato i tre modelli risultano essere molto simili tra loro (CAGR compreso tra 8,12 e 8,70 per cento, Sharpe tra 0,782 e 0,807) e lo stesso vale per il profilo conservativo (CAGR compresi tra 4,93 e 5,19 per cento), dove però si verifica anche un'altra regolarità: tutti e tre i portafogli della categoria presentano Sharpe Ratio (0,594-0,676) sensibilmente più bassi rispetto a quelli

dei profili più rischiosi, un risultato coerente con un decennio nel quale l'esposizione azionaria è stata remunerata più favorevolmente rispetto a quella obbligazionaria – una condizione di mercato, non un difetto dei portafogli.

Un altro dato meritevole di discussione: Claude presenta, su tutti e tre i profili di rischio, l'indice di Sharpe più elevato (0,806 sull'aggressivo, 0,807 sul bilanciato e 0,676 sul conservativo). La ricorrenza di questo risultato merita attenzione, seppure risulti difficile concludere che il modello sia generalmente superiore a ChatGPT e Gemini; tale considerazione verrà discussa più ampiamente nel Capitolo 5.

Anche l'indice di Calmar conferma il quadro già delineato per il profilo aggressivo: Gemini, già identificato come il caso più esposto al rischio nella Figura 4.3, presenta il valore più basso tra le strategie aggressive generate dai modelli, pari a 0,367, coerente con il maximum drawdown più profondo osservato nell'intero esperimento.

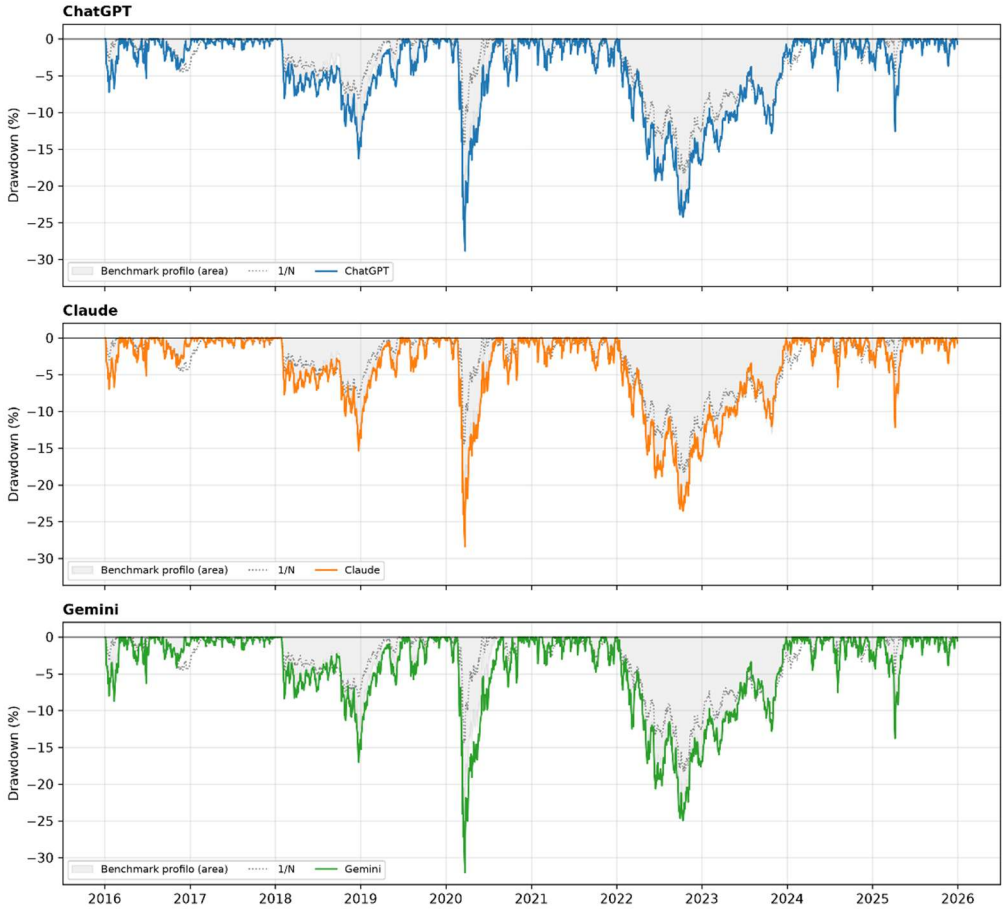
Figura 4.4: Maximum Drawdown: portafogli AI, benchmark di profilo e equal weight.



Fonte: elaborazione propria.

La Figura 4.4 offre una vista d'insieme sui drawdown dei tre profili: nel profilo aggressivo la barra di Gemini è visibilmente la più lunga, mentre per i profili bilanciato e conservativo le quattro barre di ciascun gruppo (tre modelli e il benchmark di riferimento) restano molto più vicine tra loro, coerentemente con la maggior somiglianza dei modelli già evidenziata precedentemente.

Figura 4.5: Drawdown dal picco precedente, profilo aggressivo, per modello (2016-2025).



Fonte: elaborazione propria.

La Figura 4.5 approfondisce i drawdown del profilo aggressivo, l'unico profilo a mostrare differenze rilevanti tra modelli: nei tre pannelli la curva di

Gemini si stacca sistematicamente al ribasso nei confronti degli altri modelli nei principali momenti di stress – marzo 2020 e a partire da inizio 2022 – mentre ChatGPT e Claude restano più vicini all'area del benchmark; questo è coerente con una maggiore esposizione azionaria del portafoglio Gemini, che consente ad esso maggiori rialzi, ma a fronte di una maggiore volatilità nei principali momenti di stress di mercato.

4.4 Confronto tra strategie

La Tabella 4.1 riporta le sette metriche per tutti i portafogli ed i corrispettivi benchmark, con l'equal-weight in apertura come riferimento comune e, per ciascun profilo, il benchmark seguito dai tre portafogli generati.

Tabella 4.1: Metriche di sintesi, tutti i portafogli.

Portafoglio	CAGR	Vol.	Sharpe	MaxDD	Calmar	VaR 95%	ES 95%
Equal-weight (1/N)	7,73%	6,27%	0,877	-18,25%	0,424	0,58%	0,90%
80/20 (benchmark)	10,60%	11,43%	0,751	-25,94%	0,408	1,04%	1,72%
ChatGPT - Aggressivo	11,65%	12,11%	0,793	-28,84%	0,404	1,07%	1,81%
Claude - Aggressivo	11,54%	11,75%	0,806	-28,36%	0,407	1,05%	1,76%
Gemini - Aggressivo	11,76%	13,18%	0,747	-32,05%	0,367	1,18%	1,99%
60/40 (benchmark)	8,49%	8,57%	0,743	-20,71%	0,410	0,79%	1,28%
ChatGPT - Bilanciato	8,48%	7,94%	0,793	-20,76%	0,408	0,74%	1,18%
Claude - Bilanciato	8,70%	8,07%	0,807	-20,63%	0,421	0,74%	1,20%
Gemini - Bilanciato	8,12%	7,60%	0,782	-20,60%	0,394	0,69%	1,13%

(Tabella 4.1 segue in pagina successiva)

Tabella 4.1 (prosegue dalla pagina precedente)

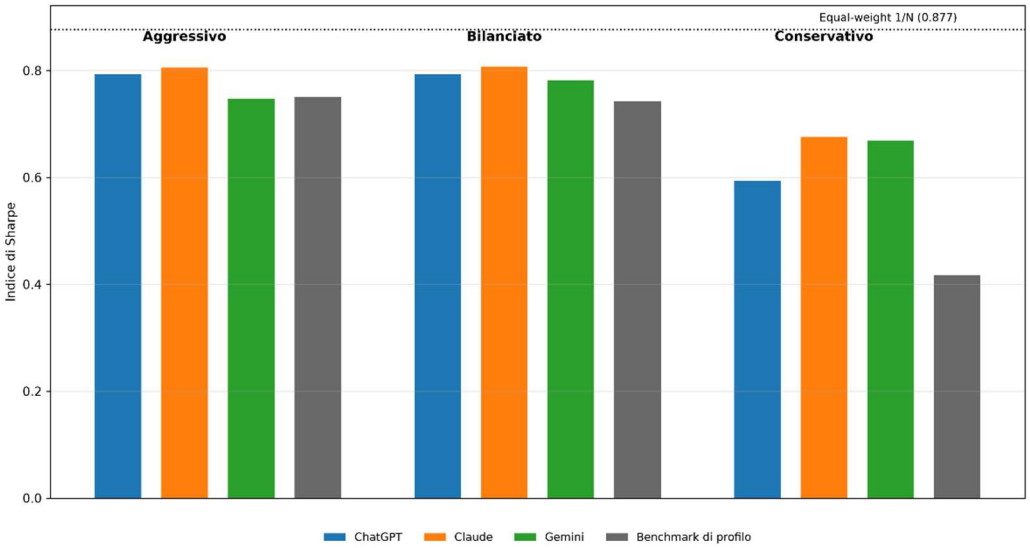
20/80 (benchmark)	3,99%	4,47%	0,417	-16,70%	0,239	0,45%	0,67%
ChatGPT - Conservativo	4,93%	4,66%	0,594	-16,18%	0,305	0,46%	0,69%
Claude - Conservativo	5,19%	4,45%	0,676	-15,27%	0,340	0,43%	0,66%
Gemini - Conservativo	4,97%	4,17%	0,669	-14,18%	0,351	0,40%	0,62%

Fonte: elaborazione propria su dati LSEG e output dei 3 modelli (Appendice B.3 e B.4)

Dalla Tabella 4.1 emergono con chiarezza due risultati principali. Primo: in otto casi su nove i portafogli generati dai modelli presentano uno Sharpe Ratio superiore rispetto al corrispettivo benchmark di riferimento. L'unica eccezione si verifica per il portafoglio aggressivo creato da Gemini, il quale presenta uno Sharpe di 0,747 contro lo 0,751 dell'80/20: una differenza molto contenuta. Secondo: nessuno dei nove portafogli generati dai modelli AI riesce ad avere uno Sharpe Ratio superiore rispetto a quello del portafoglio equal-weight, il cui indice di Sharpe, pari a 0,877, risulta essere il valore più elevato dell'intera tabella. Il vantaggio riflette, almeno in parte, un'allocazione estesa all'intero universo investibile, non vincolata ad uno specifico profilo di rischio.

Il confronto presenta, tuttavia, caratteristiche differenti a seconda del profilo che si va ad analizzare. Nel profilo aggressivo tutti e tre i portafogli AI ottengono un CAGR superiore a quello dell'80/20, ma anche una maggiore volatilità e drawdown più elevati. Nel profilo bilanciato i modelli riescono a ridurre la volatilità rispetto al benchmark 60/40, mantenendo comunque rendimenti annualizzati complessivamente simili. Nel profilo conservativo, tutti e tre superano il CAGR del 20/80 e presentano anche drawdown meno profondi, un risultato che tuttavia va contestualizzato alla luce del periodo di verifica: la finestra temporale 2016-2025 include il 2022, anno caratterizzato da forti perdite sui comparti obbligazionari, che hanno penalizzato in particolare il benchmark 20/80.

Figura 4.6: Indici di Sharpe: portafogli AI, benchmark di profilo e dell'equal-weight.



Fonte: elaborazione propria.

La Figura 4.6 rende queste osservazioni di immediata lettura: in ciascuno dei tre gruppi – ognuno per un profilo di rischio differente – le barre colorate dei modelli superano quasi sempre la barra grigia del proprio benchmark di riferimento, mentre nessuna delle dodici barre raggiunge la linea tratteggiata del portafoglio equal weight.

Il vantaggio dei portafogli AI in termini di Sharpe non si traduce, però, anche in un vantaggio in termini di rischio di coda. Nel profilo aggressivo ChatGPT e Claude presentano Sharpe Ratio più elevati rispetto al benchmark, nonostante abbiano VaR e Expected Shortfall lievemente più elevati; Gemini combina invece VaR e Expected Shortfall superiori con uno Sharpe inferiore.

Nel profilo bilanciato, invece, la relazione si inverte: tutti e tre i portafogli AI presentano VaR ed Expected Shortfall inferiori al benchmark 60/40, un vantaggio che si accompagna a Sharpe più elevati. Nel profilo conservativo i risultati variano tra modelli: ChatGPT presenta valori leggermente superiori

rispetto al benchmark di riferimento 20/80, mentre Claude e Gemini presentano valori inferiori, con Gemini che presenta il vantaggio più marcato.

In conclusione il miglioramento del rendimento corretto per la volatilità, misurato mediante lo Sharpe Ratio, emerge in otto dei nove confronti analizzati, mentre il miglioramento del rischio di coda rispetto al benchmark di profilo è comune a tutti e tre i modelli esclusivamente nel profilo bilanciato.

CAPITOLO 5
DISCUSSIONE

Questo capitolo discute i risultati presentati nel Capitolo 4, cercando di andare oltre la semplice descrizione dei dati già visti. I dati segnalano un quadro non uniforme: i portafogli generati dai modelli presentano strutture simili, ma non sempre ciò si traduce in uguali risultati corretti per il rischio; soltanto il profilo conservativo riesce a presentare vantaggi di Sharpe Ratio più ampi rispetto ai benchmark di profilo definiti in Tabella 3.2, ma nessuno di essi riesce mai a superare la semplice diversificazione equiponderata. Il capitolo propone un'interpretazione economica di questi risultati (paragrafo 5.1), ne riconosce i limiti metodologici (paragrafo 5.2) e ne trae alcune implicazioni per l'impiego pratico di strumenti di intelligenza artificiale generativa all'interno dei processi di asset allocation (paragrafo 5.3), e ne verifica infine la riproducibilità attraverso una rigenerazione dei portafogli (paragrafo 5.4).

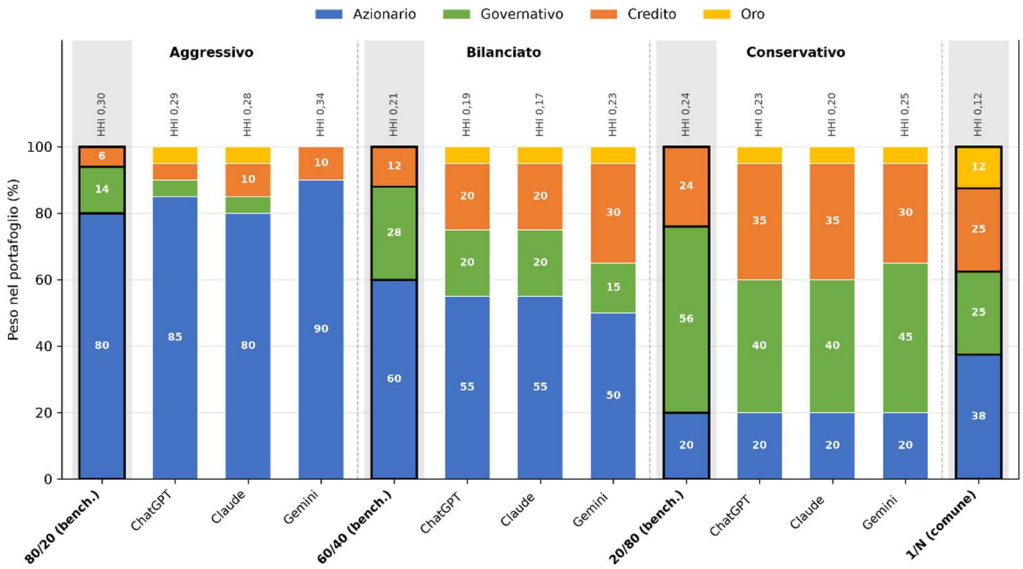
5.1 Interpretazione dei risultati

I risultati presentati nel Capitolo precedente permettono di tornare alla domanda di ricerca proposta nel paragrafo 1.3, con una risposta articolata, non riducibile ad un giudizio univoco.

Prima di entrare nel dettaglio dei singoli profili – aggressivo, bilanciato e conservativo – la Figura 5.1 può essere utile ad offrire una lettura d'insieme utile

per il confronto dei portafogli generati rispetto ai corrispettivi benchmark di riferimento. Il grafico raggruppa gli otto indici utilizzati per la generazione dei portafogli in quattro macro categorie – azionario, governativo, credito e oro – per ciascuno dei quattro benchmark utilizzati (evidenziati da contorno nero e sfondo grigio) e ciascuno dei nove portafogli AI, organizzati per profilo di rischio. Alcuni pattern emergono in maniera piuttosto evidente: nel profilo aggressivo la barra di Gemini è quasi interamente blu, priva delle porzioni verde e gialla presenti negli altri due modelli, segnalando così una forte concentrazione nei confronti del mercato azionario; nel profilo bilanciato la componente azionaria dei tre modelli è visibilmente inferiore rispetto a quella del benchmark 60/40; nel profilo conservativo la componente azionaria resta pressoché identica, mentre cambia sensibilmente l'allocazione della componente obbligazionaria, con una maggiore allocazione di capitale nella componente creditizia, piuttosto che in quella governativa.

Figura 5.1: Esposizione aggregata per asset class macro e indice di Herfindahl-Hirschman: benchmark e portafogli AI



Fonte: elaborazione propria su Tabella 3.2 e Appendice A.

La categoria "Credito" comprende l'indice FTSE US Broad Investment-Grade Bond TR, che include sia titoli di stato governativi che corporate investment grade, insieme all'indice FTSE US High-Yield Market TR. La distinzione rispetto alla categoria "Governativo" riflette quindi la scelta tra un'esposizione diretta ai Treasury e un'esposizione obbligazionaria più diversificata comprensiva sia di titoli di stato che corporate.

La Figura 5.1 non si limita a mostrare la composizione per macrocategoria: sopra ciascuna barra viene riportato anche il corrispettivo indice di Herfindahl-Hirschman (HHI)[48], il quale sintetizza il grado di concentrazione interno a ciascun portafoglio, un'informazione che la singola suddivisione per macrocategorie non riuscirebbe a fornire, dato che due portafogli con la stessa quota azionaria possono risultare molto differenti in termini di diversificazione effettiva. L'HHI è calcolato come la somma dei quadrati dei pesi di ciascun indice (non aggregati nelle quattro macro categorie): $HHI = \Sigma wi^2$, dove wi è il peso di ciascun indice. Valori più bassi indicano portafogli più diversificati; viceversa valori più elevati identificano una maggiore concentrazione. Il portafoglio equal-weight ottiene, per costruzione, il valore più basso in assoluto (HHI = 0,125). Il dato risulta essere coerente con l'ipotesi che la distribuzione uniforme dei pesi abbia contribuito alla sua elevata efficienza ex-post, ma non risulta essere sufficiente a spiegare, da solo, il vantaggio di Sharpe Ratio.

Si precisa però che l'HHI misura la concentrazione dei pesi assegnati a ciascun indice, non la concentrazione del rischio effettivo del portafoglio: un portafoglio che alloca capitale tra S&P500, EAFE e mercati emergenti riduce l'HHI, ma queste tre componenti risultano essere fortemente correlate tra loro; di conseguenza la diversificazione, che si traduce in una diminuzione del HHI,

[48] Albert O. Hirschman, "The Paternity of an Index," *The American Economic Review* 54 (1964): 761–62, https://doi.org/10.2307/1818582.

appare solo “sulla carta” non traducendosi anche in una proporzionale riduzione del rischio. L’indice risulta quindi essere un buon indicatore sintetico per comprendere quanto sia distribuito il capitale all’interno del portafoglio, ma non presenta una misura diretta del rischio sottostante.

Una parte della variabilità osservata tra i modelli, anche all’interno dello stesso profilo, trova origine a monte, a causa della natura stessa del mandato assegnato. Ai tre modelli è infatti stata fornita un’etichetta qualitativa – aggressivo, bilanciato e conservativo – e non un obiettivo di rischio quantitativo, quali potrebbero essere per esempio limiti di volatilità o di Value at Risk. La coerenza tra modelli osservata nella creazione della struttura dei portafogli (paragrafo 4.1) rappresenta, quindi, una corretta interpretazione del mandato, non di calibrazione numerica: ogni modello ha tradotto il profilo di rischio in allocazioni specifiche, ed è proprio in questa traduzione da una caratteristica qualitativa ad una quantitativa che si può concentrare gran parte della dispersione osservata tra i risultati. Questa dispersione si può osservare, ad esempio, nell’ampiezza della volatilità osservata all’interno del profilo aggressivo (11,75-13,18 per cento) a fronte di uno stesso obiettivo.

Più in particolare, nel profilo aggressivo tutti e tre i modelli sovrappesano i mercati emergenti rispetto all’allocazione del 12 per cento del benchmark 80/20: ChatGPT arriva ad un’allocazione su questi pari al 20 per cento – quota più alta dell’intero esperimento – Claude e Gemini al 15 per cento. Questa rappresenta una scelta che, nel decennio in esame, ha probabilmente pesato sul rendimento di tutti e tre i portafogli del profilo, in quanto l’azionario emergente ha sottoperformato nettamente l’azionario statunitense. Tuttavia, se il sovrappeso sugli emergenti fosse l’unica caratteristica ad aver zavorrato i portafogli, ChatGPT dovrebbe risultare il portafoglio peggiore tra i tre, invece il suo Sharpe Ratio (0,793) supera quello di Gemini (0,747), che ha un’esposizione ai mercati

emergenti inferiore. La differenza si può spiegare, più facilmente, tenendo in considerazione l'HHI: Gemini presenta il valore più alto tra i tre modelli del profilo (0,345 contro 0,290 di ChatGPT e 0,280 di Claude), risultato coerente con l'assenza totale di Treasury e oro, che tendono ad abbassare così la diversificazione complessiva del portafoglio. Il sovrappeso dei mercati emergenti non è quindi sufficiente, da solo, a spiegare il risultato più debole di Gemini. Una possibile concausa è la struttura complessivamente più concentrata del portafoglio, caratterizzato da un'esposizione azionaria pari al 90 per cento, a fronte di una totale assenza di Treasury e oro, asset class potenzialmente in grado di offrire una maggiore diversificazione rispetto ad un portafoglio che alloca la maggior parte del capitale in azioni e high yield.

In sintesi, nel profilo aggressivo il fattore critico parrebbe essere l'assenza di elementi difensivi, piuttosto che il grado di concentrazione sui mercati emergenti, comune del resto a tutti e tre i modelli.

Nel profilo bilanciato, tutti e tre i modelli destinano il 5 per cento del portafoglio all'oro, un asset class completamente assente nel benchmark di riferimento 60/40. Si tratta di una scelta che ha contribuito positivamente al rendimento dei portafogli, dato che nel periodo in esame 2016-2025 il prezzo dell'oro ha registrato una delle fasi più favorevoli della sua storia recente, con rialzi annui attorno al 25 per cento nel 2020, al 27 per cento nel 2024 e ad oltre il 60 per cento nel 2025[49]. Questo potrebbe contribuire ad aiutare a spiegare perché i tre modelli, nonostante una componente azionaria inferiore al benchmark di riferimento 60/40 (50-55 per cento contro il 60 per cento), non abbiano perso terreno in termini di rendimento: la componente aurea ha, almeno

[49] Rendimenti annui calcolati dallo scrivente sulla serie Gold Spot (USD/oncia, ticker XAU) utilizzata in questo lavoro (Appendice B.1). Su 19 anni disponibili (2007-2025), il 2025 risulta il migliore dell'intero campione (+64,4 per cento), seguito da 2007 (+30,9 per cento), 2010 (+29,6 per cento), 2024 (+27,2 per cento) e 2020 (+25,0 per cento).

in parte, probabilmente compensato la minore esposizione al rischio. Questo meccanismo - un vantaggio che dipende da un'asset class assente dal benchmark di confronto - viene ripreso criticamente nel paragrafo 5.2, dove si discute quanto i risultati qui descritti possano dipendere dalla specifica composizione del benchmark adottata. Anche in questo profilo l'HHI può fornire informazioni utili a leggere le differenze tra modelli: Claude con 0,175 risulta il portafoglio più diversificato dei tre (contro 0,190 di ChatGPT e 0,230 di Gemini). Va precisato che un HHI più contenuto non implica rendimenti più elevati. La differenza tra Claude e ChatGPT – identici in tutto fuorché la componente creditizia, dove Claude destina il 5 per cento al credito High Yield Corporate – è coerente con la combinazione di una duration generalmente più contenuta e di un maggiore carry dell'high yield. Nel 2022, in particolare, periodo già discusso a proposito del profilo conservativo, l'Aggregato IG ha perso il 13,3 per cento contro l'11 per cento dell'High Yield, un divario coerente con la maggiore sensibilità dei tassi della componente investment grade a più lunga duration. Il vantaggio dell'High Yield non si è limitato, però, esclusivamente al 2022, né sembra riducibile ad una semplice compensazione per il maggior rischio assunto: nel periodo in esame, dunque, l'High Yield presenta sì una volatilità annualizzata superiore all'Aggregato (5,46 contro 4,71 per cento), come atteso data la maggior rischiosità, ma il rendimento aggiuntivo risulta più che proporzionale rispetto al rischio assunto, con un rapporto rendimento/rischio nettamente più favorevole (1,20 contro 0,43).

Fa eccezione il massimo drawdown, più elevato per l'High Yield (-21,56 contro -18,61 per cento), verosimilmente riconducibile allo shock di marzo 2020, momento in cui il rischio di credito si è materializzato in maniera più profonda rispetto al rischio di duration. Di conseguenza, nel periodo di backtest, le caratteristiche dell'high yield si sono rivelate più favorevoli di quelle

dell'Aggregate IG, sebbene non sia possibile attribuire tali differenze ad un singolo fattore.

Nel profilo conservativo, oltre alla scelta di duration già discussa nel paragrafo 4.4 sulla ripartizione tra Treasury a dieci o a due anni, altri due elementi aiutano a spiegare il vantaggio dei portafogli generati dai modelli rispetto al benchmark. In primo luogo nessuno dei tre modelli alloca capitale ai mercati emergenti (0 per cento, contro il 3 per cento del benchmark 20/80), un'esclusione che si è rivelata vincente in quanto nel decennio del backtest l'azionario emergente ha sottoperformato in maniera evidente l'azionario statunitense. In secondo luogo, come nel profilo bilanciato, tutti e tre i portafogli dei modelli allocano il 5 per cento sull'oro, completamente assente nel benchmark di riferimento; questa rappresenta un'esposizione che fornisce bassa correlazione con le altre asset class del portafoglio e, particolarmente, nel decennio d'esame si è rivelata particolarmente remunerativa. L'HHI, in questo profilo, conferma in parte quanto già visto precedentemente per gli altri due profili: il portafoglio più diversificato, quello di Claude (HHI pari a 0,205), risulta essere anche quello a presentare il migliore CAGR (5,19 per cento) e il miglior Sharpe Ratio (0,676) dei tre. La relazione non è però pienamente lineare: Gemini, pur presentando l'HHI più elevato tra i tre modelli (0,250), supera comunque ChatGPT – che ha un HHI più basso (0,230) – sia in CAGR (4,97 per cento contro 4,93 per cento) sia in Sharpe Ratio (0,669 contro 0,594). Coerentemente con quanto appena esposto sulla duration, la spiegazione più probabile è che Gemini allochi la quota più elevata della sua componente obbligazionaria su Treasury a breve termine (35 per cento, contro il 20 per cento di ChatGPT); una scelta che nel 2022 si è rivelata particolarmente protettiva, a prescindere dal grado di concentrazione complessivo del portafoglio.

Anche nel profilo conservativo, quindi, il vantaggio sembra derivare dalla componente difensiva – duration, esclusione degli emergenti e presenza dell'oro – piuttosto che da un principio di maggiore o minore concentrazione. Come nel profilo bilanciato, va però segnalato che il ruolo dell'oro in questo vantaggio viene riconsiderato nel paragrafo 5.2, alla luce della sua assenza dal benchmark di confronto.

Un secondo risultato, trasversale a tutti e tre i profili, è che nessuno dei nove portafogli generati dai modelli di intelligenza artificiale è riuscito a superare lo Sharpe Ratio del portafoglio equiponderato. Il dato, già discusso nel paragrafo 4.2, può essere ora letto in luce anche dell'HHI. Il portafoglio equiponderato presenta infatti un valore pari a 0,125, il più basso dell'intero esperimento. Il valore immediatamente successivo è quello di Claude nel profilo bilanciato, pari a 0,175, superiore di circa il 40 per cento rispetto al portafoglio con HHI più basso. Va peraltro rilevato che, anche per il portafoglio equiponderato, il vantaggio fa riferimento al periodo di backtest e potrebbe essere meno robusto se calcolato su altri periodi. Va anche ricordato che l'equal weight risponde ad un mandato differente rispetto ai benchmark e ai portafogli dei modelli: esso non è vincolato ad alcun obiettivo di rischio specifico e di conseguenza, il confronto deve essere visto, esclusivamente, a fini comparativi.

Un ultimo elemento che merita discussione a parte: la ricorrenza, già segnalata nel capitolo precedente, di valori di Sharpe ratio più elevati per i portafogli generati da Claude rispetto agli altri modelli e ai corrispettivi benchmark di profilo, su tutti e tre i profili di rischio. Questo dato va trattato con cautela: gli scarti rispetto agli altri modelli sono piccoli ed il disegno prevede un'unica generazione per modello e profilo. Detto ciò, l'HHI potrebbe fornire un ulteriore elemento per capire, quantitativamente, i motivi di questo avvenimento: Claude presenta sempre, all'interno dei singoli profili, l'HHI più

basso (0,280 sul profilo aggressivo, 0,175 sul bilanciato, 0,205 sul conservativo). Resta comunque un limite di fondo: con una sola generazione su un solo periodo di osservazione non risulta possibile capire se si tratti di una caratteristica sistematica dell'approccio del modello o semplicemente un esito casuale di questa specifica generazione.

Nel complesso, i risultati suggeriscono che i tre modelli abbiano interpretato in maniera economicamente coerente i profili di rischio richiesti, senza tuttavia tradurli in una calibrazione quantitativa uniforme. Le differenze osservate non sembrano riconducibili a un unico fattore. Tra i modelli, all'interno dello stesso profilo, pesano soprattutto il grado di concentrazione dei pesi e la diversa combinazione di esposizione azionaria, duration e rischio di credito. Rispetto ai benchmark di profilo, pesa invece una scelta comune ai profili bilanciato e conservativo – l'inclusione dell'oro, assente in entrambi i benchmark – mentre il ruolo dei mercati emergenti varia a seconda del profilo preso in considerazione: penalizzante nell'aggressivo, dove tutti e tre i modelli li sovrappesano e favorevole nel conservativo, dove vengono invece esclusi del tutto. L'indice di Herfindahl-Hirschman offre una chiave utile per leggere alcune caratteristiche emerse nel corso dell'esperimento, in particolare la minore concentrazione dei portafogli generati da Claude, ma non è sufficiente a spiegare, autonomamente, le differenze di performance tra i vari modelli. In conclusione, le allocazioni generate dai modelli risultano economicamente plausibili e spesso competitive con i benchmark di riferimento, pur non emergendo un vantaggio capace di superare, in termini di Sharpe Ratio, la diversificazione naive. La portata di queste evidenze deve comunque essere valutata alla luce dei limiti del disegno sperimentale, discussi nel paragrafo successivo.

5.2 Limiti del metodo

Ai limiti già discussi nel paragrafo 2.4 – rischio di data leakage e la natura pseudo out-of-sample del disegno sperimentale – se ne aggiungono altri nove, più direttamente collegati ai risultati presentati nel Capitolo 4.

In primo luogo, i risultati presentati nella Tabella 4.1 non sono stati sottoposti ad un test di significatività statistica, il quale avrebbe potuto confermare se le differenze di Sharpe Ratio tra i vari modelli potessero essere ritenute statisticamente rilevanti. Un simile test avrebbe potuto quantificare se le differenze osservate fossero riconducibili esclusivamente a rumore campionario piuttosto che ad un reale vantaggio statistico informativo. Data la natura esplorativa e non ripetuta del disegno sperimentale – un'unica generazione per modello e per profilo di rischio – i confronti presentati vanno letti come descrittivi e non come statisticamente robusti. Va precisato che anche un numero limitato di generazioni non sarebbe sufficiente a superare questo limite: affinché il test si possa ritenere statisticamente valido servirebbe un campione elevato di generazioni ripetute.

In secondo luogo, l'universo investibile chiuso a otto indici, necessario per garantire la comparabilità tra i modelli, non riflette i vincoli reali di un processo di asset allocation: vincoli legislativi, costi di transazione, implicazioni fiscali e vincoli di liquidità, del tutto assenti dal disegno sperimentale.

In terzo luogo, il prompt utilizzato, pur essendo congelato per evitare cherry-picking, presenta solamente una delle possibili formulazioni realizzabili, anche sul piano linguistico; dato anche che la maggior parte dei modelli viene addestrata in lingua inglese, altre lingue potrebbero produrre output differenti.

In quarto luogo, le spiegazioni fornite nel paragrafo 5.1 – il ruolo dei mercati emergenti, dell'oro e del grado di concentrazione misurato dall'HHI – restano interpretazioni plausibili, non relazioni causali dimostrate. Il disegno del

backtest non consente di scomporre il contributo di ciascun asset class al rendimento complessivo del portafoglio: gli effetti discussi sono economicamente coerenti con i dati osservati, ma senza un'analisi di attribuzione formale non è possibile calcolare il contributo delle singole componenti al portafoglio.

In quinto luogo, l'HHI misurato nel paragrafo 5.1 ci fornisce una misura della concentrazione dei vari portafogli, non una misura di rischio associata a ciascuno di essi: asset class nominalmente distinte, potrebbero avere correlazioni elevate e dunque non fornire una diminuzione del rischio, a fronte comunque di valori di HHI più bassi.

In sesto luogo, un'eventuale nuova generazione degli stessi portafogli non permetterebbe comunque di isolare la fonte di eventuali differenze di allocazione: sarebbe possibile valutare esclusivamente i portafogli senza comprendere i motivi fondamentali dei cambiamenti di allocazione dei modelli.

In settimo luogo, le motivazioni testuali fornite dai modelli AI in Appendice A non sono state sottoposte ad alcuna verifica di coerenza interna: riflettono le spiegazioni che il modello fornisce a posteriori a seguito della propria scelta e potrebbero non coincidere con le motivazioni ex ante che hanno condotto a tale scelta.

In ottavo luogo, i tre modelli sono stati interrogati tramite le rispettive interfacce commerciali, nelle configurazioni indicate in Appendice A, e non tramite altre modalità di accesso. I risultati riflettono quindi il comportamento di quelle configurazioni in un dato momento, e potrebbero non valere per versioni successive.

In nono luogo, i benchmark di profilo (Tabella 3.2) escludono corporate high yield e oro, mentre questi restano pienamente disponibili ai modelli generativi in fase di allocazione: un'asimmetria che concede ai portafogli AI un grado di

libertà non replicato nei benchmark di confronto. Va peraltro osservato che, trattandosi di profili di rischio definiti in termini puramente qualitativi ("aggressivo", "bilanciato", "conservativo"), non esiste una composizione di benchmark univocamente corretta: quella adottata in Tabella 3.2 è una convenzione ragionevole, non l'unica possibile. Un controllo esplorativo condotto sugli stessi dati, riportato in Appendice D, mostra che il vantaggio dei portafogli AI resta sostanzialmente stabile nel profilo aggressivo, ma si ridimensiona sensibilmente nei profili bilanciato e conservativo qualora il benchmark includa anch'esso high yield e oro: un'indicazione che parte dei risultati del Capitolo 4 possa dipendere dalla specifica composizione del benchmark di confronto, oltre che dall'allocazione proposta dai modelli.

5.3 Implicazioni per l'asset management

I risultati dell'esperimento eseguito suggeriscono alcune implicazioni per l'impiego dell'intelligenza artificiale generativa nei processi di asset allocation, da leggere anche alla luce dei limiti appena esposti.

La prima implicazione riguarda il ruolo che, allo stato attuale, questi strumenti sembrano più idonei a svolgere. La similarità dei portafogli creati dai vari modelli, all'interno dello stesso profilo di rischio, suggerisce che i modelli linguistici generativi tendono a riprodurre principi di asset allocation ben consolidati nella letteratura, piuttosto che proporre soluzioni originali ed innovative. Questo risulta essere coerente con quanto esposto nel paragrafo 2.3: l'intelligenza artificiale generativa risulta essere un generatore di proposte di allocazione di portafoglio rapido e leggibile nei suoi esiti - pesi espliciti per ogni indice – e coerente con il profilo di rischio richiesto, ma non può sostituire un giudizio professionale, né costituire una fonte dimostrata di vantaggio informativo.

La seconda implicazione riguarda la validazione. Il fatto che tutti i portafogli generati abbiano ottenuto un rendimento corretto per il rischio inferiore rispetto al portafoglio equal-weight ci suggerisce che l'output generato, per quanto leggibile, richiede comunque una verifica quali-quantitativa indipendente prima di poter essere impiegato operativamente. L'assenza di questa verifica potrebbe distinguere un'allocazione semplicemente convincente da una potenzialmente utile. Un indicatore sintetico come l'HHI potrebbe offrire un primo, rapido controllo sull'effettiva diversificazione dei portafogli generati, prima di qualunque validazione.

Un'ultima considerazione riguarda il modo in cui va letta la dissimilarità tra modelli rappresentata da Gemini nel profilo aggressivo. Poiché il mandato assegnato ai modelli è di tipo qualitativo - aggressivo, bilanciato e conservativo – senza soglie quantitative, le differenze all'interno dello stesso profilo di rischio ci permettono di comprendere i confini entro i quali un'etichetta viene interpretata. Tale dispersione, peraltro, non è uniforme tra i vari profili di rischio, segno che alcune etichette lasciano più margine di interpretazione soggettiva rispetto ad altre.

Infine, il profilo conservativo risulta essere quello in cui lo scarto rispetto al benchmark di profilo risulta più ampio (Tabella 4.1), ma anche quello in cui tale scarto si rivela più sensibile alla composizione del benchmark stesso, come discusso nel nono limite del paragrafo 5.2. Ne deriva, quindi, un'indicazione più cauta: i modelli sembrano riprodurre in modo affidabile principi generali di costruzione di allocazioni difensive - gestione della duration, diversificazione geografica, ricorso ad asset poco correlati – senza però che ciò si traduca in un vantaggio rispetto ad un benchmark difensivo che possa, a sua volta, impiegare oro e high yield.

5.4 Verifica di riproducibilità

Il sesto limite discusso nel paragrafo 5.2 riguarda l'impossibilità, con un'unica generazione per modello e profilo, di comprendere se un'eventuale differenza nei pesi assegnati rifletta un cambiamento reale nell'approccio del modello o semplicemente la variabilità del processo di generazione. Al fine di ridurre, non eliminare, questa incertezza i modelli di intelligenza artificiale sono stati nuovamente interrogati con il medesimo prompt congelato (Tabella 3.3), in due varianti: la formulazione originale in italiano e una sua traduzione in inglese, per verificare anche la sensibilità dei modelli ad eventuali modifiche della lingua utilizzata. Si precisa che non è stato possibile interrogare lo stesso modello di Claude (Fable 5 Max), in quanto quest'ultimo, al momento della nuova generazione risulta di impossibile utilizzo senza l'acquisto di crediti di utilizzo; per questo motivo è stato utilizzato Sonnet 5, un modello della stessa famiglia (Appendice A.4.2-A.5.2). Successivamente i nuovi pesi sono stati sottoposti allo stesso backtest validato e sono state calcolate le stesse metriche, già precedentemente utilizzate (Appendice B.6).

Le Tabelle 5.1 e 5.2 riportano i pesi ottenuti nelle due rigenerazioni.

Tabella 5.1: Pesi dei portafogli rigenerati - prompt in italiano (%)

Portafoglio	S&P500	EAFE	EM	T10Y	T2Y	IG	HY	Oro
ChatGPT - Aggressivo	38%	22%	15%	5%	—	5%	10%	5%
Claude - Aggressivo	40%	20%	15%	5%	—	5%	10%	5%
Gemini - Aggressivo	50%	25%	15%	5%	—	—	—	5%
ChatGPT - Bilanciato	25%	15%	10%	10%	8%	18%	6%	8%
Claude - Bilanciato	25%	15%	10%	10%	10%	20%	5%	5%
Gemini - Bilanciato	35%	15%	5%	15%	—	20%	5%	5%
ChatGPT - Conservativo	10%	7%	3%	15%	25%	30%	2%	8%

(Tabella 5.1 segue in pagina successiva)

Tabella 5.1 (prosegue dalla pagina precedente)

Claude - Conservativo	10%	5%	—	20%	25%	25%	5%	10%
Gemini - Conservativo	10%	10%	—	15%	30%	25%	—	10%

Fonte: elaborazione propria su Appendice A.4.

Tabella 5.2: Pesi dei portafogli rigenerati - prompt in inglese (%)

Portafoglio	S&P500	EAFE	EM	T10Y	T2Y	IG	HY	Oro
ChatGPT - Aggressivo	42%	20%	18%	3%	2%	5%	7%	3%
Claude - Aggressivo	40%	20%	20%	5%	—	5%	5%	5%
Gemini - Aggressivo	45%	20%	15%	5%	—	—	10%	5%
ChatGPT - Bilanciato	30%	13%	7%	12%	8%	18%	6%	6%
Claude - Bilanciato	25%	10%	5%	15%	10%	20%	5%	10%
Gemini - Bilanciato	35%	15%	5%	15%	5%	15%	5%	5%
ChatGPT - Conservativo	12%	5%	3%	12%	30%	28%	3%	7%
Claude - Conservativo	10%	5%	—	20%	25%	25%	5%	10%
Gemini - Conservativo	15%	5%	—	20%	30%	20%	5%	5%

Fonte: elaborazione propria su Appendice A.5.

In media, circa l'11-12 per cento del capitale viene riallocato tra un asset e l'altro rispetto ai pesi originali – un ordine di grandezza che lascia le allocazioni comparabili rispetto a quelle di partenza, confrontabili con la Tabella 3.2 e con i pesi originali riportati in Appendice A.1-A.3. Il confronto diretto tra le due rigenerazioni – quelle in inglese e in italiano – mostra una distanza reciproca del 7,7 per cento di capitale riallocato, inferiore rispetto alla distanza con i pesi originali; ciò suggerisce quindi che il fattore dominante dei cambiamenti risulta essere maggiormente il tempo trascorso tra le rigenerazioni, piuttosto che la lingua del prompt. La Tabella 5.3 confronta alcune metriche - l'indice di Herfindahl-Hirschman, lo Sharpe Ratio e il CAGR - calcolate per i portafogli originali e per le rigenerazioni.

Tabella 5.3: Metriche di sintesi - confronto tra pesi originali e rigenerazioni

Portafoglio	HHI			Sharpe Ratio			CAGR		
	Orig.	IT	EN	Orig.	IT	EN	Orig.	IT	EN
ChatGPT - Aggressivo	0,290	0,233	0,258	0,793	0,800	0,779	11,65%	10,78%	10,96%
Claude - Aggressivo	0,280	0,240	0,250	0,806	0,806	0,792	11,54%	10,91%	11,03%
Gemini - Aggressivo	0,345	0,340	0,280	0,747	0,786	0,806	11,76%	12,05%	11,54%
ChatGPT - Bilanciato	0,190	0,154	0,172	0,793	0,844	0,830	8,48%	8,52%	8,54%
Claude - Bilanciato	0,175	0,160	0,160	0,807	0,801	0,871	8,70%	8,06%	7,99%
Gemini - Bilanciato	0,230	0,215	0,200	0,782	0,809	0,814	8,12%	9,00%	8,99%
ChatGPT - Conservativo	0,230	0,198	0,206	0,594	0,673	0,712	4,93%	5,15%	5,20%
Claude - Conservativo	0,205	0,190	0,190	0,676	0,669	0,669	5,19%	5,13%	5,13%
Gemini - Conservativo	0,250	0,205	0,200	0,669	0,697	0,676	4,97%	5,29%	5,14%

Fonte: elaborazione propria su Appendice A e B

Due pattern discussi in dettaglio nei paragrafi precedenti non trovano conferma nelle rigenerazioni. Il portafoglio Gemini Aggressivo, il quale risultava l'unico portafoglio del profilo nella generazione originale a non superare lo Sharpe Ratio del benchmark di riferimento, lo supera ora in entrambe le rigenerazioni; resta però, specialmente nella versione italiana, il portafoglio con l'allocazione azionaria più elevata ed il portafoglio più concentrato tra i tre modelli (90 per cento di allocazione azionaria, contro il 75 per cento degli altri due modelli ed un'HHI di 0,340, sostanzialmente invariato rispetto all'originale), segno che l'esito sullo Sharpe non si ripete, ma la tendenza sottostante si attenua solo in parte. Il primato di Claude nella generazione originale, il quale otteneva lo Sharpe più elevato in tutti e tre i profili non si ripete

in nessuna delle due generazioni: Claude prevale in un solo profilo su tre in ciascuna rigenerazione, e non lo stesso nelle due varianti linguistiche.

Ai modelli, è stato inoltre chiesto, a posteriori, di fornire delle motivazioni a supporto della loro allocazione. Le risposte ottenute forniscono delle stime quantitative aggiuntive (volatilità e correlazioni indicative) a sostegno delle scelte, ma rimangono comunque delle spiegazioni costruite ex-post: la loro coerenza non garantisce che riflettano fedelmente il processo che ha effettivamente generato l'allocazione, coerentemente con il settimo limite già espresso nel paragrafo 5.2.

Questi risultati, di conseguenza, non consentono conclusioni definitive - una singola rigenerazione non risolve il limite descritto nel paragrafo 5.2, che richiederebbe un campione decisamente più ampio per poter ritenere il test statisticamente valido - ma offrono comunque un riscontro concreto alla cautela già espressa nel capitolo: alcuni pattern, precedentemente discussi, si sono rivelati più legati alla singola generazione piuttosto che a una caratteristica stabile dei modelli.

CONCLUSIONI

Questo lavoro ha cercato di rispondere ad una domanda circoscritta: l'intelligenza artificiale è in grado di generare proposte di allocazione di portafoglio capaci di ottenere performance comparabili, o superiori, a strategie tradizionali? La risposta che emerge dall'analisi non può essere riducibile ad un giudizio univoco.

I portafogli generati dai tre modelli di intelligenza artificiale generativa presentano strutture di allocazione simili tra loro e coerenti con i principi consolidati di asset allocation, e in due profili su tre – il bilanciato e il conservativo – tutti e tre i modelli ottengono, rispetto ai benchmark di riferimento definiti in Tabella 3.2, rendimenti corretti per il rischio superiori; nel profilo aggressivo, l'eccezione è rappresentata da Gemini, come discusso nel Capitolo 5. Come discusso nel paragrafo 5.2, questo vantaggio non è però ugualmente robusto nei tre profili: nell'aggressivo regge anche a fronte di benchmark alternativi più esposti a high yield e oro, mentre nel bilanciato e, soprattutto, nel conservativo si ridimensiona sensibilmente qualora il benchmark venga costruito con accesso alle stesse otto asset class disponibili ai modelli (Appendice D). Nessuno dei nove portafogli generati, in ogni caso, riesce a superare la semplice diversificazione equiponderata (1/N) - risultato coerente con la letteratura sull'ottimizzazione di portafoglio - e che ridimensiona ulteriormente l'idea di un vantaggio informativo da parte dei modelli di intelligenza artificiale.

La discussione dei risultati (Capitolo 5) ha permesso di individuare alcuni meccanismi economici plausibili dietro queste differenze – l'esposizione ai mercati emergenti, l'inclusione dell'oro, il grado di concentrazione dei portafogli misurato tramite l'indice di Herfindahl-Hirschman – senza però riuscire a dimostrarne la natura causale, data la struttura sperimentale a generazione unica. La verifica di riproducibilità, condotta nel paragrafo 5.4, rigenerando i portafogli con lo stesso prompt in due lingue diverse ha, solo in parte, confermato questa cautela: alcuni pattern osservati nell'esperimento originale – in particolare l'eccezione di Gemini nel profilo aggressivo e il primato, in tutti e tre i profili, di Claude – non si sono ripetuti nelle rigenerazioni, a conferma che vanno letti come caratteristiche specifiche di quella generazione, piuttosto che come proprietà stabili del modello.

Il lavoro presenta limiti che sono stati esplicitamente dichiarati nel corso della trattazione: l'assenza di un test di significatività statistica, l'universo investibile chiuso ad otto indici, la formulazione unica del prompt, la sensibilità dei risultati alla composizione dei benchmark di confronto, l'impossibilità di isolare il contributo causale delle singole scelte di allocazione, e la natura pseudo out-of-sample dell'esperimento, che non consente di escludere con assoluta certezza una possibile influenza dei dati di addestramento dei modelli sulle decisioni di allocazione proposte.

Nel complesso, i risultati dell'esperimento suggeriscono che l'intelligenza artificiale generativa si comporti più come un generatore di proposte di allocazione coerenti con principi noti – utile per la sua rapidità, ma da sottoporre a verifiche quantitative indipendenti – che come una fonte autonoma di vantaggio informativo. Il profilo aggressivo è quello in cui questo vantaggio, per quanto descrittivo e non sottoposto a test di significatività statistica, appare più robusto; nei profili bilanciato e conservativo va invece considerato con maggiore

cautela, alla luce della sensibilità dei risultati alla composizione del benchmark discussa nel paragrafo 5.2.

Resta aperta, per ricerche future, la possibilità di approfondire questi risultati lungo diverse direttrici: la ripetizione dell'esperimento su un campione ampio di generazioni, condizione necessaria per un test statisticamente valido sulla stabilità nel tempo dell'approccio allocativo dei modelli di intelligenza artificiale; l'estensione dell'analisi a periodi diversi o più lunghi, per verificare la tenuta dei risultati al variare dei diversi contesti di mercato; la sperimentazione con formulazioni alternative del prompt, per valutare quanto le allocazioni possano essere sensibili a variazioni nel modo in cui la richiesta viene posta al modello di intelligenza artificiale; e l'estensione sistematica della verifica avviata in Appendice D a un insieme più ampio di composizioni dei benchmark di confronto, per accertare in quale misura le conclusioni raggiunte dipendano dalla specifica costruzione adottata in questo lavoro.

BIBLIOGRAFIA

Ang, Andrew. *Asset Management: A Systematic Approach to Factor Investing*. Oxford University Press, 2014.

Artzner, Philippe, Freddy Delbaen, Jean-Marc Eber, and David Heath. "Coherent Measures of Risk." *Mathematical Finance* 9, no. 3 (1999): 203–28. https://doi.org/10.1111/1467-9965.00068.

Bailey, David H., Jonathan M. Borwein, Marcos López De Prado, and Qiji Jim Zhu. "Pseudo-Mathematics and Financial Charlatanism: The Effects of Backtest Overfitting on Out-of-Sample Performance." *Notices of the American Mathematical Society* 61, no. 5 (2014): 458–71. https://doi.org/10.1090/noti1105.

Banz, Rolf W., and William J. Breen. "Sample-Dependent Results Using Accounting and Market Data: Some Evidence." *The Journal of Finance* 41, no. 4 (1986): 779–93. https://doi.org/10.1111/j.1540-6261.1986.tb04548.x.

Bartram, Söhnke M., Jürgen Branke, and Mehrshad Motahari. *Artificial Intelligence in Asset Management*. Literature Review. CFA Institute Research Foundation, 2020. https://doi.org/10.2139/ssrn.3692805.

Bodie, Zvi, Alex Kane, and Alan J. Marcus. *Investments*. 13th ed. McGraw-Hill Education, 2021.

Brinson, Gary P., L. Randolph Hood, and Gilbert L. Beebower. "Determinants of Portfolio Performance." *Financial Analysts Journal* 42, no. 4 (1986): 39–44. https://doi.org/10.2469/faj.v42.n4.39.

Cao, Longbing. "AI in Finance: Challenges, Techniques, and Opportunities." *ACM Computing Surveys* 55, no. 3 (2022): 1–38. https://doi.org/10.1145/3502289.

Carhart, Mark M. “On Persistence in Mutual Fund Performance.” *The Journal of Finance* 52, no. 1 (1997): 57–82. https://doi.org/10.1111/j.1540-6261.1997.tb03808.x.

DeMiguel, Victor, Lorenzo Garlappi, and Raman Uppal. “Optimal Versus Naive Diversification: How Inefficient Is the 1/ *N* Portfolio Strategy?” *Review of Financial Studies* 22, no. 5 (2009): 1915–53. https://doi.org/10.1093/rfs/hhm075.

Doshi-Velez, Finale, and Been Kim. “Towards A Rigorous Science of Interpretable Machine Learning.” arXiv:1702.08608. Preprint, arXiv, March 2, 2017. https://doi.org/10.48550/arXiv.1702.08608.

European Securities and Markets Authority. *AI Adoption and Trends in Securities Markets: EU Evidence*. ESMA TRV Risk Analysis – Financial Innovation Nos. ESMA50-481369926–30599. European Securities and Markets Authority, 2026.

Fama, Eugene F., and Kenneth R. French. “The Cross-Section of Expected Stock Returns.” *The Journal of Finance* 47, no. 2 (1992): 427–65. https://doi.org/10.1111/j.1540-6261.1992.tb04398.x.

Financial Stability Board. *Artificial Intelligence and Machine Learning in Financial Services: Market Developments and Financial Stability Implications*. Financial Stability Board, 2017.

Glasserman, Paul, and Caden Lin. “Assessing Look-Ahead Bias in Stock Return Predictions Generated by GPT Sentiment Analysis.” *The Journal of Financial Data Science* 6, no. 1 (2024): 25–42. https://doi.org/10.3905/jfds.2023.1.143.

Grinold, Richard C., and Ronald N. Kahn. *Active Portfolio Management: A Quantitative Approach for Producing Superior Returns and Controlling Risk*. 2nd ed. McGraw-Hill, 2000.

Gu, Shihao, Bryan Kelly, and Dacheng Xiu. “Empirical Asset Pricing via Machine Learning.” *The Review of Financial Studies* 33, no. 5 (2020): 2223–73. https://doi.org/10.1093/rfs/hhaa009.

Harvey, Campbell R., Yan Liu, and Heqing Zhu. “… and the Cross-Section of Expected Returns.” *Review of Financial Studies* 29, no. 1 (2016): 5–68. https://doi.org/10.1093/rfs/hhv059.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. Springer, 2009.

Hirschman, Albert O. "The Paternity of an Index." *The American Economic Review* 54 (1964): 761–62. https://doi.org/10.2307/1818582.

Ibbotson, Roger G., and Paul D. Kaplan. "Does Asset Allocation Policy Explain 40, 90, or 100 Percent of Performance?" *Financial Analysts Journal* 56, no. 1 (2000): 26–33. https://doi.org/10.2469/faj.v56.n1.2327.

International Monetary Fund. "Advances in Artificial Intelligence: Implications for Capital Market Activities." In *Global Financial Stability Report, October 2024: Steadying the Course: Uncertainty, Artificial Intelligence, and Financial Stability*. International Monetary Fund, 2024. https://www.imf.org/-/media/files/publications/gfsr/2024/october/english/textrevised.pdf.

Jegadeesh, Narasimhan, and Sheridan Titman. "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency." *The Journal of Finance* 48, no. 1 (1993): 65–91. https://doi.org/10.1111/j.1540-6261.1993.tb04702.x.

Kaufman, Shachar, Saharon Rosset, Claudia Perlich, and Ori Stitelman. "Leakage in Data Mining: Formulation, Detection, and Avoidance." *ACM Transactions on Knowledge Discovery from Data* 6, no. 4 (2012): 1–21. https://doi.org/10.1145/2382577.2382579.

Kelly, Bryan, and Dacheng Xiu. "Financial Machine Learning." *Foundations and Trends in Finance* 13, nos. 3–4 (2023): 205–363. https://doi.org/10.1561/500000064.

Ko, Hyungjin, and Jaewook Lee. "Can ChatGPT Improve Investment Decisions? From a Portfolio Management Perspective." *Finance Research Letters* 64 (June 2024): 105433. https://doi.org/10.1016/j.frl.2024.105433.

Lo, Andrew W., and A. Craig MacKinlay. "Data-Snooping Biases in Tests of Financial Asset Pricing Models." *Review of Financial Studies* 3, no. 3 (1990): 431–67. https://doi.org/10.1093/rfs/3.3.431.

Lopez-Lira, Alejandro, and Yuehua Tang. “Can ChatGPT Forecast Stock Price Movements? Return Predictability and Large Language Models.” *Journal of Financial Economics*, forthcoming. https://doi.org/10.2139/ssrn.4412788.

Lorenz, Philippe, Karine Perset, and Jamie Berryhill. *Initial Policy Considerations for Generative Artificial Intelligence*. No. 1. OECD Artificial Intelligence Papers. OECD Publishing, 2023. https://doi.org/10.1787/fae2d1e6-en.

Lundberg, Scott, and Su-In Lee. “A Unified Approach to Interpreting Model Predictions.” *Advances in Neural Information Processing Systems 30 (NIPS 2017)* (Long Beach, CA), 2017, 4765–74. https://doi.org/10.48550/arXiv.1705.07874.

Markowitz, Harry. “Portfolio Selection.” *The Journal of Finance* 7, no. 1 (1952): 77–91. https://doi.org/10.1111/j.1540-6261.1952.tb01525.x.

Mullainathan, Sendhil, and Jann Spiess. “Machine Learning: An Applied Econometric Approach.” *Journal of Economic Perspectives* 31, no. 2 (2017): 87–106. https://doi.org/10.1257/jep.31.2.87.

Pelster, Matthias, and Joel Val. “Can ChatGPT Assist in Picking Stocks?” *Finance Research Letters* 59 (January 2024): 104786. https://doi.org/10.1016/j.frl.2023.104786.

Ribeiro, Marco Tulio, Sameer Singh, and Carlos Guestrin. “‘Why Should I Trust You?’: Explaining the Predictions of Any Classifier.” *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, August 13, 2016, 1135–44. https://doi.org/10.1145/2939672.2939778.

Romanko, Oleksandr, Akhilesh Narayan, and Roy H. Kwon. “ChatGPT-Based Investment Portfolio Selection.” *Operations Research Forum* 4, no. 4 (2023): 91. https://doi.org/10.1007/s43069-023-00277-6.

Rudin, Cynthia. “Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead.” *Nature Machine Intelligence* 1, no. 5 (2019): 206–15. https://doi.org/10.1038/s42256-019-0048-x.

Sharpe, William F. “Mutual Fund Performance.” *The Journal of Business* 39, no. S1 (1966): 119. https://doi.org/10.1086/294846.

Singla, Alex, Alexander Sukharevsky, Bryce Hall, Lareina Yee, Michael Chui, and Tara Balakrishnan. *The State of AI in 2025: Agents, Innovation, and Transformation*. McKinsey & Company, 2025. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai#/.

Sullivan, Ryan, Allan Timmermann, and Halbert White. “Data-Snooping, Technical Trading Rule Performance, and the Bootstrap.” *The Journal of Finance* 54, no. 5 (1999): 1647–91. https://doi.org/10.1111/0022-1082.00163.

Young, Terry W. “Calmar Ratio: A Smoother Tool.” *Futures Magazine*, 1991.

SITOGRAFIA

Charles Schwab. “A Guide to Risk Profiles.” Schwab Brokerage, May 29, 2026. https://www.schwab.com/learn/story/guide-to-risk-profiles.

APPENDICE A - OUTPUT INTEGRALI DEI MODELLI

A.1 ChatGPT (OpenAI)

Modello: ChatGPT 5.6 Sol Alto - Data: 20/07/2026 - Output integrale, riportato senza modifiche.

1. PORTAFOGLIO AGGRESSIVO

Indice	**Peso %**
S&P 500 Total Return	45
MSCI EAFE Net Total Return	20
MSCI Emerging Markets Net Total Return	20
US Treasury 10Y Total Return	5
US Treasury 2Y Total Return	0
US Investment Grade Total Return	0
US High Yield Corporate Total Return	5
Gold	5

Motivazione economica (≤150 parole)

L'allocazione privilegia nettamente l'azionario (85%), coerentemente con l'obiettivo di massimizzare la crescita del capitale nel lungo periodo. L'S&P 500 rappresenta il nucleo principale grazie all'elevata qualità e dimensione delle imprese statunitensi, mentre MSCI EAFE amplia la diversificazione nei mercati sviluppati e MSCI Emerging Markets aggiunge esposizione a economie con

maggiore potenziale di crescita ma anche maggiore volatilità. Una limitata quota in Treasury decennali introduce una componente difensiva nei periodi di debolezza dei mercati azionari. L'esposizione all'High Yield incrementa il potenziale di rendimento mantenendo il reddito fisso orientato alla crescita, mentre una piccola posizione in oro contribuisce alla diversificazione complessiva.

Principale punto di forza: elevata esposizione ai motori strutturali della crescita economica globale.

Principale rischio: forte volatilità e possibilità di drawdown significativi dovuti all'elevata componente azionaria.

---

2. PORTAFOGLIO BILANCIATO

Indice	**Peso %**
S&P 500 Total Return	30
MSCI EAFE Net Total Return	15
MSCI Emerging Markets Net Total Return	10
US Treasury 10Y Total Return	15
US Treasury 2Y Total Return	5
US Investment Grade Total Return	20
US High Yield Corporate Total Return	0
Gold	5

Motivazione economica (≤150 parole)

Il portafoglio ricerca un equilibrio tra crescita e contenimento del rischio attraverso una ripartizione pressoché bilanciata tra azioni (55%) e attività difensive (45%). L'esposizione azionaria è ampiamente diversificata tra Stati Uniti, mercati sviluppati internazionali ed emergenti. La componente

obbligazionaria combina Treasury a breve e lunga scadenza con obbligazioni investment grade, ottenendo una maggiore diversificazione tra rischio di tasso e rischio di credito. L'oro rappresenta una componente di diversificazione aggiuntiva, utile per ridurre la dipendenza del portafoglio dalle sole attività finanziarie tradizionali. L'assenza di High Yield mantiene il profilo di rischio coerente con un obiettivo di equilibrio di lungo periodo.

Principale punto di forza: ampia diversificazione tra aree geografiche e principali classi di attivo.

Principale rischio: rendimenti potenzialmente inferiori a un portafoglio fortemente azionario durante lunghi mercati rialzisti.

---

3. PORTAFOGLIO CONSERVATIVO

Indice	**Peso %**
S&P 500 Total Return	15
MSCI EAFE Net Total Return	5
MSCI Emerging Markets Net Total Return	0
US Treasury 10Y Total Return	20
US Treasury 2Y Total Return	20
US Investment Grade Total Return	35
US High Yield Corporate Total Return	0
Gold	5

Motivazione economica (≤150 parole)

L'allocazione privilegia la preservazione del capitale attraverso una prevalenza di strumenti obbligazionari di elevata qualità (75%). La componente investment grade costituisce il nucleo del portafoglio grazie alla combinazione di titoli governativi e societari con merito creditizio elevato. La presenza congiunta di

Treasury a breve e lunga scadenza migliora la diversificazione rispetto ai soli rischi di tasso. L'esposizione azionaria è limitata e concentrata sui mercati sviluppati, mantenendo comunque una partecipazione al potenziale di crescita di lungo periodo. L'assenza di mercati emergenti e High Yield riduce il rischio di credito e la volatilità complessiva. Una modesta quota in oro contribuisce ulteriormente alla diversificazione.

Principale punto di forza: elevata stabilità grazie alla predominanza di obbligazioni di alta qualità.

Principale rischio: minore capacità di crescita del capitale nel lungo periodo rispetto ai portafogli con maggiore esposizione azionaria.

A.2 Claude (Anthropic)

Modello: Claude Fable 5 Max - Data: 20/07/2026 - Output integrale, riportato senza modifiche.

Portafoglio 1 - AGGRESSIVO

Indice	**Peso %**
S&P 500 Total Return	45
MSCI EAFE Net Total Return	20
MSCI Emerging Markets Net Total Return	15
US Treasury 10Y Total Return	5
US Treasury 2Y Total Return	0
US Investment Grade Total Return	0
US High Yield Corporate Total Return	10
Gold	5

Indice	Peso %
Totale	**100**

Motivazione. L'allocazione destina il 90% ad attività di rischio: 80% azionario e 10% high yield. Il peso prevalente dell'S&P 500 riflette il ruolo core del mercato statunitense per profondità, liquidità e qualità degli utili; EAFE ed emergenti ampliano la diversificazione geografica e catturano il potenziale di crescita extra-USA, accettandone volatilità e rischio valutario. L'high yield aggiunge carry con comportamento affine all'azionario ma volatilità inferiore. Il residuo 10% (Treasury decennale e oro) non ha finalità di rendimento: introduce attività storicamente poco correlate all'azionario, utili ad attenuare marginalmente i drawdown e ad alimentare i ribilanciamenti trimestrali, che riacquistano sistematicamente le classi più penalizzate. La struttura massimizza l'esposizione al premio al rischio azionario di lungo periodo, coerente con un orizzonte esteso e piena tolleranza alle oscillazioni.

Principale punto di forza: piena cattura del premio al rischio azionario globale, diversificata su tre aree geografiche, con ribilanciamento disciplinato che sfrutta la volatilità.

Principale rischio: drawdown potenzialmente severi; nelle crisi sistemiche le correlazioni tra azioni e high yield tendono a salire, riducendo la diversificazione proprio quando più necessaria.

Portafoglio 2 - BILANCIATO

Indice	Peso %
S&P 500 Total Return	30
MSCI EAFE Net Total Return	15
MSCI Emerging Markets Net Total Return	10
US Treasury 10Y Total Return	15

Indice	Peso %
US Treasury 2Y Total Return	5
US Investment Grade Total Return	15
US High Yield Corporate Total Return	5
Gold	5
Totale	**100**

Motivazione. La struttura 55/40/5 impiega tutte le otto classi. Il 55% azionario — nucleo USA, sviluppati internazionali ed emergenti — resta il motore di crescita; il 40% obbligazionario lo bilancia articolandosi su duration e credito: il decennale è storicamente la copertura più efficace nelle fasi di avversione al rischio, l'aggregato investment grade unisce reddito e qualità creditizia, il biennale stabilizza riducendo la duration complessiva, l'high yield in dose contenuta incrementa il rendimento cedolare. Il 5% di oro aggiunge un diversificatore reale, poco correlato sia alle azioni sia alle obbligazioni. Il ribilanciamento trimestrale mantiene costante il profilo di rischio e sfrutta la dispersione dei rendimenti tra le classi. Nessuna singola fonte di rischio domina: l'obiettivo è un rapporto rendimento/rischio più efficiente di quello delle singole componenti.

Principale punto di forza: diversificazione su tutte le fonti di rendimento disponibili (premio azionario, duration, credito, oro), che rende il portafoglio resiliente in una pluralità di scenari.

Principale rischio: uno scenario di rialzo simultaneo di inflazione e tassi può penalizzare contemporaneamente azioni e obbligazioni, indebolendo la funzione di copertura della componente a reddito fisso.

Portafoglio 3 - CONSERVATIVO

Indice	**Peso %**
S&P 500 Total Return	15
MSCI EAFE Net Total Return	5
MSCI Emerging Markets Net Total Return	0
US Treasury 10Y Total Return	15
US Treasury 2Y Total Return	25
US Investment Grade Total Return	30
US High Yield Corporate Total Return	5
Gold	5
Totale	**100**

Motivazione. Il 75% è investito in obbligazioni prevalentemente di alta qualità: il biennale (25%) minimizza volatilità e rischio di tasso, l'aggregato investment grade (30%) fornisce reddito con merito creditizio elevato, il decennale (15%) offre protezione nelle fasi di flight-to-quality. La quota azionaria contenuta (20%, limitata ai mercati sviluppati, con emergenti esclusi per la loro volatilità) preserva un motore di crescita reale sull'orizzonte lungo e contrasta l'erosione inflazionistica, con contributo limitato al rischio complessivo. Il 5% di high yield integra marginalmente il rendimento cedolare; il 5% di oro protegge in scenari di stress finanziario o di inflazione inattesa. La duration complessiva moderata e la prevalenza di attività a bassa volatilità contengono i drawdown attesi, coerentemente con l'obiettivo primario di preservazione del capitale.

Principale punto di forza: bassa volatilità attesa e contenimento dei drawdown, grazie alla duration moderata e all'elevata qualità creditizia della componente obbligazionaria.

Principale rischio: rendimento atteso modesto, dato il livello storicamente basso dei tassi a fine 2015, con possibilità che nel lungo periodo l'inflazione eroda il rendimento reale del portafoglio.

A.3 Gemini (Google)

Modello: Gemini 3.1 Pro - Data: 20/07/2026 - Output integrale, riportato senza modifiche.

1. PORTAFOGLIO AGGRESSIVO

Indice	Peso percentuale (%)
S&P 500 Total Return	50
MSCI EAFE Net Total Return	25
MSCI Emerging Markets Net Total Return	15
US Treasury 10Y Total Return	0
US Treasury 2Y Total Return	0
US Investment Grade Total Return	0
US High Yield Corporate Total Return	10
Gold	0

Motivazione economica

Il portafoglio "Aggressivo" è strutturato per massimizzare il premio al rischio azionario di lungo periodo. Un'esposizione del 90% in azioni intercetta la crescita economica globale, bilanciando il peso dominante delle large cap statunitensi (S&P 500) con i mercati sviluppati (MSCI EAFE) ed emergenti (MSCI EM), questi ultimi portatori di un maggiore potenziale di rendimento legato a dinamiche demografiche ed economiche favorevoli. Il restante 10% è allocato nell'US High Yield, strumento obbligazionario che storicamente si comporta in modo simile all'azionario durante i mercati rialzisti, fornendo un flusso cedolare elevato che incrementa il total return. L'assenza di bond governativi difensivi (Treasury) è coerente con l'accettazione di una volatilità estrema per ottimizzare la crescita.

Principale punto di forza dell'allocazione: Elevata capacità di rivalutazione del capitale nel lungo termine grazie all'esposizione predominante ai mercati azionari globali.

Principale rischio dell'allocazione: Sensibilità estrema alle contrazioni economiche e ai crolli azionari, con probabilità di subire drawdown profondi e prolungati a causa della totale assenza di asset decorrelati o difensivi.

2. PORTAFOGLIO BILANCIATO

Indice	**Peso percentuale (%)**
S&P 500 Total Return	30
MSCI EAFE Net Total Return	15
MSCI Emerging Markets Net Total Return	5
US Treasury 10Y Total Return	15

Indice	Peso percentuale (%)
US Treasury 2Y Total Return	0
US Investment Grade Total Return	30
US High Yield Corporate Total Return	0
Gold	5

Motivazione economica

L'asset allocation "Bilanciata" mira a un'efficiente ottimizzazione del rapporto rischio/rendimento. Il 50% azionario garantisce la necessaria partecipazione all'espansione economica globale, diversificando geograficamente le fonti di crescita. Il comparto obbligazionario (45%) è focalizzato esclusivamente su emissioni di alta qualità (US Investment Grade e Treasury a 10 anni). Queste asset class esprimono storicamente una correlazione negativa rispetto all'azionario durante gli shock di mercato (flight-to-quality), attenuando la volatilità complessiva senza sacrificare in modo eccessivo i rendimenti. Infine, l'inclusione strutturale dell'oro (5%) funge da elemento diversificatore, aggiungendo un livello di protezione contro l'inflazione imprevista o fasi di severa instabilità finanziaria che potrebbero colpire simultaneamente azioni e bond societari.

Principale punto di forza dell'allocazione: Ottima stabilità e resilienza durante cicli economici misti, ottenuta bilanciando la crescita strutturale azionaria con la protezione dei bond di alta qualità.

Principale rischio dell'allocazione: Rischio di correlazione positiva tra azioni e obbligazioni, con perdite simultanee in entrambi i comparti in caso di rapidi e imprevisti rialzi dei tassi di interesse reali.

3. PORTAFOGLIO CONSERVATIVO

Indice	Peso percentuale (%)
S&P 500 Total Return	15
MSCI EAFE Net Total Return	5
MSCI Emerging Markets Net Total Return	0
US Treasury 10Y Total Return	10
US Treasury 2Y Total Return	35
US Investment Grade Total Return	30
US High Yield Corporate Total Return	0
Gold	5

Motivazione economica

Il portafoglio "Conservativo" ha come mandato primario la difesa del capitale e la minimizzazione del rischio di perdita. Ben l'80% dell'esposizione è allocato tra strumenti obbligazionari e oro. Il pilastro centrale è il Treasury a 2 anni (35%), che minimizza il rischio di duration fornendo elevata stabilità nominale. L'US Investment Grade (30%) offre un reddito prevedibile con un merito creditizio solido, mentre il Treasury a 10 anni (10%) agisce come copertura per scenari deflattivi. La componente azionaria (20%) è limitata ai soli mercati sviluppati per evitare l'alta volatilità e il rischio valutario degli emergenti, ed è

calibrata unicamente per contrastare l'erosione del potere d'acquisto nel tempo. L'oro (5%) completa il setup difensivo complessivo.

Principale punto di forza dell'allocazione: Altissima probabilità di preservazione del capitale nominale e contenimento drastico dei drawdown, grazie alla marcata concentrazione sulla parte a breve della curva governativa.

Principale rischio dell'allocazione: Elevato rischio di shortfall rispetto all'inflazione, con la forte possibilità che i rendimenti reali del portafoglio nel lungo periodo risultino marginali o nulli.

A.4 - Rigenerazione (prompt in italiano)

A.4.1 ChatGPT (OpenAI)

Modello: ChatGPT 5.6 Sol Alto - Data: 07/08/2026 - Rigenerazione con lo stesso prompt congelato (Tabella 3.3), lingua italiana - Output integrale, riportato senza modifiche.

1. AGGRESSIVO

Indice	**Peso %**
S&P 500 Total Return	38%
MSCI EAFE Net Total Return	22%
MSCI Emerging Markets Net Total Return	15%
US Treasury 10Y Total Return	5%
US Treasury 2Y Total Return	0%
US Investment Grade Total Return	5%
US High Yield Corporate Total Return	10%
Gold	5%

Indice	Peso %
Totale	100%

Motivazione economica

L'allocazione concentra il 75% nell'azionario globale, principale motore della crescita reale del capitale nel lungo periodo. L'S&P 500 costituisce il nucleo centrale, mentre EAFE ed emergenti riducono la dipendenza dall'economia statunitense e ampliano l'esposizione a differenti cicli economici e valutari. Il 10% in high yield aumenta il potenziale di rendimento, pur introducendo un rischio di credito assimilabile in parte a quello azionario. Treasury decennali e obbligazioni investment grade forniscono una riserva difensiva limitata, utile durante alcune fasi di avversione al rischio. L'oro aggiunge una fonte di diversificazione non direttamente collegata agli utili societari. Il ribilanciamento trimestrale mantiene stabile l'elevata esposizione al rischio e impone una disciplina sistematica di riallocazione.

Principale punto di forza: elevato potenziale di crescita, con ampia diversificazione geografica dell'esposizione azionaria.

Principale rischio: drawdown molto pronunciati in fasi di contrazione globale, aggravati dalla correlazione tra azioni, mercati emergenti e high yield.

2. BILANCIATO

Indice	Peso %
S&P 500 Total Return	25%
MSCI EAFE Net Total Return	15%
MSCI Emerging Markets Net Total Return	10%
US Treasury 10Y Total Return	10%
US Treasury 2Y Total Return	8%
US Investment Grade Total Return	18%

Indice	Peso %
US High Yield Corporate Total Return	6%
Gold	8%
Totale	**100%**

Motivazione economica

Il portafoglio assegna il 50% all'azionario globale, sufficiente a sostenere la crescita del capitale senza rendere dominante il rischio azionario. La componente statunitense rimane prevalente, affiancata da mercati sviluppati internazionali ed emergenti per diversificare aree economiche, valute e fonti di rendimento. Il 36% in Treasury e investment grade costituisce il principale stabilizzatore: i titoli biennali limitano la sensibilità ai tassi, mentre Treasury decennali e obbligazioni investment grade offrono reddito e potenziale protezione nelle fasi di debolezza economica. L'high yield è mantenuto al 6% per incrementare moderatamente il rendimento atteso senza assumere un'eccessiva esposizione al rischio di credito. L'8% in oro amplia la diversificazione rispetto ai rischi azionari, creditizi e monetari.

Principale punto di forza: equilibrio tra partecipazione alla crescita globale e presenza significativa di attività difensive.

Principale rischio: azioni e obbligazioni high yield potrebbero perdere simultaneamente valore durante crisi economiche o creditizie.

3. CONSERVATIVO

Indice	Peso %
S&P 500 Total Return	10%
MSCI EAFE Net Total Return	7%
MSCI Emerging Markets Net Total Return	3%
US Treasury 10Y Total Return	15%

Indice	**Peso %**
US Treasury 2Y Total Return	25%
US Investment Grade Total Return	30%
US High Yield Corporate Total Return	2%
Gold	8%
Totale	**100%**

Motivazione economica

Il portafoglio limita l'azionario al 20%, mantenendo una partecipazione contenuta alla crescita globale e riducendo l'impatto dei ribassi dei mercati. La componente emergente è circoscritta al 3% per contenere volatilità, rischio politico e rischio valutario. Il 70% è investito in obbligazioni statunitensi: l'elevato peso dei Treasury biennali privilegia stabilità e bassa sensibilità ai tassi, mentre investment grade e Treasury decennali forniscono reddito, qualità creditizia e diversificazione di scadenza. L'high yield è ridotto al 2%, poiché il suo comportamento può avvicinarsi a quello azionario nelle fasi di stress. L'oro, con un peso dell'8%, introduce una fonte indipendente di diversificazione. Il ribilanciamento trimestrale limita l'accumulo involontario di rischio dopo forti movimenti di mercato.

Principale punto di forza: elevata diversificazione difensiva e predominanza di strumenti obbligazionari di qualità elevata.

Principale rischio: rendimenti reali contenuti e possibili perdite sulla componente obbligazionaria in caso di rialzo significativo dei tassi d'interesse.

A.4.2 Claude (Anthropic)

Modello: Claude Sonnet 5 Medio - Data: 07/08/2026 - Rigenerazione con lo stesso prompt congelato (Tabella 3.3), lingua italiana. Non è stato possibile

utilizzare la stessa configurazione dell'esperimento originale (Claude Fable 5 Max), non più disponibile al momento della rigenerazione. Output integrale, riportato senza modifiche.

Portafoglio 1 - AGGRESSIVO

Indice	**Peso %**
S&P 500 Total Return	40
MSCI EAFE Net Total Return	20
MSCI Emerging Markets Net Total Return	15
US Treasury 10Y Total Return	5
US Treasury 2Y Total Return	0
US Investment Grade Total Return	5
US High Yield Corporate Total Return	10
Gold	5
Totale	**100%**

Motivazione economica

L'allocazione destina circa il 75% del capitale ad asset azionari globali, ripartiti tra mercati sviluppati USA, altri mercati sviluppati e mercati emergenti, in coerenza con l'obiettivo di massimizzare la crescita nel lungo periodo. La componente obbligazionaria è ridotta e concentrata su titoli a maggior rendimento atteso (high yield) piuttosto che su strumenti puramente difensivi come i Treasury a breve termine, che vengono esclusi in quanto poco funzionali a un profilo orientato alla crescita. Una piccola quota in oro fornisce diversificazione strutturale senza incidere significativamente sul potenziale di rendimento complessivo. L'assenza di leva e la natura long-only mantengono il rischio entro i limiti di un portafoglio interamente investito in asset rischiosi.

Punto di forza principale: elevata esposizione a fonti di crescita diversificate geograficamente (USA, sviluppati, emergenti).

Rischio principale: elevata volatilità e drawdown potenzialmente ampi in fasi di correzione azionaria globale sincronizzata.

Portafoglio 2 - BILANCIATO

Indice	**Peso %**
S&P 500 Total Return	25
MSCI EAFE Net Total Return	15
MSCI Emerging Markets Net Total Return	10
US Treasury 10Y Total Return	10
US Treasury 2Y Total Return	10
US Investment Grade Total Return	20
US High Yield Corporate Total Return	5
Gold	5
Totale	**100%**

Motivazione economica

La componente azionaria (50%) è bilanciata da una componente obbligazionaria equivalente, articolata su diverse duration e profili di credito: Treasury a 10 e 2 anni per la stabilità, investment grade per un reddito moderato con rischio di credito contenuto, e una quota limitata di high yield per un contributo incrementale di rendimento. La diversificazione geografica azionaria tra USA, mercati sviluppati esteri ed emergenti riduce la concentrazione su una singola area economica. L'oro, in quota contenuta, agisce da elemento di decorrelazione. Questa struttura riflette l'obiettivo di ricercare crescita del capitale mantenendo un profilo di rischio moderato, senza affidarsi a previsioni di mercato o scelte tattiche.

Punto di forza principale: equilibrio strutturale tra fonti di crescita e fonti di stabilità, con buona diversificazione tra asset class.

Rischio principale: in fasi di stress simultaneo su azionario e credito, la correlazione tra le componenti rischiose può ridurre l'effetto di diversificazione atteso.

Portafoglio 3 - CONSERVATIVO

Indice	**Peso %**
S&P 500 Total Return	10
MSCI EAFE Net Total Return	5
MSCI Emerging Markets Net Total Return	0
US Treasury 10Y Total Return	20
US Treasury 2Y Total Return	25
US Investment Grade Total Return	25
US High Yield Corporate Total Return	5
Gold	10
Totale	**100%**

Motivazione economica

L'80% del portafoglio è allocato su strumenti a reddito fisso di elevata qualità creditizia (Treasury a 2 e 10 anni, investment grade), privilegiando la preservazione del capitale e il contenimento della volatilità. I Treasury a 2 anni, con la quota più elevata, riducono la sensibilità complessiva ai tassi d'interesse rispetto a un'esposizione concentrata sulla scadenza decennale. La componente azionaria è limitata e priva di mercati emergenti, esclusi per il maggiore rischio economico e valutario incompatibile con l'obiettivo difensivo. L'oro, in quota moderata, offre un ulteriore elemento di diversificazione rispetto ai rischi di

mercato tradizionali. La struttura risultante minimizza l'esposizione a fonti di rischio elevato mantenendo un contributo residuale di crescita.

Punto di forza principale: bassa volatilità attesa grazie alla forte prevalenza di strumenti obbligazionari a elevato merito creditizio.

Rischio principale: rendimento atteso contenuto e vulnerabilità a fasi prolungate di rialzo dei tassi d'interesse, che penalizzerebbero la componente obbligazionaria a duration più lunga.

A.4.3 Gemini (Google)

Modello: Gemini 3.1 Pro - Data: 07/08/2026 - Rigenerazione con lo stesso prompt congelato (Tabella 3.3), lingua italiana - Output integrale, riportato senza modifiche.

1. AGGRESSIVO

Indice	**Peso %**
S&P 500 Total Return	50
MSCI EAFE Net Total Return	25
MSCI Emerging Markets Net Total Return	15
US Treasury 10Y Total Return	5
Gold	5
US Treasury 2Y Total Return	0
US Investment Grade Total Return	0
US High Yield Corporate Total Return	0
Totale	**100%**

Motivazione economica: L'allocazione destina il 90% del portafoglio al comparto azionario per

massimizzare il premio per il rischio nel lungo periodo. L'S&P 500 funge da motore principale per la creazione di valore, mentre l'esposizione ai mercati sviluppati (EAFE) ed emergenti garantisce un'ampia diversificazione geografica e permette di catturare i trend di crescita globale. La quota residua del 10%, equamente divisa tra US Treasury 10Y e Oro, svolge una funzione di decorrelazione strategica. Questa componente agisce come ammortizzatore di liquidità per attenuare marginalmente gli shock sistemici, senza compromettere il potenziale di rendimento composto richiesto da un profilo orientato alla crescita aggressiva.

Principale punto di forza: Elevato potenziale di rendimento reale nel lungo termine, trainato dalla forte esposizione strutturale ai mercati azionari globali.

Principale rischio: Elevata volatilità e suscettibilità a profondi drawdown durante le fasi di recessione economica o di grave stress dei mercati finanziari.

2. BILANCIATO

Indice	**Peso %**
S&P 500 Total Return	35
US Investment Grade Total Return	20
MSCI EAFE Net Total Return	15
US Treasury 10Y Total Return	15
MSCI Emerging Markets Net Total Return	5
US High Yield Corporate Total Return	5
Gold	5
US Treasury 2Y Total Return	0
Totale	**100%**

Motivazione economica: La struttura riflette un approccio strategico che bilancia equamente la ricerca di

rendimento e il controllo del rischio. L'azionario (55%) assicura la crescita del capitale, ma con un'esposizione più misurata rispetto al profilo aggressivo. Il comparto obbligazionario core (35%), composto da Investment Grade e Treasury a 10 anni, è essenziale per generare un flusso di reddito stabile e fornire una zavorra fondamentale durante le correzioni del mercato azionario. L'inclusione di un 5% in High Yield potenzia il rendimento cedolare assumendo un moderato rischio di credito, mentre il 5% in Oro inserisce un asset reale utile come diversificatore contro potenziali pressioni inflazionistiche.

Principale punto di forza: Eccellente diversificazione delle fonti di rischio che storicamente ottimizza il rapporto rischio/rendimento e attutisce la volatilità ciclica.

Principale rischio: Vulnerabilità a shock macroeconomici che comportano una simultanea correlazione positiva tra azionario e obbligazionario, come improvvisi rialzi dei tassi di interesse.

3. CONSERVATIVO

Indice	**Peso %**
US Treasury 2Y Total Return	30
US Investment Grade Total Return	25
US Treasury 10Y Total Return	15
S&P 500 Total Return	10
MSCI EAFE Net Total Return	10
Gold	10
MSCI Emerging Markets Net Total Return	0
US High Yield Corporate Total Return	0
Totale	**100%**

Motivazione economica: Il portafoglio alloca il 70% in strumenti a reddito fisso di alta qualità (Treasury e Investment Grade), con una marcata concentrazione sulla scadenza a 2 anni. Questa scelta minimizza l'esposizione al rischio di tasso d'interesse e privilegia la preservazione nominale del capitale. L'allocazione azionaria è limitata al 20% ed esclude volutamente i mercati emergenti, concentrandosi sulle aree sviluppate per mitigare l'erosione del potere d'acquisto nel lungo termine senza introdurre volatilità eccessiva. Il 10% in Oro rafforza la natura difensiva del portafoglio, operando come bene rifugio primario nelle fasi di avversione al rischio.

Principale punto di forza: Massima stabilità del capitale e ampiezza estremamente contenuta dei drawdown anche negli scenari di mercato più avversi.

Principale rischio: Rischio di erosione del potere d'acquisto nel lungo termine, in quanto i rendimenti attesi potrebbero faticare a superare significativamente i livelli di inflazione.

A.5 - Rigenerazione (prompt in inglese)

A.5.1 ChatGPT (OpenAI)

Modello: ChatGPT 5.6 Sol Alto - Data: 07/08/2026 - Rigenerazione con lo stesso prompt congelato (Tabella 3.3), tradotto in inglese - Output integrale, riportato senza modifiche.

1. AGGRESSIVE

Index	**Weight %**
S&P 500 Total Return	42%
MSCI EAFE Net Total Return	20%

Index	Weight %
MSCI Emerging Markets Net Total Return	18%
US Treasury 10Y Total Return	3%
US Treasury 2Y Total Return	2%
US Investment Grade Total Return	5%
US High Yield Corporate Total Return	7%
Gold	3%
Total	**100%**

Economic rationale:

The portfolio allocates 80% to equities, emphasizing long-term capital appreciation while diversifying across the United States, developed international markets and emerging markets. The S&P 500 is the largest position because of its broad exposure to established, profitable companies. Emerging markets receive a substantial allocation due to their higher long-term growth potential, despite greater political, currency and economic risk. High-yield bonds add income and return potential but retain significant sensitivity to economic conditions. Investment-grade bonds, Treasuries and gold provide limited diversification and liquidity without materially reducing the portfolio's growth orientation. Quarterly rebalancing systematically restores the strategic risk exposures after market movements.

Main strength: High long-term growth potential combined with meaningful geographic diversification.

Main risk: Severe drawdowns during broad equity and credit-market contractions.

2. BALANCED

Index	Weight %
S&P 500 Total Return	30%
MSCI EAFE Net Total Return	13%
MSCI Emerging Markets Net Total Return	7%
US Treasury 10Y Total Return	12%
US Treasury 2Y Total Return	8%
US Investment Grade Total Return	18%
US High Yield Corporate Total Return	6%
Gold	6%
Total	**100%**

Economic rationale: The portfolio divides its capital between 50% equities and 44% fixed income, with the remaining 6% allocated to gold. Equities provide the principal source of long-term real capital growth, led by US large-cap companies and complemented by developed and emerging-market diversification. Investment-grade bonds form the fixed-income core, providing income and lower expected volatility than equities. Ten-year Treasuries offer potential diversification during periods of economic weakness, while two-year Treasuries reduce overall duration and volatility. High-yield bonds provide additional income but are limited because their risk characteristics can resemble equities. Gold introduces a distinct source of diversification during certain inflationary or financial-stress environments. Quarterly rebalancing maintains an approximately even balance between growth and defensive assets.

Main strength: Balanced exposure to capital growth, income and multiple sources of diversification.

Main risk: Simultaneous equity and bond weakness could reduce the effectiveness of diversification.

3. CONSERVATIVE

Index	**Weight %**
S&P 500 Total Return	12%
MSCI EAFE Net Total Return	5%
MSCI Emerging Markets Net Total Return	3%
US Treasury 10Y Total Return	12%
US Treasury 2Y Total Return	30%
US Investment Grade Total Return	28%
US High Yield Corporate Total Return	3%
Gold	7%
Total	**100%**

Economic rationale: The portfolio assigns 70% to high-quality fixed income, emphasizing stability, income and drawdown containment. Two-year Treasuries are the largest allocation because their short maturity limits interest-rate sensitivity and supports capital stability. Investment-grade bonds provide additional income while maintaining relatively high credit quality. Ten-year Treasuries add duration-based diversification but are constrained to limit losses from rising interest rates. A 20% equity allocation preserves some long-term growth and inflation-offsetting potential, with emerging markets kept small because of their higher volatility. High-yield exposure is minimal due to its credit and equity-like risks. Gold provides diversification against selected inflationary, currency and financial-stress conditions. Quarterly rebalancing prevents riskier assets from becoming disproportionately large.

Main strength: Strong emphasis on high-quality, relatively low-volatility fixed-income assets.

Main risk: Limited long-term capital growth and vulnerability to losses in real purchasing power.

A.5.2 Claude (Anthropic)

Modello: Claude Sonnet 5 Medio - Data: 07/08/2026 - Rigenerazione con lo stesso prompt congelato (Tabella 3.3), tradotto in inglese. Non è stato possibile utilizzare la stessa configurazione dell'esperimento originale (Claude Fable 5 Max), non più disponibile al momento della rigenerazione - Output integrale, riportato senza modifiche.

1. AGGRESSIVE

Index	**Weight %**
S&P 500 Total Return	40
MSCI EAFE Net Total Return	20
MSCI Emerging Markets Net Total Return	20
US Treasury 10Y Total Return	5
US Treasury 2Y Total Return	0
US Investment Grade Total Return	5
US High Yield Corporate Total Return	5
Gold	5
Total	**100%**

Rationale:

The portfolio concentrates 80% in global equities (US, developed ex-US, emerging markets), reflecting the long-term equity risk premium as the primary

growth driver. Broad geographic diversification across developed and emerging markets spreads country and currency risk while retaining full participation in growth cycles. A small High Yield allocation adds a credit-risk-driven return source correlated with equity cycles. Minimal exposure to Treasuries and Investment Grade bonds serves only as marginal shock absorption, not as a core stabilizer, consistent with an investor prioritizing capital appreciation over drawdown control. A small Gold position provides diversification against tail risks without diluting the growth orientation.

Main strength: Maximum long-term exposure to global equity risk premia, supporting the highest expected compounding growth among the three profiles.

Main risk: High sensitivity to global equity drawdowns and prolonged bear markets, with limited buffering from fixed income.

2. BALANCED

Index	**Weight %**
S&P 500 Total Return	25
MSCI EAFE Net Total Return	10
MSCI Emerging Markets Net Total Return	5
US Treasury 10Y Total Return	15
US Treasury 2Y Total Return	10
US Investment Grade Total Return	20
US High Yield Corporate Total Return	5
Gold	10
Total	**100%**

Rationale:

Equities represent 40% of the portfolio, providing growth while fixed income (50%) offers stability and income across the maturity and credit spectrum

(Treasury 10Y, Treasury 2Y, Investment Grade). This layered bond exposure balances interest rate risk with reduced volatility. A modest High Yield allocation adds incremental return without significant credit concentration. Gold at 10% enhances diversification, given its historically low correlation with traditional asset classes. Geographic equity diversification is retained but tilted toward developed markets, reducing emerging market volatility exposure relative to the aggressive profile. The overall structure aims to capture reasonable long-term growth while meaningfully containing drawdowns.

Main strength: Diversification across equities, multiple fixed income segments, and gold reduces reliance on any single risk factor.

Main risk: Moderate equity and credit exposure still leaves the portfolio vulnerable to simultaneous equity and credit market stress.

3. CONSERVATIVE

Index	**Weight %**
S&P 500 Total Return	10
MSCI EAFE Net Total Return	5
MSCI Emerging Markets Net Total Return	0
US Treasury 10Y Total Return	20
US Treasury 2Y Total Return	25
US Investment Grade Total Return	25
US High Yield Corporate Total Return	5
Gold	10
Total	**100%**

Rationale:

Fixed income dominates the allocation (75%), with a strong tilt toward Treasury 2Y and Investment Grade for capital stability and lower interest rate sensitivity

relative to longer duration alone. Treasury 10Y adds moderate duration for diversification within the rate curve. Limited equity exposure (15%), concentrated in developed markets, provides a modest long-term growth contribution without materially raising volatility. Emerging markets are excluded due to their higher risk profile, inconsistent with capital preservation objectives. A small High Yield position and a 10% Gold allocation offer marginal diversification benefits. The structure prioritizes volatility containment and principal stability over growth maximization.

Main strength: High allocation to short- and investment-grade fixed income provides strong capital stability and low volatility.

Main risk: Elevated sensitivity to interest rate movements affecting the bond-heavy allocation, alongside limited long-term growth potential.

A.5.3 Gemini (Google)

Modello: Gemini 3.1 Pro - Data: 07/08/2026 - Rigenerazione con lo stesso prompt congelato (Tabella 3.3), tradotto in inglese - Output integrale, riportato senza modifiche.

AGGRESSIVE

Index	**Weight %**
S&P 500 Total Return	45%
MSCI EAFE Net Total Return	20%
MSCI Emerging Markets Net Total Return	15%
US High Yield Corporate Total Return	10%
US Treasury 10Y Total Return	5%
Gold	5%

Index	Weight %
US Treasury 2Y Total Return	0%
US Investment Grade Total Return	0%
Total	**100%**

Economic Rationale

This allocation heavily favors risk assets to maximize long-term compounding, committing 80% to global equities. The S&P 500 acts as the core growth engine, complemented by substantial allocations to EAFE and Emerging Markets for capturing varied global economic cycles and demographic trends. A 10% allocation to High Yield adds an equity-correlated income stream. To prevent ruinous drawdowns during severe deflationary shocks or market panics, a minor 10% allocation is split between 10-Year Treasuries and Gold. This defensive sliver provides structural diversification and liquidity for quarterly rebalancing without significantly dragging down long-term expected returns.

Main strength of the allocation: High capacity for capital accumulation through broad exposure to global equity risk premiums.

Main risk of the allocation: Severe vulnerability to synchronized global equity drawdowns and prolonged bear markets.

BALANCED

Index	Weight %
S&P 500 Total Return	35%
MSCI EAFE Net Total Return	15%
US Treasury 10Y Total Return	15%
US Investment Grade Total Return	15%
MSCI Emerging Markets Net Total Return	5%

Index	Weight %
US Treasury 2Y Total Return	5%
US High Yield Corporate Total Return	5%
Gold	5%
Total	**100%**

Economic Rationale

This portfolio is constructed to capture steady growth while dampening volatility through a diversified structural approach. The S&P 500 remains the cornerstone, paired with moderate EAFE and Emerging Markets exposure to ensure global equity participation. The fixed-income block is strategically layered: US Investment Grade provides core income, while 10-Year Treasuries offer crucial duration protection against growth shocks. A small allocation to 2-Year Treasuries acts as a volatility buffer. High Yield and Gold provide satellite diversification, contributing extra yield and real-asset protection, respectively. This mix harnesses equity beta for growth but relies on high-quality bonds to absorb shocks, smoothing the long-term compounding trajectory.

Main strength of the allocation: Robust structural resilience that balances risk-adjusted returns across a wide variety of economic environments.

Main risk of the allocation: Potential correlation spikes between equities and bonds during inflationary shocks, reducing the effectiveness of the portfolio's core diversification.

CONSERVATIVE

Index	Weight %
US Treasury 2Y Total Return	30%
US Treasury 10Y Total Return	20%
US Investment Grade Total Return	20%

Index	Weight %
S&P 500 Total Return	15%
MSCI EAFE Net Total Return	5%
US High Yield Corporate Total Return	5%
Gold	5%
MSCI Emerging Markets Net Total Return	0%
Total	**100%**

Economic Rationale

Capital preservation is achieved by dedicating 70% of the portfolio to high-quality fixed income. A heavy 30% allocation to 2-Year Treasuries anchors the portfolio, severely restricting volatility and providing extreme liquidity. US Investment Grade and 10-Year Treasuries provide the necessary income and duration to protect against deflationary pressures. Equities are strictly limited to 20%, heavily tilted toward the S&P 500 for stability, serving primarily to defend purchasing power against inflation rather than drive aggressive growth. Emerging Markets are entirely excluded due to their outsized drawdown risk. Small allocations to High Yield and Gold offer marginal yield enhancement and tail-risk diversification without compromising the mandate of stability.

Main strength of the allocation: Superior capital protection and minimal drawdown risk, particularly during severe equity market contractions.

Main risk of the allocation: Low real returns that may struggle to outpace higher inflation environments over the long term, resulting in purchasing power erosion.

APPENDICE B - CODICE

B.0 - Istruzioni di replica

Gli script riportati in questa appendice sono stati sviluppati in Python 3.14.3, con le librerie pandas 3.0.3, numpy 2.5.1, matplotlib 3.11.0 e openpyxl 3.1.5. Il codice è stato sviluppato con il supporto di uno strumento di intelligenza artificiale generativa, sotto la supervisione e verifica dello scrivente, che ne ha validato la correttezza tramite i test su casi a soluzione nota descritti nel paragrafo 3.4 (Appendice B.2).

Gli script presuppongono la struttura di cartelle descritta nel paragrafo 3.1: una cartella «dati_grezzi» con i file di origine così come esportati dalle piattaforme (LSEG e FRED), una cartella «codice» contenente gli script, e una cartella «output» in cui vengono salvati i file prodotti (dataset puliti, tabelle di risultati, grafici).

Gli script vanno eseguiti nell'ordine seguente, poiché ciascuno utilizza gli output prodotti dal precedente:

B.1 (carica_dati.py) - legge i file grezzi, applica le correzioni documentate in Appendice C e produce il dataset giornaliero 2016-2025;

B.2 (backtest.py) - definisce il motore di backtest e le metriche di valutazione, e ne esegue la validazione su casi sintetici;

B.3 (backtest_benchmark.py) - applica il motore ai quattro benchmark;

B.4 (backtest_portafogli_ai.py) - verifica la validità formale dei nove portafogli generati dai modelli e applica loro il medesimo motore di backtest.

Il quinto script, B.5 (grafici.py), produce le figure impiegate nel Capitolo 4 a partire dagli output già calcolati da B.3 e B.4, senza ripetere alcun calcolo. Il

codice sorgente dei cinque script è riportato per esteso da B.1 a B.5: come il registro delle correzioni in Appendice C.

B.1 - carica_dati.py

Codice sorgente (218 righe).

```
# -*- coding: utf-8 -*-
"""
======================================================================
 CARICAMENTO E PULIZIA DATI - Tesi: AI generativa e strategie di investimento
======================================================================
 Cosa fa questo script:
   1. Legge i 9 file grezzi (8 serie LSEG Workspace + risk-free FRED)
      dalla cartella ../dati_grezzi, cosi' come esportati (mai modificati a
mano).
   2. Applica le correzioni documentate (bad tick del 18/01/2022 sulle
      serie FTSE) e scrive un LOG delle correzioni in ../output.
   3. Allinea tutte le serie sul calendario di borsa USA (date S&P 500),
      colmando i festivi esteri/FX con l'ultimo valore disponibile.
   4. Produce:
      - ../output/master_giornaliero_2016_2025.csv  (backtest)
      - ../output/master_mensile_pre2016.csv        (statistiche opzionali)
      - ../output/log_correzioni.txt                (appendice tesi)
======================================================================
"""

import os
import numpy as np
import pandas as pd
import openpyxl

CARTELLA_GREZZI = os.path.join("..", "dati_grezzi", "Eikon")
CARTELLA_OUTPUT = os.path.join("..", "output")

# ----------------------------------------------------------------
# Anagrafica dei file: nome file -> (nome breve, colonna prezzo attesa)
# ----------------------------------------------------------------
SERIE = {
    "1_SPXTR.xlsx":         ("SP500",      None),
    "2_MSCI_EAFE.xlsx":     ("EAFE",       None),
    "3_MSCI_EM.xlsx":       ("EM",         None),
    "4b_Treasury_10y.xlsx": ("TREAS_10Y",  None),
    "5_FTSE_Treasury_2y.xlsx":   ("TREAS_2Y",   None),
    "6_FTSE_IG.xlsx":             ("IG",         None),
    "7_Corporate_HY.xlsx":  ("CORP_HY",    None),
    "8_Gold.xlsx":          ("ORO",        None),
}
FILE_RISKFREE = "9_DTB3.csv"   # export FRED
```

```
# Correzioni documentate: (nome breve, data)
# Bad tick 18/01/2022: valore ribassato del 3-11% con pieno recupero il
giorno
# successivo, sincronizzato su piu' serie FTSE e incompatibile con la
duration
# (il rendimento decennale USA vario' di ~9 pb quel giorno, non ~125 pb).
CORREZIONI = [
    ("TREAS_10Y", "2022-01-18", "Errore di pubblicazione indice FTSE (bad
tick, reversal completo il 19/01)"),
    ("TREAS_2Y",  "2022-01-18", "Errore di pubblicazione indice FTSE (bad
tick, reversal completo il 19/01)"),
    ("IG",   "2022-01-18", "Errore di pubblicazione indice FTSE (bad tick,
reversal completo il 19/01)"),
]

INIZIO_TEST, FINE_TEST = "2016-01-01", "2025-12-31"


def leggi_export_workspace(percorso):
    """Legge un export Price History di LSEG Workspace.

    Trova la riga di intestazione ('Exchange Date'), individua la colonna
    prezzo (Close > Trade Price > Bid) e restituisce una Series giornaliera
    ordinata in senso crescente.
    """
    wb = openpyxl.load_workbook(percorso, read_only=True)
    ws = wb[wb.sheetnames[0]]
    riga_header = None
    for i, row in enumerate(ws.iter_rows(min_row=1, max_row=60,
values_only=True), start=1):
        if row and row[0] in ("Exchange Date", "Date"):
            riga_header = i
            break
    wb.close()
    if riga_header is None:
        raise ValueError(f"Intestazione 'Exchange Date' non trovata in
{percorso}")

    df = pd.read_excel(percorso, skiprows=riga_header - 1)
    df = df.rename(columns=lambda c: str(c).strip())
    col_prezzo = next((c for c in ("Close", "Trade Price", "Bid") if c in
df.columns), None)
    if col_prezzo is None:
        raise ValueError(f"Nessuna colonna prezzo riconosciuta in {percorso}:
{list(df.columns)}")

    # alcuni export FTSE hanno due colonne data: si usa 'Exchange Date' se
presente, altrimenti 'Date'
    col_data = "Exchange Date" if "Exchange Date" in df.columns else "Date"
    df = df.rename(columns={col_data: "Exchange Date"})
    df["Exchange Date"] = pd.to_datetime(df["Exchange Date"],
errors="coerce")
    df = (df.dropna(subset=["Exchange Date", col_prezzo])
            .sort_values("Exchange Date")
```

```
            .set_index("Exchange Date"))
    s = df[col_prezzo].astype(float)
    s = s[~s.index.duplicated(keep="last")]
    return s, col_prezzo


def main():
    os.makedirs(CARTELLA_OUTPUT, exist_ok=True)
    log = []
    log.append("LOG DELLE CORREZIONI E DELLE SCELTE DI PULIZIA DATI")
    log.append("=" * 60)

    # ------------------------------------------------------------
    # 1. Lettura dei grezzi
    # ------------------------------------------------------------
    print("1) Lettura file grezzi da", CARTELLA_GREZZI)
    prezzi = {}
    for nome_file, (nome, _) in SERIE.items():
        percorso = os.path.join(CARTELLA_GREZZI, nome_file)
        s, col = leggi_export_workspace(percorso)
        prezzi[nome] = s
        print(f"   {nome:<10} {nome_file:<24} colonna '{col}', "
              f"{len(s)} oss., {s.index.min().date()} ->
{s.index.max().date()}")
        log.append(f"{nome}: file '{nome_file}', colonna prezzo '{col}', "
                   f"{len(s)} osservazioni ({s.index.min().date()} -
{s.index.max().date()})")

    # ------------------------------------------------------------
    # 2. Correzioni documentate (bad tick) con interpolazione
    # ------------------------------------------------------------
    print("\n2) Correzioni documentate")
    log.append("\nCORREZIONI PUNTUALI (valore rimosso e interpolato
linearmente):")
    for nome, data, motivo in CORREZIONI:
        data = pd.Timestamp(data)
        s = prezzi[nome]
        if data not in s.index:
            print(f"   ATTENZIONE: {data.date()} non presente in {nome},
nessuna correzione")
            continue
        vecchio = s.loc[data]
        s.loc[data] = np.nan
        s = s.interpolate(method="linear", limit_area="inside")
        nuovo = s.loc[data]
        prezzi[nome] = s
        print(f"   {nome:<10} {data.date()}: {vecchio:,.2f} -> {nuovo:,.2f}
({motivo})")
        log.append(f"  {nome} {data.date()}: rimosso {vecchio:,.4f},
interpolato {nuovo:,.4f}. Motivo: {motivo}")

    # ------------------------------------------------------------
    # 3. Calendario master = giorni di borsa USA (date S&P 500)
    #    Le altre serie vengono riallineate; i buchi da festivita'
```

```
    #    estere/FX si colmano con l'ultimo valore disponibile (ffill,
    #    max 5 giorni). Il conteggio dei riempimenti finisce nel log.
    # ------------------------------------------------------------
    print("\n3) Allineamento sul calendario di borsa USA")
    log.append("\nALLINEAMENTO CALENDARI (calendario master: giorni di
negoziazione S&P 500):")
    cal_test = prezzi["SP500"].loc[INIZIO_TEST:FINE_TEST].index

    master = pd.DataFrame(index=cal_test)
    for nome, s in prezzi.items():
        riallineata = s.reindex(cal_test)
        n_buchi = int(riallineata.isna().sum())
        riallineata = riallineata.ffill(limit=5)
        residui = int(riallineata.isna().sum())
        master[nome] = riallineata
        print(f"   {nome:<10} giorni colmati con ffill: {n_buchi:>3} | NA
residui: {residui}")
        log.append(f"  {nome}: {n_buchi} date colmate con ultimo valore
disponibile (festivita' locali), {residui} NA residui")

    if master.isna().any().any():
        raise ValueError("NA residui nel master giornaliero: verificare i
dati")

    # ------------------------------------------------------------
    # 4. Risk-free (FRED DTB3): tasso % annuo, ffill dei festivi
    # ------------------------------------------------------------
    print("\n4) Risk-free FRED DTB3")
    rf = pd.read_csv(os.path.join(CARTELLA_GREZZI, FILE_RISKFREE))
    rf.columns = ["data", "tasso"]
    rf["data"] = pd.to_datetime(rf["data"])
    rf = rf.set_index("data")["tasso"]
    rf = pd.to_numeric(rf, errors="coerce")
    rf = rf.reindex(cal_test).ffill(limit=10)
    print(f"   DTB3: {rf.notna().sum()} valori sul calendario master, "
          f"media periodo test {rf.mean():.2f}%")
    log.append(f"\nRISK-FREE: FRED DTB3 (discount basis, % annua), festivita'
colmate con ffill. Media 2016-2025: {rf.mean():.2f}%")
    master["RF_DTB3"] = rf

    # ------------------------------------------------------------
    # 5. Master mensile pre-cutoff (statistiche storiche opzionali)
    #    Fine mese, dal primo dato disponibile di ciascuna serie al
31/12/2015.
    # ------------------------------------------------------------
    print("\n5) Master mensile pre-cutoff (fino al 31/12/2015)")
    mensile = {}
    for nome, s in prezzi.items():
        pre = s.loc[:"2015-12-31"]
        if len(pre) == 0:
            continue
        m = pre.resample("ME").last()
        mensile[nome] = m
    pre_master = pd.DataFrame(mensile)
```

```
    print(f"   Serie coperte: {list(pre_master.columns)}")
    log.append("\nMASTER MENSILE PRE-CUTOFF: fine mese, dal primo dato
disponibile per serie al 31/12/2015.")
    log.append("  Nota: le serie FTSE Treasury (10Y, 2Y) partono dal 2009 e
presentano due lacune nel 2011;")
    log.append("  per eventuali statistiche storiche sui Treasury usare la
finestra dal 2012.")

    # ------------------------------------------------------------
    # 6. Salvataggi
    # ------------------------------------------------------------
    p1 = os.path.join(CARTELLA_OUTPUT, "master_giornaliero_2016_2025.csv")
    p2 = os.path.join(CARTELLA_OUTPUT, "master_mensile_pre2016.csv")
    p3 = os.path.join(CARTELLA_OUTPUT, "log_correzioni.txt")
    master.to_csv(p1, float_format="%.6f")
    pre_master.to_csv(p2, float_format="%.6f")
    with open(p3, "w", encoding="utf-8") as f:
        f.write("\n".join(log) + "\n")

    print("\n6) File prodotti:")
    print("  ", p1, f"({master.shape[0]} righe x {master.shape[1]} colonne)")
    print("  ", p2, f"({pre_master.shape[0]} righe x {pre_master.shape[1]}
colonne)")
    print("  ", p3)

    # ------------------------------------------------------------
    # 7. Riepilogo di controllo
    # ------------------------------------------------------------
    print("\n7) Controllo finale sul master giornaliero 2016-2025:")
    rend = master.drop(columns=["RF_DTB3"]).pct_change().dropna()
    anni = len(master) / 252
    riep = pd.DataFrame({
        "CAGR": (master.drop(columns=["RF_DTB3"]).iloc[-1] /
master.drop(columns=["RF_DTB3"]).iloc[0]) ** (1 / anni) - 1,
        "Vol_ann": rend.std() * np.sqrt(252),
    })
    print((riep * 100).round(2).astype(str) + "%")
    print("\nCOMPLETATO: dataset master pronto per il motore di backtest.")


if __name__ == "__main__":
    main()
```

B.2 - backtest.py

Codice sorgente (377 righe).

```
# -*- coding: utf-8 -*-
"""
======================================================================
 MOTORE DI BACKTEST - Tesi: AI generativa e strategie di investimento
```

```
================================================================
 Cosa fa questo script:
   1. Definisce il motore di backtest: dato un'insieme di serie di
      Prezzo nella versione total return e un vettore di pesi, simula un
      portafoglio statico con ribilanciamento trimestrale.
   2. Calcola le metriche del progetto: CAGR, volatilita' annualizzata,
      Sharpe, Maximum Drawdown, Calmar.
   3. Esegue una serie di TEST DI VALIDAZIONE su dati sintetici con
      proprieta' note, per dimostrare che il motore calcola i valori
      corretti PRIMA di usarlo sui dati reali.


 Non serve alcun file di input: i dati di collaudo sono generati
 dallo script stesso. Quando arriveranno i dati Eikon, questo stesso
 motore verra' richiamato da un secondo script di caricamento dati.
================================================================
"""

import numpy as np
import pandas as pd
import matplotlib
matplotlib.use("Agg")  # salva i grafici su file senza aprire finestre
import matplotlib.pyplot as plt
import os

# ----------------------------------------------------------------
# Impostazioni generali
# ----------------------------------------------------------------
GIORNI_ANNO = 252          # convenzione: giorni di borsa in un anno
CARTELLA_OUTPUT = os.path.join("..", "output")  # ../output rispetto a
"codice"

# ================================================================
# 1. MOTORE DI BACKTEST
# ================================================================

def date_ribilanciamento(indice_date):
    """Restituisce le date di ribilanciamento trimestrale.

    Regola: il PRIMO giorno di borsa di ogni trimestre solare
    (gennaio, aprile, luglio, ottobre). La prima data della serie
    e' sempre inclusa (costituzione del portafoglio).
    """
    df = pd.DataFrame(index=indice_date)
    df["anno"] = df.index.year
    df["trimestre"] = df.index.quarter
    # primo giorno di borsa di ciascun (anno, trimestre)
    prime_date = df.groupby(["anno", "trimestre"]).apply(lambda g:
g.index.min())
    date_rib = set(pd.DatetimeIndex(prime_date.values))
    date_rib.add(indice_date[0])
    return date_rib
```

```python
def backtest_portafoglio(prezzi, pesi, capitale_iniziale=100.0):
    """Simula un portafoglio statico con ribilanciamento trimestrale.

    Parametri
    ---------
    prezzi : DataFrame giornaliero (righe=date, colonne=indici),
             serie di prezzo TOTAL RETURN, gia' allineate e senza NA.
    pesi   : dict {nome_colonna: peso in %}. I pesi devono sommare a 100.
             Gli indici non citati hanno peso 0.

    Ritorna
    -------
    valore    : Series giornaliera del valore del portafoglio.
    pesi_eff  : DataFrame giornaliero dei pesi effettivi (deriva inclusa),
                utile per verificare il ribilanciamento.
    """
    # --- controlli di coerenza sull'input ---
    somma = sum(pesi.values())
    if abs(somma - 100.0) > 1e-9:
        raise ValueError(f"I pesi devono sommare a 100, non a {somma}")
    for nome in pesi:
        if nome not in prezzi.columns:
            raise ValueError(f"Indice '{nome}' assente dai dati")
    if prezzi.isna().any().any():
        raise ValueError("I prezzi contengono valori mancanti: pulire prima i
dati")

    colonne = list(prezzi.columns)
    w_target = np.array([pesi.get(c, 0.0) / 100.0 for c in colonne])
    px = prezzi[colonne].to_numpy(dtype=float)
    date_rib = date_ribilanciamento(prezzi.index)

    n_giorni, n_asset = px.shape
    valore = np.empty(n_giorni)
    pesi_eff = np.empty((n_giorni, n_asset))

    quote = np.zeros(n_asset)  # numero di "quote" detenute per ciascun
indice
    for t in range(n_giorni):
        data = prezzi.index[t]
        if t == 0:
            val = capitale_iniziale
        else:
            val = float(np.dot(quote, px[t]))
        # nel giorno di ribilanciamento si riportano i pesi al target
        if data in date_rib:
            quote = val * w_target / px[t]
        valore[t] = val
        pesi_eff[t] = (quote * px[t]) / val if val > 0 else 0.0

    valore = pd.Series(valore, index=prezzi.index, name="valore")
    pesi_eff = pd.DataFrame(pesi_eff, index=prezzi.index, columns=colonne)
    return valore, pesi_eff
```

```
# ============================================================
# 2. METRICHE
# ============================================================

def var_es_storico(rendimenti, livello=0.95):
    """VaR ed Expected Shortfall storici (non parametrici), su base
GIORNALIERA.

    Metodo: si ordinano i rendimenti giornalieri dal peggiore al migliore e
si
    prende il k-esimo peggiore, dove k = floor((1-livello) * n). Nessuna
    assunzione sulla forma della distribuzione (a differenza del VaR
    parametrico, che assume normalita').

    Non vengono annualizzati: la regola di scala radice-del-tempo presuppone
    rendimenti indipendenti e normali, l'assunzione che il metodo storico
    e' pensato per evitare. Restano quindi misure di perdita potenziale in
    un singolo giorno di negoziazione.

    Ritorna (VaR, ES) come numeri POSITIVI (frazione di perdita, es. 0.025 =
2,5%).
    """
    r = np.sort(np.asarray(rendimenti))
    n = len(r)
    q = 1.0 - livello
    k = max(int(np.floor(q * n)), 1)
    var = -r[k - 1]          # k-esimo rendimento peggiore
    es = -r[:k].mean()       # media dei k rendimenti peggiori (coda)
    return var, es


def calcola_metriche(valore, risk_free_annuo=0.0):
    """Calcola le metriche del progetto da una serie giornaliera di valore.

    risk_free_annuo: tasso privo di rischio annuo (es. 0.02 = 2%).
    Ritorna un dict con: CAGR, Volatilita', Sharpe, MaxDD, Calmar.
    """
    rend = valore.pct_change().dropna()
    n = len(rend)
    anni = n / GIORNI_ANNO

    # CAGR: dipende solo dal primo e dall'ultimo valore
    cagr = (valore.iloc[-1] / valore.iloc[0]) ** (1.0 / anni) - 1.0

    # Volatilita' annualizzata
    vol = rend.std(ddof=1) * np.sqrt(GIORNI_ANNO)

    # Rendimenti in eccesso sul risk-free (convertito a giornaliero)
    rf_g = (1.0 + risk_free_annuo) ** (1.0 / GIORNI_ANNO) - 1.0
    ecc = rend - rf_g

    # Sharpe annualizzato
    sharpe = np.nan
```

```
    if ecc.std(ddof=1) > 0:
        sharpe = ecc.mean() / ecc.std(ddof=1) * np.sqrt(GIORNI_ANNO)

    # Maximum drawdown su serie giornaliera
    massimo_prec = valore.cummax()
    drawdown = valore / massimo_prec - 1.0
    max_dd = drawdown.min()

    # Calmar
    calmar = cagr / abs(max_dd) if max_dd < 0 else np.nan

    # VaR ed Expected Shortfall storici al 95%, giornalieri (vedi
var_es_storico)
    var_95, es_95 = var_es_storico(rend, livello=0.95)

    return {
        "CAGR": cagr,
        "Volatilita_ann": vol,
        "Sharpe": sharpe,
        "Max_Drawdown": max_dd,
        "Calmar": calmar,
        "VaR_95": var_95,
        "ES_95": es_95,
    }


def tabella_metriche(dizionario_serie, risk_free_annuo=0.0):
    """Da un dict {nome: serie di valore} produce la tabella delle
metriche."""
    righe = {}
    for nome, serie in dizionario_serie.items():
        righe[nome] = calcola_metriche(serie, risk_free_annuo)
    tab = pd.DataFrame(righe).T
    return tab

# ============================================================
# 3. TEST DI VALIDAZIONE SU DATI SINTETICI
# ============================================================

def _esito(nome, condizione, dettaglio=""):
    stato = "OK " if condizione else "FALLITO"
    print(f"  [{stato}] {nome}" + (f"  ({dettaglio})" if dettaglio else ""))
    return bool(condizione)


def esegui_test():
    print("=" * 68)
    print("TEST DI VALIDAZIONE DEL MOTORE DI BACKTEST (dati sintetici)")
    print("=" * 68)
    tutti_ok = True

    # Calendario di borsa fittizio: giorni feriali 2016-2025
    date = pd.date_range("2016-01-01", "2025-12-31", freq="B")
    n = len(date)
```

```
    rng = np.random.default_rng(42)  # seme fisso: risultati riproducibili

    # ------------------------------------------------------------
    # TEST 1: crescita costante nota -> CAGR esatto
    # Serie costruita per rendere esattamente il 5% annuo composto.
    # ------------------------------------------------------------
    print("\nTEST 1 - CAGR su crescita costante del 5% annuo")
    g = (1.05) ** (1.0 / GIORNI_ANNO) - 1.0
    serie_5pct = pd.Series(100 * np.cumprod(np.full(n, 1 + g)), index=date)
    m = calcola_metriche(serie_5pct)
    tutti_ok &= _esito("CAGR atteso 5,00%", abs(m["CAGR"] - 0.05) < 1e-6,
                       f"calcolato {m['CAGR']:.4%}")
    tutti_ok &= _esito("Volatilita' attesa 0", m["Volatilita_ann"] < 1e-9,
                       f"calcolata {m['Volatilita_ann']:.2e}")

    # ------------------------------------------------------------
    # TEST 2: drawdown costruito a tavolino -> MDD esatto
    # La serie sale a 130, scende esattamente a 91 (-30%), poi risale.
    # ------------------------------------------------------------
    print("\nTEST 2 - Maximum Drawdown noto del -30%")
    v = np.concatenate([
        np.linspace(100, 130, n // 3),
        np.linspace(130, 91, n // 3),          # -30% esatto dal picco
        np.linspace(91, 140, n - 2 * (n // 3)),
    ])
    serie_dd = pd.Series(v, index=date)
    m = calcola_metriche(serie_dd)
    tutti_ok &= _esito("MDD atteso -30,00%", abs(m["Max_Drawdown"] + 0.30) <
1e-9,
                       f"calcolato {m['Max_Drawdown']:.4%}")

    # ------------------------------------------------------------
    # TEST 3: portafoglio 100% su un indice = l'indice stesso
    # ------------------------------------------------------------
    print("\nTEST 3 - Portafoglio 100% su un solo indice replica l'indice")
    prezzi = pd.DataFrame({
        "Indice_A": 100 * np.cumprod(1 + rng.normal(0.0004, 0.010, n)),
        "Indice_B": 100 * np.cumprod(1 + rng.normal(0.0002, 0.006, n)),
        "Indice_C": 100 * np.cumprod(1 + rng.normal(0.0003, 0.012, n)),
    }, index=date)
    val_100A, _ = backtest_portafoglio(prezzi, {"Indice_A": 100})
    rend_port = val_100A.pct_change().dropna()
    rend_A = prezzi["Indice_A"].pct_change().dropna()
    scarto = (rend_port - rend_A).abs().max()
    tutti_ok &= _esito("Scarto massimo dei rendimenti ~ 0", scarto < 1e-12,
                       f"scarto {scarto:.2e}")

    # ------------------------------------------------------------
    # TEST 4: i pesi tornano al target a ogni ribilanciamento
    # Portafoglio 60/40 su due indici con andamenti divergenti.
    # ------------------------------------------------------------
    print("\nTEST 4 - Ribilanciamento trimestrale: pesi riportati al target")
    val_6040, pesi_eff = backtest_portafoglio(
        prezzi, {"Indice_A": 60, "Indice_B": 40})
```

```
    d_rib = sorted(date_ribilanciamento(prezzi.index))
    # nei giorni di ribilanciamento il peso di A deve essere esattamente 60%
    pesi_ai_rib = pesi_eff.loc[pesi_eff.index.isin(d_rib), "Indice_A"]
    err_rib = (pesi_ai_rib - 0.60).abs().max()
    tutti_ok &= _esito("Peso al target nei giorni di ribilanciamento",
                       err_rib < 1e-12, f"errore max {err_rib:.2e}")
    # tra un ribilanciamento e l'altro i pesi DEVONO poter derivare
    deriva = (pesi_eff["Indice_A"] - 0.60).abs().max()
    tutti_ok &= _esito("Deriva dei pesi presente tra i ribilanciamenti",
                       deriva > 0.001, f"deriva max {deriva:.2%}")
    # numero atteso di ribilanciamenti: 4 trimestri x 10 anni = 40 date
    tutti_ok &= _esito("Numero date di ribilanciamento = 40",
                       len(d_rib) == 40, f"trovate {len(d_rib)}")

    # ------------------------------------------------------------
    # TEST 5: coerenza contabile - nessun valore creato dal nulla
    # Un portafoglio ribilanciato di indici che rendono tutti
    # esattamente il 5% deve rendere esattamente il 5%.
    # ------------------------------------------------------------
    print("\nTEST 5 - Coerenza contabile del ribilanciamento")
    stesso = pd.Series(100 * np.cumprod(np.full(n, 1 + g)), index=date)
    prezzi_uguali = pd.DataFrame({"X": stesso, "Y": stesso * 2, "Z": stesso *
0.5},
                                 index=date)
    val_mix, _ = backtest_portafoglio(prezzi_uguali, {"X": 50, "Y": 30, "Z":
20})
    m = calcola_metriche(val_mix)
    tutti_ok &= _esito("CAGR del mix = 5,00%", abs(m["CAGR"] - 0.05) < 1e-6,
                       f"calcolato {m['CAGR']:.4%}")

    # ------------------------------------------------------------
    # TEST 6: Sharpe ~ 0 quando il rendimento medio = risk-free
    # ------------------------------------------------------------
    print("\nTEST 6 - Sharpe nullo se il rendimento non supera il risk-free")
    rf = 0.03
    rf_g = (1 + rf) ** (1 / GIORNI_ANNO) - 1
    rumore = rng.normal(0.0, 0.008, n)
    rend_neutri = rf_g + rumore - rumore.mean()  # media ESATTAMENTE = rf
    serie_neutra = pd.Series(100 * np.cumprod(1 + rend_neutri), index=date)
    m = calcola_metriche(serie_neutra, risk_free_annuo=rf)
    tutti_ok &= _esito("Sharpe atteso ~ 0", abs(m["Sharpe"]) < 0.02,
                       f"calcolato {m['Sharpe']:.4f}")

    # ------------------------------------------------------------
    # TEST 7: VaR ed Expected Shortfall storici su distribuzione nota
    # 200 rendimenti costruiti a tavolino: 190 pari a +0,1%, i 10 peggiori
    # (5% esatto di 200) pari a -8%. Essendo la coda tutta identica,
    # sia il VaR che l'ES al 95% devono risultare esattamente 8%.
    # ------------------------------------------------------------
    print("\nTEST 7 - VaR ed ES storici al 95% su distribuzione nota")
    r_test = np.concatenate([np.full(10, -0.08), np.full(190, 0.001)])
    var_t, es_t = var_es_storico(r_test, livello=0.95)
    tutti_ok &= _esito("VaR_95 atteso 8,00%", abs(var_t - 0.08) < 1e-9,
                       f"calcolato {var_t:.4%}")
```

```
    tutti_ok &= _esito("ES_95 atteso 8,00%", abs(es_t - 0.08) < 1e-9,
                       f"calcolato {es_t:.4%}")
    # coda asimmetrica: 8 valori a -10%, 2 valori a -2% (i 10 peggiori su
200)
    r_test2 = np.concatenate([np.full(8, -0.10), np.full(2, -0.02),
np.full(190, 0.001)])
    var_t2, es_t2 = var_es_storico(r_test2, livello=0.95)
    tutti_ok &= _esito("VaR_95 (coda asimmetrica) atteso 2,00%", abs(var_t2 -
0.02) < 1e-9,
                       f"calcolato {var_t2:.4%}")
    es_atteso2 = (8 * 0.10 + 2 * 0.02) / 10  # media dei 10 peggiori
    tutti_ok &= _esito("ES_95 (coda asimmetrica) atteso 8,40%", abs(es_t2 -
es_atteso2) < 1e-9,
                       f"calcolato {es_t2:.4%}")

    # ------------------------------------------------------------
    # DIMOSTRAZIONE FINALE: mini-esperimento completo
    # 3 portafogli su 3 indici sintetici + tabella metriche + grafico
    # ------------------------------------------------------------
    print("\nDIMOSTRAZIONE - Mini-esperimento su 3 indici sintetici")
    portafogli = {
        "Aggressivo (80/10/10)":  {"Indice_A": 80, "Indice_B": 10,
"Indice_C": 10},
        "Bilanciato (50/40/10)":  {"Indice_A": 50, "Indice_B": 40,
"Indice_C": 10},
        "Conservativo (20/70/10)": {"Indice_A": 20, "Indice_B": 70,
"Indice_C": 10},
    }
    serie_valori = {}
    for nome, w in portafogli.items():
        serie_valori[nome], _ = backtest_portafoglio(prezzi, w)
    tab = tabella_metriche(serie_valori, risk_free_annuo=0.02)
    formattata = tab.copy()
    for col in ["CAGR", "Volatilita_ann", "Max_Drawdown", "VaR_95", "ES_95"]:
        formattata[col] = (formattata[col] * 100).round(2).astype(str) + "%"
    formattata["Sharpe"] = tab["Sharpe"].round(3)
    formattata["Calmar"] = tab["Calmar"].round(3)
    print("\n" + formattata.to_string())

    # grafico di prova salvato in ../output
    os.makedirs(CARTELLA_OUTPUT, exist_ok=True)
    fig, ax = plt.subplots(figsize=(10, 5))
    for nome, serie in serie_valori.items():
        ax.plot(serie.index, serie.values, label=nome, linewidth=1.2)
    ax.set_ylabel("Valore (base 100)")
    ax.legend()
    ax.grid(alpha=0.3)
    for con_titolo in (True, False):
        ax.set_title("Collaudo motore di backtest - portafogli su dati
sintetici" if con_titolo else "")
        suffisso = "con_titolo" if con_titolo else "senza_titolo"
        percorso_png = os.path.join(CARTELLA_OUTPUT,
f"collaudo_backtest_{suffisso}.png")
        fig.tight_layout()
```

```
        fig.savefig(percorso_png, dpi=300)
        print(f"Grafico di collaudo salvato in: {percorso_png}")
    plt.close(fig)

    # ------------------------------------------------------------
    print("\n" + "=" * 68)
    if tutti_ok:
        print("ESITO COMPLESSIVO: TUTTI I TEST SUPERATI - motore validato.")
    else:
        print("ESITO COMPLESSIVO: ALMENO UN TEST FALLITO - non usare il
motore.")
    print("=" * 68)
    return tutti_ok


if __name__ == "__main__":
    esegui_test()
```

B.3 - backtest_benchmark.py

Codice sorgente (111 righe).

```
# -*- coding: utf-8 -*-
"""
======================================================================
 BACKTEST DEI BENCHMARK - Tesi: AI generativa e strategie di investimento
======================================================================
 Esegue il backtest dei 4 benchmark congelati (composizione definita
 PRIMA della generazione dei portafogli AI) sul dataset master
 2016-2025, usando il motore validato in backtest.py.

 Benchmark (pesi %, congelati il 16/07/2026) - uno per profilo + 1/N:
   - 80/20 (aggressivo)      : 80 az. / 20 obb.
   - 60/40 (bilanciato)      : 60 az. / 40 obb.
   - 20/80 (conservativo)    : 20 az. / 80 obb.
   - Equal-weight 1/N        : 12,5 su ciascuno degli 8 indici (comune)
 Blocco azionario: 60 SP500 / 25 EAFE / 15 EM (pesi ~mercato mondiale)
 Blocco obbligaz.: 40 T10Y / 30 T2Y / 30 IG   (~aggregate USA)
 Quote azionarie 80/60/20 = profili Growth / Moderate Growth /
 Conservative (Charles Schwab, 2026).
======================================================================
"""

import os
import numpy as np
import pandas as pd
import matplotlib
matplotlib.use("Agg")
import matplotlib.pyplot as plt

from backtest import backtest_portafoglio, calcola_metriche, GIORNI_ANNO
```

```
CARTELLA_OUTPUT = os.path.join("..", "output")
FILE_MASTER = os.path.join(CARTELLA_OUTPUT,
"master_giornaliero_2016_2025.csv")

# ---------------------------------------------------------------
# Benchmark congelati (pesi interi o mezzi punti, somma 100)
# ---------------------------------------------------------------
BENCHMARK = {
    "80/20 aggressivo": {"SP500": 48, "EAFE": 20, "EM": 12, "TREAS_10Y": 8,
"TREAS_2Y": 6,  "IG": 6},
    "60/40 bilanciato": {"SP500": 36, "EAFE": 15, "EM": 9,  "TREAS_10Y": 16,
"TREAS_2Y": 12, "IG": 12},
    "20/80 conservativo": {"SP500": 12, "EAFE": 5, "EM": 3, "TREAS_10Y": 32,
"TREAS_2Y": 24, "IG": 24},
    "Equal-weight 1/N": {"SP500": 12.5, "EAFE": 12.5, "EM": 12.5,
"TREAS_10Y": 12.5,
                         "TREAS_2Y": 12.5, "IG": 12.5, "CORP_HY": 12.5,
"ORO": 12.5},
}


def rf_medio_annuo(master):
    """Risk-free medio del periodo (DTB3 in % annua -> frazione annua).

    Nota metodologica: DTB3 varia molto nel periodo (0% -> 5%); per lo
    Sharpe si usa il tasso medio del periodo come costante. Scelta
    dichiarata nel Capitolo 3; l'ordinamento tra portafogli non ne
    risente perche' identica per tutti.
    """
    return master["RF_DTB3"].mean() / 100.0


def main():
    print("=" * 70)
    print("BACKTEST DEI BENCHMARK (2016-2025, ribilanciamento trimestrale)")
    print("=" * 70)

    master = pd.read_csv(FILE_MASTER, index_col=0, parse_dates=True)
    prezzi = master.drop(columns=["RF_DTB3"])
    rf = rf_medio_annuo(master)
    print(f"\nDataset: {len(prezzi)} giorni, {prezzi.index.min().date()} ->
{prezzi.index.max().date()}")
    print(f"Risk-free medio periodo (DTB3): {rf:.2%}\n")

    serie_valori, tabella = {}, {}
    for nome, pesi in BENCHMARK.items():
        val, _ = backtest_portafoglio(prezzi, pesi)
        serie_valori[nome] = val
        tabella[nome] = calcola_metriche(val, risk_free_annuo=rf)

    tab = pd.DataFrame(tabella).T
    out = tab.copy()
    for c in ("CAGR", "Volatilita_ann", "Max_Drawdown", "VaR_95", "ES_95"):
```

```
        out[c] = (tab[c] * 100).round(2).astype(str) + "%"
    out["Sharpe"] = tab["Sharpe"].round(3)
    out["Calmar"] = tab["Calmar"].round(3)
    print(out.to_string())

    # salvataggi: tabella CSV + grafico equity line
    os.makedirs(CARTELLA_OUTPUT, exist_ok=True)
    p_csv = os.path.join(CARTELLA_OUTPUT, "benchmark_metriche.csv")
    tab.to_csv(p_csv, float_format="%.6f")

    fig, ax = plt.subplots(figsize=(11, 6))
    for nome, s in serie_valori.items():
        ax.plot(s.index, s.values, label=nome, linewidth=1.3)
    ax.set_ylabel("Valore")
    ax.legend(); ax.grid(alpha=0.3)
    for con_titolo in (True, False):
        ax.set_title("Benchmark 2016-2025 (base 100, ribilanciamento
trimestrale)" if con_titolo else "")
        suffisso = "con_titolo" if con_titolo else "senza_titolo"
        p_png = os.path.join(CARTELLA_OUTPUT,
f"benchmark_equity_line_{suffisso}.png")
        fig.tight_layout()
        fig.savefig(p_png, dpi=300)
    plt.close(fig)

    p_val = os.path.join(CARTELLA_OUTPUT, "benchmark_serie_valori.csv")
    pd.DataFrame(serie_valori).to_csv(p_val, float_format="%.6f")

    print(f"\nFile prodotti:\n  {p_csv}")
    print(f"  {os.path.join(CARTELLA_OUTPUT,
'benchmark_equity_line_con_titolo.png')}")
    print(f"  {os.path.join(CARTELLA_OUTPUT,
'benchmark_equity_line_senza_titolo.png')}")
    print(f"  {p_val}")
    print("\nCOMPLETATO: benchmark backtestati con il motore validato.")


if __name__ == "__main__":
    main()
```

B.4 - backtest_portafogli_ai.py

Codice sorgente (225 righe).

```
# -*- coding: utf-8 -*-
"""
======================================================================
 BACKTEST DEI PORTAFOGLI AI - Tesi: AI generativa e strategie di investimento
======================================================================
 Esegue il backtest dei 9 portafogli generati dai tre modelli (ChatGPT,
 Claude, Gemini - 3 profili ciascuno) sul dataset master 2016-2025,
```

```
 usando lo stesso motore validato impiegato per i benchmark.

 I pesi sono trascritti dagli output integrali dei modelli
 (Appendice A), generati a partire dal prompt congelato.

======================================================================
"""

import os
import numpy as np
import pandas as pd
import matplotlib
matplotlib.use("Agg")
import matplotlib.pyplot as plt

from backtest import backtest_portafoglio, calcola_metriche

CARTELLA_OUTPUT = os.path.join("..", "output")
FILE_MASTER = os.path.join(CARTELLA_OUTPUT,
"master_giornaliero_2016_2025.csv")
FILE_BENCH_SERIE = os.path.join(CARTELLA_OUTPUT,
"benchmark_serie_valori.csv")

# ----------------------------------------------------------------
# Pesi dagli output dei modelli (20/07/2026)
# ----------------------------------------------------------------
PORTAFOGLI_AI = {
    "ChatGPT - Aggressivo":   {"SP500": 45, "EAFE": 20, "EM": 20,
"TREAS_10Y": 5,  "TREAS_2Y": 0,  "IG": 0,  "CORP_HY": 5,  "ORO": 5},
    "ChatGPT - Bilanciato":   {"SP500": 30, "EAFE": 15, "EM": 10,
"TREAS_10Y": 15, "TREAS_2Y": 5,  "IG": 20, "CORP_HY": 0,  "ORO": 5},
    "ChatGPT - Conservativo": {"SP500": 15, "EAFE": 5,  "EM": 0,
"TREAS_10Y": 20, "TREAS_2Y": 20, "IG": 35, "CORP_HY": 0,  "ORO": 5},

    "Claude - Aggressivo":    {"SP500": 45, "EAFE": 20, "EM": 15,
"TREAS_10Y": 5,  "TREAS_2Y": 0,  "IG": 0,  "CORP_HY": 10, "ORO": 5},
    "Claude - Bilanciato":    {"SP500": 30, "EAFE": 15, "EM": 10,
"TREAS_10Y": 15, "TREAS_2Y": 5,  "IG": 15, "CORP_HY": 5,  "ORO": 5},
    "Claude - Conservativo":  {"SP500": 15, "EAFE": 5,  "EM": 0,
"TREAS_10Y": 15, "TREAS_2Y": 25, "IG": 30, "CORP_HY": 5,  "ORO": 5},

    "Gemini - Aggressivo":    {"SP500": 50, "EAFE": 25, "EM": 15,
"TREAS_10Y": 0,  "TREAS_2Y": 0,  "IG": 0,  "CORP_HY": 10, "ORO": 0},
    "Gemini - Bilanciato":    {"SP500": 30, "EAFE": 15, "EM": 5,
"TREAS_10Y": 15, "TREAS_2Y": 0,  "IG": 30, "CORP_HY": 0,  "ORO": 5},
    "Gemini - Conservativo":  {"SP500": 15, "EAFE": 5,  "EM": 0,
"TREAS_10Y": 10, "TREAS_2Y": 35, "IG": 30, "CORP_HY": 0,  "ORO": 5},
}


COLORI_MODELLO = {"ChatGPT": "#1f77b4", "Claude": "#ff7f0e", "Gemini":
"#2ca02c"}
```

```
def grafico_valore_cumulato_ai(serie_valori, bench, bench_map,
con_titolo=True):

    """Figura 4.3: valore cumulato dei portafogli AI in disposizione 2+1.
    """
    print("\nGrafico valore cumulato AI: layout 2 sopra + 1 centrato
sotto...")

    fig = plt.figure(figsize=(11, 8.5))
    gs = fig.add_gridspec(nrows=2, ncols=4, height_ratios=[1, 1],
hspace=0.32, wspace=0.28)

    assi = {
        "Aggressivo": fig.add_subplot(gs[0, 0:2]),
        "Bilanciato": fig.add_subplot(gs[0, 2:4]),
        "Conservativo": fig.add_subplot(gs[1, 1:3]),
    }

    for profilo, ax in assi.items():
        for modello in ["ChatGPT", "Claude", "Gemini"]:
            nome = f"{modello} - {profilo}"
            ax.plot(serie_valori[nome].index, serie_valori[nome].values,
                    label=modello, linewidth=1.35,
color=COLORI_MODELLO[modello])

        nome_benchmark = bench_map[profilo]
        ax.plot(bench.index, bench[nome_benchmark], label="Benchmark
profilo",
                color="dimgray", linewidth=1.2, linestyle="--")
        ax.plot(bench.index, bench["Equal-weight 1/N"], label="1/N",
                color="black", linewidth=1.1, linestyle=":")

        ax.set_title(profilo, fontsize=11, fontweight="bold")
        ax.grid(alpha=0.3)
        ax.tick_params(axis="both", labelsize=8)
        ax.legend(fontsize=7.5, loc="upper left")

    assi["Aggressivo"].set_ylabel("Valore cumulato (base 100)")
    assi["Conservativo"].set_ylabel("Valore cumulato (base 100)")

    if con_titolo:
        fig.suptitle("Valore cumulato dei portafogli AI, per profilo, "
                     "confrontato con benchmark ed equal-weight (2016-
2025)",
                     fontsize=13, y=0.98)

    fig.subplots_adjust(left=0.08, right=0.98, bottom=0.08, top=0.92 if
con_titolo else 0.97)

    suffisso = "con_titolo" if con_titolo else "senza_titolo"
    p = os.path.join(CARTELLA_OUTPUT,
f"valore_cumulato_ai_layout_2x1_{suffisso}.png")
    fig.savefig(p, dpi=300, bbox_inches="tight")
    plt.close(fig)
```

```
    print(f"   salvato: {p}")
    return p


def verifica_formale(nome, pesi):
    """Controlli di validita' formale richiesti dal protocollo (par. 3.3 /
3.7 progetto)."""
    somma = sum(pesi.values())
    interi = all(float(w).is_integer() for w in pesi.values())
    non_negativi = all(w >= 0 for w in pesi.values())
    ok = abs(somma - 100) < 1e-9 and interi and non_negativi
    stato = "OK" if ok else "NON VALIDO"
    print(f"  [{stato}] {nome:<24} somma={somma:.0f}  interi={interi}  non-
negativi={non_negativi}")
    return ok


def main():
    print("=" * 78)
    print("VERIFICA FORMALE DEI 9 PORTAFOGLI AI (prima del backtest)")
    print("=" * 78)
    for nome, pesi in PORTAFOGLI_AI.items():
        if not verifica_formale(nome, pesi):
            raise ValueError(f"Portafoglio non valido: {nome} - verificare
prima di procedere")
    print("\nTutti i 9 portafogli superano la verifica formale: nessuna
esclusione necessaria.\n")

    master = pd.read_csv(FILE_MASTER, index_col=0, parse_dates=True)
    prezzi = master.drop(columns=["RF_DTB3"])
    rf = master["RF_DTB3"].mean() / 100.0

    print("=" * 78)
    print("BACKTEST (2016-2025, ribilanciamento trimestrale)")
    print("=" * 78)
    serie_valori, tabella, pesi_tab = {}, {}, {}
    for nome, pesi in PORTAFOGLI_AI.items():
        val, _ = backtest_portafoglio(prezzi, pesi)
        serie_valori[nome] = val
        tabella[nome] = calcola_metriche(val, risk_free_annuo=rf)
        pesi_tab[nome] = pesi

    tab = pd.DataFrame(tabella).T
    out = tab.copy()
    for c in ("CAGR", "Volatilita_ann", "Max_Drawdown", "VaR_95", "ES_95"):
        out[c] = (tab[c] * 100).round(2).astype(str) + "%"
    out["Sharpe"] = tab["Sharpe"].round(3)
    out["Calmar"] = tab["Calmar"].round(3)
    print(out.to_string())

    # ------------------------------------------------------------
    # Confronto con i benchmark di profilo
    # ------------------------------------------------------------
    if os.path.exists(FILE_BENCH_SERIE):
```

```
        bench = pd.read_csv(FILE_BENCH_SERIE, index_col=0, parse_dates=True)
        mappa_profilo = {"Aggressivo": "80/20 aggressivo", "Bilanciato":
"60/40 bilanciato",
                         "Conservativo": "20/80 conservativo"}
        print("\n" + "=" * 78)
        print("CONFRONTO CON BENCHMARK DI PROFILO E EQUAL-WEIGHT (Sharpe
ratio)")
        print("=" * 78)
        sh_ew = calcola_metriche(bench["Equal-weight 1/N"],
risk_free_annuo=rf)["Sharpe"]
        righe = []
        for nome, m in tabella.items():
            profilo = nome.split(" - ")[1]
            bmk = mappa_profilo[profilo]
            sh_bmk = calcola_metriche(bench[bmk],
risk_free_annuo=rf)["Sharpe"]
            righe.append({"Portafoglio": nome, "Sharpe_AI":
round(m["Sharpe"], 3),
                          "Benchmark_profilo": bmk, "Sharpe_benchmark":
round(sh_bmk, 3),
                          "Sharpe_1/N": round(sh_ew, 3)})
        conf = pd.DataFrame(righe).set_index("Portafoglio")
        print(conf.to_string())

    # ------------------------------------------------------------
    # Tabella pesi 9x8 (per il capitolo 4 e per l'analisi di coerenza)
    # ------------------------------------------------------------
    pesi_df = pd.DataFrame(pesi_tab).T

    # ------------------------------------------------------------
    # Salvataggi
    # ------------------------------------------------------------
    os.makedirs(CARTELLA_OUTPUT, exist_ok=True)
    p_metriche = os.path.join(CARTELLA_OUTPUT, "ai_metriche.csv")
    p_pesi = os.path.join(CARTELLA_OUTPUT, "ai_pesi.csv")
    p_val = os.path.join(CARTELLA_OUTPUT, "ai_serie_valori.csv")
    tab.to_csv(p_metriche, float_format="%.6f")
    pesi_df.to_csv(p_pesi)
    pd.DataFrame(serie_valori).to_csv(p_val, float_format="%.6f")

    # grafico combinato: layout 2 pannelli sopra (Aggressivo, Bilanciato) + 1
centrato
    # sotto (Conservativo)
    profili = ["Aggressivo", "Bilanciato", "Conservativo"]
    bench_map = {"Aggressivo": "80/20 aggressivo", "Bilanciato": "60/40
bilanciato",
                 "Conservativo": "20/80 conservativo"}
    bench = pd.read_csv(FILE_BENCH_SERIE, index_col=0, parse_dates=True) if
os.path.exists(FILE_BENCH_SERIE) else None
    file_combinato = {}
    if bench is not None:
        for con_titolo in (False, True):
            percorso = grafico_valore_cumulato_ai(serie_valori, bench,
bench_map, con_titolo=con_titolo)
```

```
            suffisso = "con_titolo" if con_titolo else "senza_titolo"
            file_combinato[suffisso] = percorso

    # grafici singoli, uno per profilo: stesso contenuto del pannello, ma
piu' grandi
    file_singoli = {}
    for profilo in profili:
        fig_s, ax_s = plt.subplots(figsize=(11, 7))
        for modello in ["ChatGPT", "Claude", "Gemini"]:
            nome = f"{modello} - {profilo}"
            ax_s.plot(serie_valori[nome].index, serie_valori[nome].values,
label=modello, linewidth=1.6)
        if bench is not None:
            ax_s.plot(bench.index, bench[bench_map[profilo]],
label="Benchmark profilo", color="black", linewidth=1.3, linestyle="--")
            ax_s.plot(bench.index, bench["Equal-weight 1/N"], label="1/N",
color="gray", linewidth=1.3, linestyle=":")
        ax_s.set_ylabel("Valore (base 100)")
        ax_s.grid(alpha=0.3)
        ax_s.legend(fontsize=10)
        for con_titolo in (True, False):
            ax_s.set_title(f"Profilo {profilo} \u2013 Portafogli AI vs
benchmark (2016-2025)" if con_titolo else "", fontsize=13)
            suffisso = "con_titolo" if con_titolo else "senza_titolo"
            nome_file = f"ai_equity_line_{profilo.lower()}_{suffisso}.png"
            p_singolo = os.path.join(CARTELLA_OUTPUT, nome_file)
            fig_s.tight_layout()
            fig_s.savefig(p_singolo, dpi=300)
            file_singoli[f"{profilo} ({suffisso})"] = p_singolo
        plt.close(fig_s)

    print(f"\nFile prodotti:\n  {p_metriche}\n  {p_pesi}\n  {p_val}")
    for suffisso, percorso in file_combinato.items():
        print(f"  {percorso}")
    for profilo, percorso in file_singoli.items():
        print(f"  {percorso}")
    print("\nCOMPLETATO: backtest dei 9 portafogli AI.")


if __name__ == "__main__":
    main()
```

B.5 - grafici.py

Codice sorgente (436 righe).

```
# -*- coding: utf-8 -*-
"""
======================================================================
 GRAFICI PER IL CAPITOLO 4 - Tesi: AI generativa e strategie di investimento
======================================================================
```

```
Produce cinque gruppi di grafici a partire dagli output gia' calcolati
dagli script precedenti:

  1. Drawdown nel tempo - un grafico per profilo, in DUE varianti:
     - "diviso": tre pannelli separati (uno per modello);
     - "unico": un solo pannello con tutte le linee insieme, piu'
       compatto.
  2. Mappa di calore dei pesi - i 9 portafogli AI (righe) sugli 8
     indici (colonne), per leggere a colpo d'occhio le similarità o
     dissimilarità tra i tre modelli.
  3. Confronto Sharpe ratio - grafico a barre dei 13 portafogli
     (9 AI + 4 benchmark), raggruppati per profilo.
  4. Confronto Max Drawdown - stessa struttura del grafico 3, sul
     massimo drawdown anziche' sullo Sharpe Ratio.
  5. Esposizione aggregata per asset class macro - gli otto indici
     raggruppati in quattro macro categorie (Azionario, Governativo,
     Credito, Oro), con l'indice di Herfindahl-Hirschman (HHI)
     annotato sopra ciascuna barra.
======================================================================
"""

import os
import numpy as np
import pandas as pd
import matplotlib
matplotlib.use("Agg")
import matplotlib.pyplot as plt
from matplotlib.patches import Patch

CARTELLA_OUTPUT = os.path.join("..", "output")

PROFILI = ["Aggressivo", "Bilanciato", "Conservativo"]
BENCH_DI_PROFILO = {"Aggressivo": "80/20 aggressivo", "Bilanciato": "60/40
bilanciato",
                    "Conservativo": "20/80 conservativo"}
COLORI_MODELLO = {"ChatGPT": "#1f77b4", "Claude": "#ff7f0e", "Gemini":
"#2ca02c"}

# ----------------------------------------------------------------------
# Pesi per il grafico 5 (Tabella 3.2 e Appendice B.4) - non
# derivano dai CSV di output, a differenza dei grafici 1-4.
# ----------------------------------------------------------------------
BENCHMARK = {
    "80/20": {"SP500": 48, "EAFE": 20, "EM": 12, "TREAS_10Y": 8, "TREAS_2Y":
6, "IG": 6, "CORP_HY": 0, "ORO": 0},
    "60/40": {"SP500": 36, "EAFE": 15, "EM": 9, "TREAS_10Y": 16, "TREAS_2Y":
12, "IG": 12, "CORP_HY": 0, "ORO": 0},
    "20/80": {"SP500": 12, "EAFE": 5, "EM": 3, "TREAS_10Y": 32, "TREAS_2Y":
24, "IG": 24, "CORP_HY": 0, "ORO": 0},
    "1/N": {"SP500": 12.5, "EAFE": 12.5, "EM": 12.5, "TREAS_10Y": 12.5,
"TREAS_2Y": 12.5, "IG": 12.5, "CORP_HY": 12.5, "ORO": 12.5},
}

PORTAFOGLI_AI = {
```

```
    "ChatGPT - Aggressivo": {"SP500": 45, "EAFE": 20, "EM": 20, "TREAS_10Y":
5, "TREAS_2Y": 0, "IG": 0, "CORP_HY": 5, "ORO": 5},
    "Claude - Aggressivo": {"SP500": 45, "EAFE": 20, "EM": 15, "TREAS_10Y":
5, "TREAS_2Y": 0, "IG": 0, "CORP_HY": 10, "ORO": 5},
    "Gemini - Aggressivo": {"SP500": 50, "EAFE": 25, "EM": 15, "TREAS_10Y":
0, "TREAS_2Y": 0, "IG": 0, "CORP_HY": 10, "ORO": 0},
    "ChatGPT - Bilanciato": {"SP500": 30, "EAFE": 15, "EM": 10, "TREAS_10Y":
15, "TREAS_2Y": 5, "IG": 20, "CORP_HY": 0, "ORO": 5},
    "Claude - Bilanciato": {"SP500": 30, "EAFE": 15, "EM": 10, "TREAS_10Y":
15, "TREAS_2Y": 5, "IG": 15, "CORP_HY": 5, "ORO": 5},
    "Gemini - Bilanciato": {"SP500": 30, "EAFE": 15, "EM": 5, "TREAS_10Y":
15, "TREAS_2Y": 0, "IG": 30, "CORP_HY": 0, "ORO": 5},
    "ChatGPT - Conservativo": {"SP500": 15, "EAFE": 5, "EM": 0, "TREAS_10Y":
20, "TREAS_2Y": 20, "IG": 35, "CORP_HY": 0, "ORO": 5},
    "Claude - Conservativo": {"SP500": 15, "EAFE": 5, "EM": 0, "TREAS_10Y":
15, "TREAS_2Y": 25, "IG": 30, "CORP_HY": 5, "ORO": 5},
    "Gemini - Conservativo": {"SP500": 15, "EAFE": 5, "EM": 0, "TREAS_10Y":
10, "TREAS_2Y": 35, "IG": 30, "CORP_HY": 0, "ORO": 5},
}

COLORI_CATEGORIA = {
    "Azionario": "#4472C4",
    "Governativo": "#70ad47",
    "Credito": "#ed7d31",
    "Oro": "#ffc000",
}


def drawdown(serie_valore):
    """Serie di drawdown (%) dal picco precedente, punto per punto."""
    massimo = serie_valore.cummax()
    return serie_valore / massimo - 1.0


def grafico_drawdown_diviso(ai_val, bench_val, con_titolo=True):
    """Versione a 3 pannelli (uno per modello): meno linee sovrapposte."""
    print("1a) Grafici di drawdown DIVISI per profilo (un pannello per
modello)...")
    for profilo in PROFILI:
        dd_bmk = drawdown(bench_val[BENCH_DI_PROFILO[profilo]]) * 100
        dd_ew = drawdown(bench_val["Equal-weight 1/N"]) * 100
        y_min = min(dd_bmk.min(), dd_ew.min())

        fig, axes = plt.subplots(3, 1, figsize=(11, 10), sharex=True,
sharey=True)
        for ax, modello in zip(axes, ["ChatGPT", "Claude", "Gemini"]):
            nome = f"{modello} - {profilo}"
            dd_mod = drawdown(ai_val[nome]) * 100
            y_min = min(y_min, dd_mod.min())

            ax.fill_between(dd_bmk.index, dd_bmk.values, 0, color="black",
alpha=0.06,
                            label="Benchmark profilo (area)")
```

```
            ax.plot(dd_ew.index, dd_ew.values, color="gray", linewidth=1,
linestyle=":", label="1/N")
            ax.plot(dd_mod.index, dd_mod.values,
color=COLORI_MODELLO[modello], linewidth=1.3, label=modello)

            ax.set_title(modello, fontsize=11, loc="left", fontweight="bold")
            ax.axhline(0, color="black", linewidth=0.6)
            ax.grid(alpha=0.3)
            ax.set_ylabel("Drawdown (%)")
            ax.legend(fontsize=8, loc="lower left", ncol=3)

        for ax in axes:
            ax.set_ylim(y_min * 1.05, 2)
        if con_titolo:
            fig.suptitle(f"Profilo {profilo} \u2013 Drawdown dal picco
precedente, per modello (2016-2025)", fontsize=13)
        fig.tight_layout()
        suffisso = "con_titolo" if con_titolo else "senza_titolo"
        p = os.path.join(CARTELLA_OUTPUT,
f"drawdown_{profilo.lower()}_diviso_{suffisso}.png")
        fig.savefig(p, dpi=300, bbox_inches='tight')
        plt.close(fig)
        print(f"   salvato: {p}")


def grafico_drawdown_unico(ai_val, bench_val, con_titolo=True):
    """Versione compatta: un solo pannello per profilo, tutte le linee
insieme.
    """
    print("\n1b) Grafici di drawdown UNICI per profilo (un solo
pannello)...")
    for profilo in PROFILI:
        fig, ax = plt.subplots(figsize=(11, 6))
        for modello in ["ChatGPT", "Claude", "Gemini"]:
            nome = f"{modello} - {profilo}"
            dd = drawdown(ai_val[nome]) * 100
            ax.plot(dd.index, dd.values, label=modello, linewidth=1.3,
color=COLORI_MODELLO[modello])
        dd_bmk = drawdown(bench_val[BENCH_DI_PROFILO[profilo]]) * 100
        ax.plot(dd_bmk.index, dd_bmk.values, label="Benchmark profilo",
color="black",
                linewidth=1.2, linestyle="--")
        dd_ew = drawdown(bench_val["Equal-weight 1/N"]) * 100
        ax.plot(dd_ew.index, dd_ew.values, label="1/N", color="gray",
linewidth=1.2, linestyle=":")

        ax.fill_between(dd_bmk.index, dd_bmk.values, 0, color="black",
alpha=0.03)
        if con_titolo:
            ax.set_title(f"Profilo {profilo} \u2013 Drawdown dal picco
precedente (2016-2025)", fontsize=13)
        ax.set_ylabel("Drawdown (%)")
        ax.axhline(0, color="black", linewidth=0.6)
        ax.grid(alpha=0.3)
```

```
        ax.legend(fontsize=9, loc="lower left")
        fig.tight_layout()
        suffisso = "con_titolo" if con_titolo else "senza_titolo"
        p = os.path.join(CARTELLA_OUTPUT,
f"drawdown_{profilo.lower()}_unico_{suffisso}.png")
        fig.savefig(p, dpi=300, bbox_inches='tight')
        plt.close(fig)
        print(f"   salvato: {p}")


def grafico_heatmap_pesi(pesi_df, con_titolo=True):
    print("\n2) Mappa di calore dei pesi (9 portafogli AI x 8 indici)...")
    colonne_ordine = ["SP500", "EAFE", "EM", "TREAS_10Y", "TREAS_2Y", "IG",
"CORP_HY", "ORO"]
    etichette_asset = ["S&P 500", "EAFE", "EM", "Treasury 10Y", "Treasury
2Y",
                       "Investment Grade", "Corp. HY", "Oro"]

    ordine_righe = [f"{modello} - {profilo}" for profilo in PROFILI
                    for modello in ["ChatGPT", "Claude", "Gemini"]]
    dati = pesi_df.loc[ordine_righe, colonne_ordine]

    fig, ax = plt.subplots(figsize=(10.5, 8.0))
    im = ax.imshow(dati.values, cmap="YlOrRd", aspect="auto", vmin=0,
vmax=dati.values.max())

    # Etichette delle colonne spostate in alto (piu' leggibile: si leggono
    # subito insieme al titolo, senza dover scorrere fino in fondo)
    ax.xaxis.tick_top()
    ax.xaxis.set_label_position("top")
    ax.set_xticks(range(len(colonne_ordine)))
    ax.set_xticklabels(etichette_asset, rotation=30, ha="left",
fontweight="bold")

    etichette_righe = ["ChatGPT", "Claude", "Gemini"] * len(PROFILI)
    ax.set_yticks(range(len(dati.index)))
    ax.set_yticklabels(etichette_righe, fontweight="bold")

    for i in range(dati.shape[0]):
        for j in range(dati.shape[1]):
            v = dati.values[i, j]
            colore_testo = "white" if v > dati.values.max() * 0.55 else
"black"
            ax.text(j, i, f"{v:.0f}", ha="center", va="center", fontsize=9,
color=colore_testo)

    for y in (2.5, 5.5):
        ax.axhline(y, color="black", linewidth=1.2)

    for centro, profilo in zip((1, 4, 7), PROFILI):
        ax.text(-0.24, 1 - (centro + 0.5) / 9, profilo,
transform=ax.transAxes,
                ha="center", va="center", fontsize=10, fontweight="bold",
rotation=90)
```

```
    if con_titolo:
        # pad maggiore: con le etichette delle colonne in alto serve piu'
        # spazio tra il titolo e la figura
        ax.set_title("Pesi assegnati dai nove portafogli AI (%), raggruppati
per profilo di rischio",
                     fontsize=12, pad=75)
    # colorbar molto vicina alla heatmap (pad ridotto rispetto all'originale
0.22)
    fig.colorbar(im, ax=ax, label="Peso (%)", orientation="horizontal",
fraction=0.045, pad=0.02, aspect=35)
    fig.subplots_adjust(left=0.26, bottom=0.04, right=0.96, top=0.72 if
con_titolo else 0.80)
    suffisso = "con_titolo" if con_titolo else "senza_titolo"
    p = os.path.join(CARTELLA_OUTPUT, f"heatmap_pesi_ai_{suffisso}.png")
    fig.savefig(p, dpi=300, bbox_inches='tight')
    plt.close(fig)
    print(f"   salvato: {p}")


def grafico_sharpe_confronto(ai_metriche, bench_metriche, con_titolo=True):
    print("\n3) Confronto Sharpe ratio (9 portafogli AI + 4 benchmark)...")
    fig, ax = plt.subplots(figsize=(12, 6.5))

    posizione = 0
    xticks, xlabels = [], []
    larghezza = 0.7
    colori_barre = []
    valori = []
    etichette = []

    for profilo in PROFILI:
        for modello in ["ChatGPT", "Claude", "Gemini"]:
            nome = f"{modello} - {profilo}"
            valori.append(ai_metriche.loc[nome, "Sharpe"])
            etichette.append(modello)
            colori_barre.append(COLORI_MODELLO[modello])
            xticks.append(posizione)
            posizione += 1
        valori.append(bench_metriche.loc[BENCH_DI_PROFILO[profilo],
"Sharpe"])
        etichette.append("Benchmark")
        colori_barre.append("dimgray")
        xticks.append(posizione)
        posizione += 1.8

    barre = ax.bar(xticks, valori, width=larghezza, color=colori_barre)
    ax.set_xticks([])

    legend_handles = [Patch(facecolor=COLORI_MODELLO[m], label=m) for m in
["ChatGPT", "Claude", "Gemini"]]
    legend_handles.append(Patch(facecolor="dimgray", label="Benchmark di
profilo"))
```

```
    ax.legend(handles=legend_handles, loc="upper center",
bbox_to_anchor=(0.5, -0.05),
              ncol=4, fontsize=9, frameon=False)

    sh_ew = bench_metriche.loc["Equal-weight 1/N", "Sharpe"]
    ax.axhline(sh_ew, color="black", linewidth=1.2, linestyle=":")
    ax.text(xticks[-1], sh_ew + 0.01, f"Equal-weight 1/N ({sh_ew:.3f})",
            ha="right", va="bottom", fontsize=9)

    centri = [1, 5.8, 10.6]
    for centro, profilo in zip(centri, PROFILI):
        ax.text(centro, max(valori) * 1.06, profilo, ha="center",
fontsize=11, fontweight="bold")

    ax.set_ylabel("Indice di Sharpe")
    if con_titolo:
        ax.set_title("Indice di Sharpe: portafogli AI, benchmark di profilo
ed equal-weight", fontsize=13)
    ax.grid(alpha=0.3, axis="y")
    fig.tight_layout()
    suffisso = "con_titolo" if con_titolo else "senza_titolo"
    p = os.path.join(CARTELLA_OUTPUT, f"sharpe_confronto_{suffisso}.png")
    fig.savefig(p, dpi=300, bbox_inches='tight')
    plt.close(fig)
    print(f"   salvato: {p}")


def grafico_maxdd_confronto(ai_metriche, bench_metriche, con_titolo=True):
    print("\n4) Confronto Max Drawdown (9 portafogli AI + 4 benchmark)...")
    fig, ax = plt.subplots(figsize=(12, 6.5))

    posizione = 0
    xticks = []
    larghezza = 0.7
    colori_barre = []
    valori = []
    etichette = []

    for profilo in PROFILI:
        for modello in ["ChatGPT", "Claude", "Gemini"]:
            nome = f"{modello} - {profilo}"
            valori.append(ai_metriche.loc[nome, "Max_Drawdown"] * 100)
            etichette.append(modello)
            colori_barre.append(COLORI_MODELLO[modello])
            xticks.append(posizione)
            posizione += 1
        valori.append(bench_metriche.loc[BENCH_DI_PROFILO[profilo],
"Max_Drawdown"] * 100)
        etichette.append("Benchmark")
        colori_barre.append("dimgray")
        xticks.append(posizione)
        posizione += 1.8

    ax.bar(xticks, valori, width=larghezza, color=colori_barre)
```

```
    ax.set_xticks([])

    legend_handles = [Patch(facecolor=COLORI_MODELLO[m], label=m) for m in
["ChatGPT", "Claude", "Gemini"]]
    legend_handles.append(Patch(facecolor="dimgray", label="Benchmark di
profilo"))
    ax.legend(handles=legend_handles, loc="upper center",
bbox_to_anchor=(0.5, -0.06),
              ncol=4, fontsize=9, frameon=False)

    dd_ew = bench_metriche.loc["Equal-weight 1/N", "Max_Drawdown"] * 100
    ax.axhline(dd_ew, color="black", linewidth=1.2, linestyle=":")
    ax.text(xticks[-1], dd_ew - 0.5, f"Equal-weight 1/N ({dd_ew:.1f}%)",
            ha="right", va="top", fontsize=9)

    centri = [1, 5.8, 10.6]
    for centro, profilo in zip(centri, PROFILI):
        ax.text(centro, 1.2, profilo, ha="center", va="bottom", fontsize=11,
fontweight="bold")

    ax.set_ylabel("Maximum Drawdown (%)")
    ax.axhline(0, color="black", linewidth=0.6)
    ax.set_ylim(min(valori) * 1.08, 4.5)
    if con_titolo:
        ax.set_title("Maximum Drawdown: portafogli AI, benchmark di profilo
ed equal-weight", fontsize=13)
    ax.grid(alpha=0.3, axis="y")
    fig.tight_layout()
    suffisso = "con_titolo" if con_titolo else "senza_titolo"
    p = os.path.join(CARTELLA_OUTPUT, f"maxdd_confronto_{suffisso}.png")
    fig.savefig(p, dpi=300, bbox_inches='tight')
    plt.close(fig)
    print(f"   salvato: {p}")


def hhi(pesi):
    """Indice di Herfindahl-Hirschman calcolato sugli otto indici originali
    (non aggregati): somma dei quadrati dei pesi (in quota, non percentuale).
    Valori piu' bassi indicano maggiore diversificazione."""
    return sum((w / 100) ** 2 for w in pesi.values())


def aggrega(pesi):
    """Riaggrega i pesi sui singoli indici nelle
    quattro macro categorie."""
    return {
        "Azionario": pesi["SP500"] + pesi["EAFE"] + pesi["EM"],
        "Governativo": pesi["TREAS_10Y"] + pesi["TREAS_2Y"],
        "Credito": pesi["IG"] + pesi["CORP_HY"],
        "Oro": pesi["ORO"],
    }


def grafico_esposizioni_aggregate(con_titolo=True):
```

```
    """Grafico 5: esposizione aggregata per macro categoria (Azionario,
    Governativo, Credito, Oro) con indice di Herfindahl-Hirschman (HHI)
    annotato sopra ciascuna barra."""
    print("\n5) Esposizione aggregata per asset class macro (con HHI)...")

    ordine = [
        ("80/20 (bench.)", BENCHMARK["80/20"]),
        ("ChatGPT", PORTAFOGLI_AI["ChatGPT - Aggressivo"]),
        ("Claude", PORTAFOGLI_AI["Claude - Aggressivo"]),
        ("Gemini", PORTAFOGLI_AI["Gemini - Aggressivo"]),
        ("60/40 (bench.)", BENCHMARK["60/40"]),
        ("ChatGPT ", PORTAFOGLI_AI["ChatGPT - Bilanciato"]),
        ("Claude ", PORTAFOGLI_AI["Claude - Bilanciato"]),
        ("Gemini ", PORTAFOGLI_AI["Gemini - Bilanciato"]),
        ("20/80 (bench.)", BENCHMARK["20/80"]),
        ("ChatGPT  ", PORTAFOGLI_AI["ChatGPT - Conservativo"]),
        ("Claude  ", PORTAFOGLI_AI["Claude - Conservativo"]),
        ("Gemini  ", PORTAFOGLI_AI["Gemini - Conservativo"]),
        ("1/N (comune)", BENCHMARK["1/N"]),
    ]

    etichette = [nome for nome, _ in ordine]
    aggregati = [aggrega(pesi) for _, pesi in ordine]
    indici_benchmark = {0, 4, 8, 12}

    categorie = ["Azionario", "Governativo", "Credito", "Oro"]
    x = np.arange(len(etichette))

    fig, ax = plt.subplots(figsize=(10.8, 6.2))

    for i in indici_benchmark:
        ax.axvspan(i - 0.40, i + 0.40, color="#e8e8e8", zorder=0)

    base = np.zeros(len(etichette))
    for cat in categorie:
        valori = np.array([a[cat] for a in aggregati])
        edgecolors = ["black" if i in indici_benchmark else "white" for i in
range(len(etichette))]
        linewidths = [1.6 if i in indici_benchmark else 0.5 for i in
range(len(etichette))]
        for xi, v, b, ec, lw in zip(x, valori, base, edgecolors, linewidths):
            ax.bar(xi, v, bottom=b, width=0.62, color=COLORI_CATEGORIA[cat],
                   edgecolor=ec, linewidth=lw, zorder=2)
            if v >= 6:
                ax.text(xi, b + v / 2, f"{v:.0f}", ha="center", va="center",
                        fontsize=8, color="white", fontweight="bold",
zorder=3)
        base += valori
    for cat in categorie:
        ax.bar(0, 0, color=COLORI_CATEGORIA[cat], label=cat)

    ax.set_xticks(x)
    ax.set_xlim(-0.55, len(etichette) - 1 + 0.55)
    ax.set_xticklabels(etichette, rotation=45, ha="right", fontsize=9)
```

```
    for i, tick_label in enumerate(ax.get_xticklabels()):
        if i in indici_benchmark:
            tick_label.set_fontweight("bold")
    ax.set_ylabel("Peso nel portafoglio (%)")
    ax.set_ylim(0, 132)
    ax.set_yticks(range(0, 101, 20))
    ax.legend(loc="upper center", bbox_to_anchor=(0.5, 1.10), ncol=4,
fontsize=10, frameon=False)
    ax.set_axisbelow(True)
    ax.grid(axis="y", alpha=0.3)
    ax.spines["top"].set_visible(False)

    for pos in (3.5, 7.5, 11.5):
        ax.axvline(pos, color="grey", linewidth=0.8, linestyle="--",
alpha=0.6)
    centri_profilo = {"Aggressivo": 1.5, "Bilanciato": 5.5, "Conservativo":
9.5}
    for profilo, centro in centri_profilo.items():
        ax.text(centro, 127, profilo, ha="center", fontsize=10,
fontweight="bold")

    for xi, (nome, pesi) in zip(x, ordine):
        valore_hhi = hhi(pesi)
        ax.text(xi, 103, f"HHI {valore_hhi:.2f}".replace(".", ","),
rotation=90,
                ha="center", va="bottom", fontsize=7.8, color="#444444")

    if con_titolo:
        ax.set_title("Esposizione aggregata per asset class macro: benchmark
e portafogli AI", fontsize=12)
    fig.tight_layout()
    suffisso = "con_titolo" if con_titolo else "senza_titolo"
    p = os.path.join(CARTELLA_OUTPUT,
f"esposizioni_aggregate_{suffisso}.png")
    fig.savefig(p, dpi=300, bbox_inches="tight", pad_inches=0.05)
    plt.close(fig)
    print(f"   salvato: {p}")


def main():
    print("=" * 70)
    print("GRAFICI PER IL CAPITOLO 4")
    print("=" * 70)

    ai_val = pd.read_csv(os.path.join(CARTELLA_OUTPUT,
"ai_serie_valori.csv"), index_col=0, parse_dates=True)
    bench_val = pd.read_csv(os.path.join(CARTELLA_OUTPUT,
"benchmark_serie_valori.csv"), index_col=0, parse_dates=True)
    ai_pesi = pd.read_csv(os.path.join(CARTELLA_OUTPUT, "ai_pesi.csv"),
index_col=0)
    ai_metriche = pd.read_csv(os.path.join(CARTELLA_OUTPUT,
"ai_metriche.csv"), index_col=0)
    bench_metriche = pd.read_csv(os.path.join(CARTELLA_OUTPUT,
"benchmark_metriche.csv"), index_col=0)
```

```
    for con_titolo in (True, False):
        etichetta = "CON titolo" if con_titolo else "SENZA titolo (per la
tesi)"
        print(f"\n{'='*70}\nVersione {etichetta}\n{'='*70}")
        grafico_drawdown_diviso(ai_val, bench_val, con_titolo=con_titolo)
        grafico_drawdown_unico(ai_val, bench_val, con_titolo=con_titolo)
        grafico_heatmap_pesi(ai_pesi, con_titolo=con_titolo)
        grafico_sharpe_confronto(ai_metriche, bench_metriche,
con_titolo=con_titolo)
        grafico_maxdd_confronto(ai_metriche, bench_metriche,
con_titolo=con_titolo)
        grafico_esposizioni_aggregate(con_titolo=con_titolo)

    print("\nCOMPLETATO: grafici del Capitolo 4 salvati in ../output.")


if __name__ == "__main__":
    main()
```

B.6 - backtest_rigenerazione.py

Codice sorgente (156 righe).

```
# -*- coding: utf-8 -*-
"""
======================================================================
 BACKTEST DELLE RIGENERAZIONI - Tesi: AI generativa e strategie di
investimento
======================================================================
 Esegue il backtest dei 18 portafogli ottenuti rigenerando i tre
 modelli con il medesimo prompt congelato (Tabella 3.3), in due
 varianti linguistiche: italiano (identico all'originale) e inglese
 (traduzione dello stesso prompt). Usa lo stesso motore di backtest
 gia' validato in backtest.py, sullo stesso dataset 2016-2025.

 I pesi sono trascritti dagli output integrali dei modelli
 (Appendice A.4-A.9), generati a partire dallo stesso prompt congelato
 usato per l'esperimento originale (paragrafo 5.4).
======================================================================
"""

import os
import numpy as np
import pandas as pd

from backtest import backtest_portafoglio, calcola_metriche

CARTELLA_OUTPUT = os.path.join("..", "output")
```

```
FILE_MASTER = os.path.join(CARTELLA_OUTPUT,
"master_giornaliero_2016_2025.csv")

# ---------------------------------------------------------------
# Pesi - rigenerazione con prompt in ITALIANO (Appendice A.4-A.6)
# ---------------------------------------------------------------
PORTAFOGLI_IT = {
    "ChatGPT - Aggressivo":   {"SP500": 38, "EAFE": 22, "EM": 15,
"TREAS_10Y": 5,  "TREAS_2Y": 0,  "IG": 5,  "CORP_HY": 10, "ORO": 5},
    "ChatGPT - Bilanciato":   {"SP500": 25, "EAFE": 15, "EM": 10,
"TREAS_10Y": 10, "TREAS_2Y": 8,  "IG": 18, "CORP_HY": 6,  "ORO": 8},
    "ChatGPT - Conservativo": {"SP500": 10, "EAFE": 7,  "EM": 3,
"TREAS_10Y": 15, "TREAS_2Y": 25, "IG": 30, "CORP_HY": 2,  "ORO": 8},

    "Claude - Aggressivo":    {"SP500": 40, "EAFE": 20, "EM": 15,
"TREAS_10Y": 5,  "TREAS_2Y": 0,  "IG": 5,  "CORP_HY": 10, "ORO": 5},
    "Claude - Bilanciato":    {"SP500": 25, "EAFE": 15, "EM": 10,
"TREAS_10Y": 10, "TREAS_2Y": 10, "IG": 20, "CORP_HY": 5,  "ORO": 5},
    "Claude - Conservativo":  {"SP500": 10, "EAFE": 5,  "EM": 0,
"TREAS_10Y": 20, "TREAS_2Y": 25, "IG": 25, "CORP_HY": 5,  "ORO": 10},

    "Gemini - Aggressivo":    {"SP500": 50, "EAFE": 25, "EM": 15,
"TREAS_10Y": 5,  "TREAS_2Y": 0,  "IG": 0,  "CORP_HY": 0,  "ORO": 5},
    "Gemini - Bilanciato":    {"SP500": 35, "EAFE": 15, "EM": 5,
"TREAS_10Y": 15, "TREAS_2Y": 0,  "IG": 20, "CORP_HY": 5,  "ORO": 5},
    "Gemini - Conservativo":  {"SP500": 10, "EAFE": 10, "EM": 0,
"TREAS_10Y": 15, "TREAS_2Y": 30, "IG": 25, "CORP_HY": 0,  "ORO": 10},
}

# ---------------------------------------------------------------
# Pesi - rigenerazione con prompt in INGLESE (Appendice A.7-A.9)
# ---------------------------------------------------------------
PORTAFOGLI_EN = {
    "ChatGPT - Aggressivo":   {"SP500": 42, "EAFE": 20, "EM": 18,
"TREAS_10Y": 3,  "TREAS_2Y": 2,  "IG": 5,  "CORP_HY": 7,  "ORO": 3},
    "ChatGPT - Bilanciato":   {"SP500": 30, "EAFE": 13, "EM": 7,
"TREAS_10Y": 12, "TREAS_2Y": 8,  "IG": 18, "CORP_HY": 6,  "ORO": 6},
    "ChatGPT - Conservativo": {"SP500": 12, "EAFE": 5,  "EM": 3,
"TREAS_10Y": 12, "TREAS_2Y": 30, "IG": 28, "CORP_HY": 3,  "ORO": 7},

    "Claude - Aggressivo":    {"SP500": 40, "EAFE": 20, "EM": 20,
"TREAS_10Y": 5,  "TREAS_2Y": 0,  "IG": 5,  "CORP_HY": 5,  "ORO": 5},
    "Claude - Bilanciato":    {"SP500": 25, "EAFE": 10, "EM": 5,
"TREAS_10Y": 15, "TREAS_2Y": 10, "IG": 20, "CORP_HY": 5,  "ORO": 10},
    "Claude - Conservativo":  {"SP500": 10, "EAFE": 5,  "EM": 0,
"TREAS_10Y": 20, "TREAS_2Y": 25, "IG": 25, "CORP_HY": 5,  "ORO": 10},

    "Gemini - Aggressivo":    {"SP500": 45, "EAFE": 20, "EM": 15,
"TREAS_10Y": 5,  "TREAS_2Y": 0,  "IG": 0,  "CORP_HY": 10, "ORO": 5},
    "Gemini - Bilanciato":    {"SP500": 35, "EAFE": 15, "EM": 5,
"TREAS_10Y": 15, "TREAS_2Y": 5,  "IG": 15, "CORP_HY": 5,  "ORO": 5},
    "Gemini - Conservativo":  {"SP500": 15, "EAFE": 5,  "EM": 0,
"TREAS_10Y": 20, "TREAS_2Y": 30, "IG": 20, "CORP_HY": 5,  "ORO": 5},
}
```

```
# ------------------------------------------------------------
# Pesi originali (Appendice A.1-A.3), per il confronto
# ------------------------------------------------------------
PORTAFOGLI_ORIGINALI = {
    "ChatGPT - Aggressivo":   {"SP500": 45, "EAFE": 20, "EM": 20,
"TREAS_10Y": 5,  "TREAS_2Y": 0,  "IG": 0,  "CORP_HY": 5,  "ORO": 5},
    "ChatGPT - Bilanciato":   {"SP500": 30, "EAFE": 15, "EM": 10,
"TREAS_10Y": 15, "TREAS_2Y": 5,  "IG": 20, "CORP_HY": 0,  "ORO": 5},
    "ChatGPT - Conservativo": {"SP500": 15, "EAFE": 5,  "EM": 0,
"TREAS_10Y": 20, "TREAS_2Y": 20, "IG": 35, "CORP_HY": 0,  "ORO": 5},
    "Claude - Aggressivo":    {"SP500": 45, "EAFE": 20, "EM": 15,
"TREAS_10Y": 5,  "TREAS_2Y": 0,  "IG": 0,  "CORP_HY": 10, "ORO": 5},
    "Claude - Bilanciato":    {"SP500": 30, "EAFE": 15, "EM": 10,
"TREAS_10Y": 15, "TREAS_2Y": 5,  "IG": 15, "CORP_HY": 5,  "ORO": 5},
    "Claude - Conservativo":  {"SP500": 15, "EAFE": 5,  "EM": 0,
"TREAS_10Y": 15, "TREAS_2Y": 25, "IG": 30, "CORP_HY": 5,  "ORO": 5},
    "Gemini - Aggressivo":    {"SP500": 50, "EAFE": 25, "EM": 15,
"TREAS_10Y": 0,  "TREAS_2Y": 0,  "IG": 0,  "CORP_HY": 10, "ORO": 0},
    "Gemini - Bilanciato":    {"SP500": 30, "EAFE": 15, "EM": 5,
"TREAS_10Y": 15, "TREAS_2Y": 0,  "IG": 30, "CORP_HY": 0,  "ORO": 5},
    "Gemini - Conservativo":  {"SP500": 15, "EAFE": 5,  "EM": 0,
"TREAS_10Y": 10, "TREAS_2Y": 35, "IG": 30, "CORP_HY": 0,  "ORO": 5},
}


def hhi(pesi):
    """Indice di Herfindahl-Hirschman sugli otto indici originali (non
    aggregati): somma dei quadrati dei pesi in quota."""
    return sum((w / 100) ** 2 for w in pesi.values())


def verifica_formale(nome, pesi):
    somma = sum(pesi.values())
    ok = abs(somma - 100) < 1e-9
    stato = "OK" if ok else "NON VALIDO"
    print(f"  [{stato}] {nome:<24} somma={somma:.0f}")
    return ok


def main():
    print("=" * 78)
    print("VERIFICA FORMALE DEI 18 PORTAFOGLI RIGENERATI (prima del
backtest)")
    print("=" * 78)
    tutti_validi = True
    for etichetta, portafogli in [("IT", PORTAFOGLI_IT), ("EN",
PORTAFOGLI_EN)]:
        for nome, pesi in portafogli.items():
            if not verifica_formale(f"{nome} ({etichetta})", pesi):
                tutti_validi = False
    if not tutti_validi:
        raise ValueError("Almeno un portafoglio non e' valido - verificare
prima di procedere")
```

```
    print("\nTutti i 18 portafogli rigenerati superano la verifica
formale.\n")

    master = pd.read_csv(FILE_MASTER, index_col=0, parse_dates=True)
    prezzi = master.drop(columns=["RF_DTB3"])
    rf = master["RF_DTB3"].mean() / 100.0

    print("=" * 78)
    print("BACKTEST (2016-2025, ribilanciamento trimestrale) - stesso motore
di backtest.py")
    print("=" * 78)

    righe = []
    for nome in PORTAFOGLI_ORIGINALI:
        riga = {"Portafoglio": nome}
        for etichetta, portafogli in [("orig", PORTAFOGLI_ORIGINALI), ("IT",
PORTAFOGLI_IT), ("EN", PORTAFOGLI_EN)]:
            pesi = portafogli[nome]
            val, _ = backtest_portafoglio(prezzi, pesi)
            m = calcola_metriche(val, risk_free_annuo=rf)
            riga[f"HHI_{etichetta}"] = round(hhi(pesi), 3)
            riga[f"Sharpe_{etichetta}"] = round(m["Sharpe"], 3)
            riga[f"CAGR_{etichetta}"] = round(m["CAGR"] * 100, 2)
        righe.append(riga)

    tab = pd.DataFrame(righe).set_index("Portafoglio")
    print(tab.to_string())

    # ---------------------------------------------------------------
    # Percentuale di capitale riallocato rispetto all'originale
    # (distanza L1 tra i due vettori di pesi, divisa per 2)
    # ---------------------------------------------------------------
    print("\n" + "=" * 78)
    print("PERCENTUALE DI CAPITALE RIALLOCATO RISPETTO ALL'ORIGINALE")
    print("=" * 78)
    colonne = ["SP500", "EAFE", "EM", "TREAS_10Y", "TREAS_2Y", "IG",
"CORP_HY", "ORO"]
    scarti = []
    for nome in PORTAFOGLI_ORIGINALI:
        orig = PORTAFOGLI_ORIGINALI[nome]
        for etichetta, portafogli in [("IT", PORTAFOGLI_IT), ("EN",
PORTAFOGLI_EN)]:
            nuovo = portafogli[nome]
            distanza = sum(abs(orig[c] - nuovo[c]) for c in colonne) / 2
            scarti.append(distanza)
            print(f"  {nome:<24} ({etichetta}): {distanza:.1f}% del capitale
riallocato")
    print(f"\nMedia su tutti i confronti: {sum(scarti)/len(scarti):.1f}%")

    # ---------------------------------------------------------------
    # Salvataggio
    # ---------------------------------------------------------------
    os.makedirs(CARTELLA_OUTPUT, exist_ok=True)
    p_csv = os.path.join(CARTELLA_OUTPUT, "rigenerazione_metriche.csv")
```

```
    tab.to_csv(p_csv, float_format="%.6f")
    print(f"\nFile prodotto:\n  {p_csv}")
    print("\nCOMPLETATO: backtest delle rigenerazioni (paragrafo 5.4).")


if __name__ == "__main__":
    main()
```

APPENDICE C - LOG DELLE CORREZIONI SUI DATI

Registro generato automaticamente dalla procedura di caricamento e pulizia dei dati (Appendice B.1) al momento dell'esecuzione. Si riporta integralmente:

```
LOG DELLE CORREZIONI E DELLE SCELTE DI PULIZIA DATI
============================================================
SP500: file '1_SPXTR.xlsx', colonna prezzo 'Close', 5031 osservazioni (2006-
01-03 - 2025-12-31)
EAFE: file '2_MSCI_EAFE.xlsx', colonna prezzo 'Close', 5218 osservazioni
(2006-01-02 - 2025-12-31)
EM: file '3_MSCI_EM.xlsx', colonna prezzo 'Close', 5216 osservazioni (2006-
01-02 - 2025-12-31)
TREAS_10Y: file '4b_Treasury_10y.xlsx', colonna prezzo 'Trade Price', 4303
osservazioni (2009-01-01 - 2025-12-31)
TREAS_2Y: file '5_FTSE_Treasury_2y.xlsx', colonna prezzo 'Trade Price', 4303
osservazioni (2009-01-01 - 2025-12-31)
IG: file '6_FTSE_IG.xlsx', colonna prezzo 'Trade Price', 5132 osservazioni
(2006-01-03 - 2025-12-31)
CORP_HY: file '7_Corporate_HY.xlsx', colonna prezzo 'Trade Price', 5179
osservazioni (2006-01-03 - 2025-12-31)
ORO: file '8_Gold.xlsx', colonna prezzo 'Bid', 5201 osservazioni (2006-01-03
- 2025-12-31)

CORREZIONI PUNTUALI (valore rimosso e interpolato linearmente):
  TREAS_10Y 2022-01-18: rimosso 708.0048, interpolato 791.1549. Motivo:
Errore di pubblicazione indice FTSE (bad tick, reversal completo il 19/01)
  TREAS_2Y 2022-01-18: rimosso 410.4919, interpolato 422.1766. Motivo: Errore
di pubblicazione indice FTSE (bad tick, reversal completo il 19/01)
  IG 2022-01-18: rimosso 1,745.3586, interpolato 1,904.2449. Motivo: Errore
di pubblicazione indice FTSE (bad tick, reversal completo il 19/01)

ALLINEAMENTO CALENDARI (calendario master: giorni di negoziazione S&P 500):
  SP500: 0 date colmate con ultimo valore disponibile (festivita' locali), 0
NA residui
  EAFE: 0 date colmate con ultimo valore disponibile (festivita' locali), 0
NA residui
  EM: 0 date colmate con ultimo valore disponibile (festivita' locali), 0 NA
residui
  TREAS_10Y: 32 date colmate con ultimo valore disponibile (festivita'
locali), 0 NA residui
  TREAS_2Y: 32 date colmate con ultimo valore disponibile (festivita'
locali), 0 NA residui
```

```
  IG: 32 date colmate con ultimo valore disponibile (festivita' locali), 0 NA
residui
  CORP_HY: 2 date colmate con ultimo valore disponibile (festivita' locali),
0 NA residui
  ORO: 0 date colmate con ultimo valore disponibile (festivita' locali), 0 NA
residui

RISK-FREE: FRED DTB3 (discount basis, % annua), festivita' colmate con ffill.
Media 2016-2025: 2.17%

MASTER MENSILE PRE-CUTOFF: fine mese, dal primo dato disponibile per serie al
31/12/2015.
  Nota: le serie FTSE Treasury (10Y, 2Y) partono dal 2009 e presentano due
lacune nel 2011;
  per eventuali statistiche storiche sui Treasury usare la finestra dal 2012.
```

APPENDICE D - VERIFICA DI ROBUSTEZZA SUI BENCHMARK DI CONFRONTO

Questa appendice riporta il controllo esplorativo richiamato nel nono limite del paragrafo 5.2. I benchmark di profilo definiti in Tabella 3.2 impiegano sei degli otto indici dell'universo investibile, escludendo corporate high yield e oro, mentre i modelli generativi hanno avuto accesso a tutte e otto le asset class in fase di allocazione. La verifica ricostruisce quindi i tre benchmark di profilo concedendo loro lo stesso universo investibile disponibile ai modelli, e ripete il confronto del Capitolo 4 su questa base.

Metodologia, dati e motore di backtest restano identici a quelli descritti nel Capitolo 3: il calcolo impiega lo stesso script di backtest riportato in Appendice B.2, applicato senza modifiche ai soli pesi della Tabella D.1, stesso dataset giornaliero 2016-2025, stesso ribilanciamento trimestrale, stesse formule per CAGR, volatilità, Sharpe Ratio, Maximum Drawdown e Calmar. Si sottolinea che i pesi dei benchmark alternativi non sono in alcun modo ottimizzati sui rendimenti del periodo: non vi è stima di rendimenti attesi, matrice di covarianza o procedura di ottimizzazione, e dunque nessuna forma di look-ahead bias. La costruzione segue la stessa logica qualitativa con cui erano stati definiti i benchmark originali.

D.1 - Costruzione dei benchmark alternativi

L'architettura di base resta quella dei benchmark originali: un blocco azionario ripartito 60/25/15 tra S&P 500, mercati sviluppati ed emergenti, e un blocco obbligazionario ripartito 40/30/30 tra Treasury a 10 anni, Treasury a 2 anni e credito investment grade. Resta invariato anche il rapporto azionario/obbligazionario che qualifica i tre profili - 80/20 per l'aggressivo, 60/40 per il bilanciato, 20/80 per il conservativo - il quale continua però a valere all'interno della sola componente core.

A questa componente core si affianca una componente satellite composta da corporate high yield e oro, di dimensione crescente al calare del profilo di rischio: 8 per cento nell'aggressivo, 12 per cento nel bilanciato, 18 per cento nel conservativo. La composizione interna del satellite riflette la funzione che le due asset class svolgono nei diversi profili: prevalenza di high yield nell'aggressivo (6 contro 2 per cento), e prevalenza dell'oro nel conservativo (15 contro 3 per cento). Il capitale destinato alla componente satellite viene sottratto in proporzione a tutte e sei le posizioni originarie della componente core, così da non alterarne i rapporti interni: ogni peso originario è moltiplicato per il complemento a uno della quota satellite. Ne consegue che l'esposizione azionaria complessiva risulta lievemente inferiore a quella nominale del profilo - 73,6 per cento nell'aggressivo, 52,8 nel bilanciato, 16,4 nel conservativo - pur restando immutato il rapporto tra azionario e obbligazionario all'interno della componente core.

Tabella D.1: Pesi dei benchmark originali e alternativi, per profilo.[50]

Benchmark	S&P500	EAFE	EM	T10Y	T2Y	IG	HY	Oro
Aggressivo - originale	48,0%	20,0%	12,0%	8,0%	6,0%	6,0%	—	—
Aggressivo - alternativo	44,2%	18,4%	11,0%	7,4%	5,5%	5,5%	6,0%	2,0%
Bilanciato - originale	36,0%	15,0%	9,0%	16,0%	12,0%	12,0%	—	—
Bilanciato - alternativo	31,7%	13,2%	7,9%	14,1%	10,6%	10,6%	5,0%	7,0%
Conservativo - originale	12,0%	5,0%	3,0%	32,0%	24,0%	24,0%	—	—
Conservativo - alternativo	9,8%	4,1%	2,5%	26,2%	19,7%	19,7%	3,0%	15,0%

Fonte: elaborazione propria su Tabella 3.2.

D.2 - Risultati

La Tabella D.2 riporta le metriche di sintesi per i tre benchmark originali, i tre benchmark alternativi e i nove portafogli generati dai modelli, con l'equal-weight in apertura come termine di riferimento comune.

Tabella D.2: Metriche di sintesi, benchmark originali e alternativi a confronto con i portafogli AI.

Portafoglio	CAGR	Volatilità	Sharpe	Max DD	Calmar
Equal-weight (1/N)	7,73%	6,27%	0,877	-18,25%	0,424
Aggressivo - originale	10,60%	11,43%	0,751	-25,94%	0,408
Aggressivo - alternativo	10,48%	10,76%	0,780	-25,17%	0,416
ChatGPT - Aggressivo	11,65%	12,11%	0,793	-28,84%	0,404
Claude - Aggressivo	11,54%	11,75%	0,806	-28,36%	0,407
Gemini - Aggressivo	11,76%	13,18%	0,747	-32,05%	0,367

(Tabella D.2 segue in pagina successiva)

[50] I pesi sono arrotondati al primo decimale; le righe possono quindi sommare a 100,1 anziché a 100.

Tabella D.2 (prosegue dalla pagina precedente)

Bilanciato - originale	8,49%	8,57%	0,743	-20,71%	0,410
Bilanciato - alternativo	8,92%	7,94%	0,844	-19,57%	0,456
ChatGPT - Bilanciato	8,48%	7,94%	0,793	-20,76%	0,408
Claude - Bilanciato	8,70%	8,07%	0,807	-20,63%	0,421
Gemini - Bilanciato	8,12%	7,60%	0,782	-20,60%	0,394
Conservativo - originale	3,99%	4,47%	0,417	-16,70%	0,239
Conservativo - alternativo	5,76%	4,97%	0,718	-15,68%	0,367
ChatGPT - Conservativo	4,93%	4,66%	0,594	-16,18%	0,305
Claude - Conservativo	5,19%	4,45%	0,676	-15,27%	0,340
Gemini - Conservativo	4,97%	4,17%	0,669	-14,18%	0,351

Fonte: elaborazione propria su dati LSEG e output dei 3 modelli (Appendice A e B).

Nel profilo aggressivo la conclusione del Capitolo 4 non cambia in modo sostanziale. Il benchmark alternativo migliora rispetto all'originale (Sharpe Ratio da 0,751 a 0,780), ma i portafogli generati da ChatGPT e Claude restano comunque avanti (0,793 e 0,806), mentre Gemini (0,747) resta leggermente indietro sia rispetto all'originale sia rispetto all'alternativo - un esito coerente con quanto già osservato nel paragrafo 5.1 a proposito dell'assenza di oro nel portafoglio aggressivo di quel modello.

Nel profilo bilanciato e, in misura più marcata, nel conservativo il quadro si modifica invece in modo rilevante. Nel bilanciato il benchmark alternativo raggiunge uno Sharpe Ratio di 0,844, superiore a tutti e tre i modelli (0,782-0,807); nel conservativo passa da 0,417 a 0,718, contro lo 0,594-0,676 dei modelli. In entrambi i profili, dunque, il vantaggio dei portafogli generati osservato nel Capitolo 4 si annulla o si inverte non appena il benchmark può impiegare le stesse otto asset class.

Merita una precisazione il caso del profilo conservativo, dove il miglioramento potrebbe essere attribuito a un semplice effetto di riduzione del

rischio: così non è. Il benchmark alternativo presenta infatti una volatilità annualizzata superiore a quella dell'originale (4,97 contro 4,47 per cento), a fronte di un'esposizione azionaria più contenuta (16,4 contro 20 per cento). Il miglioramento dello Sharpe Ratio deriva quindi dal rendimento (CAGR dal 3,99 al 5,76 per cento) e non da un profilo più prudente: nel decennio in esame l'oro e l'high yield, poco correlati con l'azionario e il governativo, hanno aggiunto rendimento più di quanto abbiano aggiunto rischio complessivo. Si osserva inoltre che il Maximum Drawdown migliora in tutti e tre i profili (da -25,94 a -25,17 per cento nell'aggressivo, da -20,71 a -19,57 nel bilanciato, da -16,70 a -15,68 nel conservativo).

Resta infine confermato il risultato principale del lavoro: nessun portafoglio, né generato dai modelli né costruito come benchmark alternativo, raggiunge lo Sharpe Ratio della diversificazione equiponderata (0,877).

D.3 - Limiti della verifica

Questa verifica non costituisce una revisione del disegno sperimentale descritto nel Capitolo 3, che resta congelato nella forma in cui è stato definito prima della generazione dei portafogli, e i suoi esiti non sono stati utilizzati per modificare in alcun modo i risultati del Capitolo 4. Va inoltre ribadito che i benchmark alternativi qui impiegati non sono portafogli ottimizzati in senso media-varianza: la loro composizione riflette una convenzione ragionevole, come lo era quella dei benchmark originali, e non l'unica costruzione possibile. L'esercizio ha quindi valore esplorativo e serve a delimitare, non a quantificare, la sensibilità dei risultati alla scelta del termine di paragone. Come indicato nelle conclusioni, una verifica sistematica richiederebbe l'esplorazione di un insieme ampio di composizioni alternative.

RINGRAZIAMENTI

Desidero innanzitutto ringraziare la mia relatrice, la Prof.ssa Caterina Lucarelli, per la disponibilità, l'attenzione e il supporto dimostrati durante la realizzazione di questo lavoro. I suoi suggerimenti e il confronto avuto in questi mesi hanno contribuito in maniera importante allo sviluppo della tesi.

Un ringraziamento speciale va ai miei genitori, Sergio ed Emanuela, che mi hanno sempre sostenuto nelle mie scelte e mi hanno permesso di affrontare questi anni con serenità e determinazione. Il loro supporto, la loro fiducia e la loro presenza sono stati fondamentali per raggiungere questo traguardo.

Grazie a mia sorella Elisa, per avermi insegnato, soprattutto con il suo esempio, cosa significhi perseverare anche nei momenti più difficili e continuare a credere nei propri obiettivi senza arrendersi.

Grazie a Giulia, per essermi stata accanto durante tutto questo percorso, per aver condiviso con me le soddisfazioni e i momenti più difficili e per avermi sempre incoraggiato a continuare a dare il massimo.

Un grazie anche a Vittoria, che con i suoi due anni è riuscita a ricordarmi, spesso senza lasciarmi molta possibilità di scelta, quando fosse arrivato il momento di fare una pausa tra una sessione di studio e l'altra.

Infine, grazie a tutte le persone che, in modi diversi, hanno fatto parte di questi anni e hanno contribuito a rendere questo percorso qualcosa che porterò con me anche oltre l'università.