AI & ETHICS LITERACY for Teachers

Un manuale per l'educazione all'etica dell'Intelligenza Artificiale nella scuola secondaria di secondo grado

a cura di
Aldo Pisano, Ines Crispini, Tiziana Catarci, Maria Cristina Carrisi, Piergiorgio Donatelli, Colin de la Higuera



UNICApress/didattica

Come si educa una generazione che crescerà prendendo decisioni insieme all'Intelligenza Artificiale? E come si formano cittadini capaci non soltanto di utilizzare l'IA, ma di comprenderne criticamente il funzionamento, riconoscerne i limiti e valutarne le implicazioni etiche?

AI & Ethics Literacy for Teachers nasce per rispondere a questa sfida, proponendo un modello educativo che integra competenze tecnologiche, riflessione filosofica e pratica didattica. Il volume presenta il framework dell'AI & Ethics Literacy attraverso un approccio autenticamente interdisciplinare, nel quale filosofia morale, etica dell'Intelligenza Artificiale, ingegneria informatica, didattica della matematica, pedagogia speciale, scienze dell'educazione e metodologia della Philosophical Enquiry concorrono a offrire ai docenti una prospettiva unitaria per leggere la complessità dell'ecosistema digitale contemporaneo.

Pensato principalmente per la scuola secondaria di secondo grado, ma facilmente adattabile anche alla scuola secondaria di primo grado, il manuale unisce rigore scientifico, chiarezza espositiva e immediatezza operativa. Accanto ai fondamenti teorici – dal Large Ethics Model alla tassonomia PAIA, dall'etica normativa ai principi dell'Intelligenza Artificiale – propone laboratori, schede didattiche, percorsi interdisciplinari, domande guida e attività immediatamente trasferibili nella pratica quotidiana dell'insegnamento.

Il volume parte da una convinzione precisa: le nuove esigenze educative non devono necessariamente tradursi in ulteriori contenuti, discipline o carico di lavoro aggiuntivo per i docenti né tantomeno richiedono di abbandonare ciò che funziona. Servono strumenti capaci di aiutare a rileggere e a orientare ciò che già si insegna: una cassetta degli attrezzi agile, flessibile e scientificamente fondata. Per questo il textbook non propone un curricolo separato, ma un innesto flessibile all'interno dei percorsi esistenti, adattabile alle diverse discipline, alla progettazione interdisciplinare e all'educazione civica, offrendo una nuova chiave di lettura dei contenuti disciplinari e facendo emergere le questioni etiche, sociali ed epistemologiche implicate dall'uso delle tecnologie di IA. Si vuole promuovere capacità argomentative, pensiero critico, consapevolezza epistemica e cittadinanza digitale per formare cittadini capaci non soltanto di usare le tecnologie, ma di comprenderle, discuterle e orientarle consapevolmente.

Frutto di una sperimentazione pilota nelle scuole secondarie italiane e validato da docenti appartenenti a differenti aree disciplinari, il volume propone un modello trasferibile di innovazione educativa, orientato alla costruzione di una cultura dell'Intelligenza Artificiale che metta al centro la persona, il dialogo e la responsabilità.

UNICApress/didattica

AI & ETHICS LITERACY
for Teachers

Un manuale per l'educazione all'etica dell'Intelligenza Artificiale nella scuola secondaria di secondo grado

a cura di
Aldo PISANO, Ines CRISPINI, Tiziana CATARCI,
Maria Cristina CARRISI, Piergiorgio DONATELLI, Colin DE LA HIGUERA

Cagliari
UNICApress
2026

© 2026 — Gli autori, le autrici e UNICApress

Quest'opera è distribuita con licenza Creative Commons Attribuzione – Non commerciale – Condividi allo stesso modo 4.0 Internazionale (CC BY-NC-SA 4.0).

Prima edizione: UNICApress, 2026.

ISBN: 978-88-3312-248-9

DOI: https://doi.org/10.13125/unicapress.978-88-3312-248-9

In copertina: immagine ottenuta tramite programma di intelligenza artificiale, a cura degli autori

---

NOTA SULLA VALIDAZIONE

Il presente manuale è stato validato da 28 docenti della scuola secondaria di secondo grado provenienti da aree disciplinari diverse.

La fase di validazione, condotta nella primavera 2026 tramite scheda strutturata (valutazioni su scala 1–5 e commenti qualitativi), ha orientato la revisione finale del testo. Il confronto con i docenti validatori ha permesso di intervenire sulla densità del linguaggio, sugli esempi disciplinari, sull'accessibilità del framework teorico e sull'articolazione dei laboratori didattici.

Citazione consigliata: Pisano A., Crispini I., Catarci T., Carrisi M.C., Donatelli P., De la Higuera, C., (a cura di), *AI & Ethics Literacy for Teachers. Un manuale per l'educazione all'etica dell'Intelligenza Artificiale nella scuola secondaria di secondo grado*, UNICApress, 2026.

CON IL CONTRIBUTO E IL PATROCINIO DI

SIpEIA — Società Italiana per l'Etica dell'Intelligenza Artificiale

Università della Calabria

Università di Foggia

UNESCO Chair RELIA, Université de Nantes

Consulta di Bioetica

University of Ljubljana — Faculty of Theology

Centre for Human-Centred AI

The College of Staten Island of The City University of New York

CON IL PATROCINIO E LA COLLABORAZIONE DI

Liceo "T. Campanella" · ITIS "E. Fermi" — ITC "L. Brutium"

The College of Staten Island, CUNY · UNESCO Chair RELIA, Université de Nantes

CURATORI E COLLABORATORI

CURATELA

Aldo Pisano, Università della Calabria — SIpEIA.
Ines Crispini, Università della Calabria — SIpEIA.
Tiziana Catarci, Sapienza Università di Roma — CNR — SIpEIA.
Maria Cristina Carrisi, Università degli Studi di Cagliari.
Piergriogio Donatelli, Sapienza Università di Roma.
Colin de la Higuera (versione inglese), UNESCO Chair RELIA, Université de Nantes.

CONTRIBUTI E COLLABORAZIONI

Loredana Cardullo (Università di Catania) · Maria Concetta Carruba (Università Pegaso) · Stephen Ferst (College of Staten Island — CUNY) · Dario Lombardi (Università di Foggia) · Massimiliano Mancini (Università di Firenze) · Emiliana Murgia (Università di Genova) · Giorgia Nieddu (Università di Cagliari) · Rossana Adele Rossi (Università della Calabria) · Nunziatina Sanfilippo (Università di Catania) · Vojko Strahovnik (University of Ljubljana) · Luca Tenneriello (Sapienza Università di Roma) · Antonella Valenti (Università della Calabria).

PARTNER ISTITUZIONALI

SIpEIA — Società Italiana per l'Etica dell'Intelligenza Artificiale · Consulta di Bioetica · UNESCO Chair RELIA (Université de Nantes) · Center for Human-Centred AI and New Ethics of Technologies (University of Ljubljana) · Dipartimento di Culture, Educazione e Società — Università della Calabria · Dipartimento di Scienze della Formazione — Università di Catania · College of Staten Island, CUNY.

DOCENTI VALIDATORI

Si ringraziano i docenti della scuola secondaria di secondo grado che hanno contribuito alla fase di validazione del manuale: Antonella Amodeo · Carla Armigero · Ernesto Astorino · Maria Grazia Barbanera · Paola Burza · Giuseppe Catalano · Annalisa Cervello · Graziella Cosentino · Vincenzo Curcio · Silvana De Cicco · Francesca De Luca · Giovanna Di Tria · Alessandro Donato · Rosanna Fiorino · Angela Fortino · Annabella Gallina · Emiliano Iusi · Gaetano Marco Lanzafame · Benedetta Leonetti · Sabina Lettori · Cinzia Nicoletti · Rita Pisano · Maria Pontuale · Anna Ruocco · Lucia Giuseppina Scudo · Carla Sollazzo · Ida Stumpo · Anna Ziviello.

DIRIGENTI SCOLASTICI

Si ringraziano le Dirigenti Scolastiche Rosita Paradiso e Annina Carnevale per il supporto alla fase pilota da cui è nato il presente manuale.

INDICE

Introduzione....12
1. AI & Ethics Literacy: un framework per l'educazione del futuro....14
1.1 Scopo e framework etico-educativo....14
Perché un'educazione all'etica dell'IA....14
La vulnerabilità epistemica: una sfida strutturale....14
Vulnerabilità epistemica e altre forme di vulnerabilità....15
Le conseguenze: l'ingiustizia epistemica....15
1.2 Il Large Ethics Model e la tassonomia PAIA....15
L'architettura del Large Ethics Model (LEM)....15
Il modello PAIA: quattro dimensioni di rischio....16
L'architettura multi-stakeholder della responsabilità....17
Due bias da combattere: etico e di automazione....17
1.3 Educare all'IA: competenze digitali e pensiero critico....18
L'educazione come risposta alla vulnerabilità epistemica....18
Le competenze chiave dell'AI & Ethics Literacy....18
Le evidenze empiriche: il pilot study italiano....19
Conclusione: verso una resilienza epistemica....20
Bibliografia essenziale....20
2. Etica e IA: i contenuti....20
2.1 Le etiche normative: orientarsi tra teoria e pratica....21
2.1.1 Il consequenzialismo: conta solo il risultato?....21
2.1.2 La deontologia: ci sono azioni sempre sbagliate?....22
2.1.3 L'etica della virtù: che persona voglio essere?....22
2.1.4 Confronto sinottico: tre sguardi sulla stessa azione....23
2.1.5 Dall'etica teorica all'etica applicata....23
2.1.6 I quattro principi di Beauchamp e Childress....24
2.1.7 Sacralità e qualità della vita....24
2.1.8 Portare l'etica in classe: suggerimenti pratici....24
2.1.9 Conclusione: l'etica come pratica del pensiero....25
2.2 Fondamenti di Intelligenza Artificiale....25
2.2.1 Il problema della definizione....26
2.2.2 Una classificazione degli agenti IA....27
2.2.3 IA debole e IA forte....28
2.2.4 Le tecnologie chiave: dal Machine Learning ai LLM....28
2.2.5 Come nascono i bias negli algoritmi....29

2.2.6 Amplificazione dei bias e cicli di retroazione 30
2.3 Etica dell'Intelligenza Artificiale 32
Tre modi di intendere l'educazione e l'IA 32
2.3.1 L'etica dell'IA è un'etica applicata? 32
2.3.2 I principi per un'IA affidabile 33
2.3.3 Casi concreti: le sfide etiche dell'IA oggi 35
2.3.4 Conclusione: la sfida educativa 35
3. Metodologia, strumenti e risorse didattiche 37
3.1 La Philosophical Enquiry come metodologia per l'AI & Ethics Literacy 37
Le origini: da Matthew Lipman alla Philosophical Enquiry 37
I principi pedagogici della PhiE 38
Le pratiche dialogiche: tre metodologie a confronto 38
Gli strumenti operativi della PhiE 38
Il mito in classe: Arianna, il Colosso, Medusa 38
3.2 Un curriculum verticale per il biennio e il triennio 39
Suggerimenti per il docente di sostegno (BES/DSA) 40
3.3 Philosophical practices e cultura della domanda 40
4. Schede operative: esempi didattici 43
4.1 Il Trolley Problem: etica normativa e intuizioni morali 43
Fasi — versione standard (90 min) 44
Variante breve (50 min) 44
4.2 Agenti IA: analitici, esecutivi, interattivi 45
Fasi — versione standard (90 min) 45
Variante breve (50 min) 46
4.3 PAIA all'opera: analisi di un sistema reale 46
Fasi — versione standard (90 min) 47
Variante breve (50 min) 47
4.4 Costruire un'IA affidabile: casi dal mondo reale 48
Fasi — versione standard (90 min) 48
Variante breve (50 min) 48
4.5 Dialogare con un LLM: prompt, bias e attendibilità 49
Fasi — versione standard (90 min) 49
Variante breve (50 min) 50
4.6 "Di chi ci fidiamo?" Statistiche, IA e disinformazione 50
Fasi — versione standard (90 min) 51
Variante breve (50 min) 51
4.7 Dualità etiche: i "tre coltelli" della tecnologia 52
Fasi — versione standard (90 min) 52

Variante breve (50 min) 53
4.8 Governance dell'IA: chi decide? 53
Fasi — versione standard (90 min) 54
Variante breve (50 min) 54
Nota finale — condivisione dei materiali 54
Glossario 55
Bibliografia generale 58
Ringraziamenti 60
Profili curatori e collaboratori 61

Introduzione

Il manuale *AI & Ethics Literacy for Teachers* presenta gli scopi e l'impianto teorico-metodologico dell'omonimo progetto, a partire dai risultati di uno studio pilota condotto nella scuola secondaria di secondo grado (Pisano, Crispini, Rossi, 2026). Il pilot study ha evidenziato come un approccio integrato all'alfabetizzazione all'intelligenza artificiale e all'etica favorisca un incremento significativo della consapevolezza critica, della capacità argomentativa e dell'atteggiamento riflessivo di studenti e docenti rispetto all'uso delle tecnologie di IA.

Muovendo da tali evidenze, il progetto si propone di estendere e strutturare un modello formativo trasferibile, fondato sull'intreccio tra etica come metodo di indagine e come contenuto, e sull'uso di pratiche dialogiche e inquiry-based. L'obiettivo principale è l'innesto sistematico di tale modello all'interno dell'asse delle competenze digitali, superando una concezione meramente strumentale dell'AI literacy e promuovendo una competenza digitale critica, responsabile e valorialmente orientata.

La sfida è promettente. Il pensiero critico — spesso evocato in astratto come obiettivo educativo generico — necessita di strumenti concreti per divenire operativo. La Philosophical Enquiry (PhiE) risponde a questa esigenza passando, paradossalmente, attraverso un'ulteriore presunta astrazione: la filosofia. Eppure è proprio nella pratica filosofica che il pensiero critico trova il suo esercizio più autentico. La PhiE applica la filosofia, attraverso il dialogo, come esercizio alla contraddizione e alla destrutturazione: il confronto con l'altro — il dialogo fra pari — diviene il terreno in cui si impara a sostenere una tesi e, al tempo stesso, a pensarne il contrario. Ciò che nasce come processo esterno (il dialogo nella comunità di ricerca) si trasforma così in processo interno: pensiero critico propriamente detto, ossia la capacità di pensare una cosa e il suo contrario, di sospendere il giudizio, di riformulare le proprie posizioni alla luce di ragioni migliori.

AI & Ethics Literacy for Teachers mira così a fornire agli insegnanti strumenti concettuali, didattici e metodologici per integrare stabilmente la riflessione etica sull'IA nella progettazione curricolare, contribuendo alla formazione di cittadini digitali capaci di comprendere, valutare e governare le implicazioni sociali, culturali e morali delle tecnologie emergenti. Il manuale si rivolge in particolare ai docenti della scuola secondaria di secondo grado, ma i suoi contenuti e il suo metodo sono pensati per essere adattabili a diversi contesti formativi.

La struttura del testo segue una logica progressiva. Il primo capitolo introduce il framework teorico: la vulnerabilità epistemica, il Large Ethics Model e le competenze chiave dell'AI & Ethics Literacy. Il secondo capitolo offre i contenuti disciplinari: le etiche normative, i fondamenti dell'IA e l'etica dell'IA come etica applicata. Il terzo capitolo presenta il metodo: la Philosophical Enquiry, le tecniche di facilitazione e il curricolo per classi. Il quarto capitolo, infine, fornisce schede operative ed esempi didattici pronti all'uso. Chiude il volume un glossario essenziale dei termini filosofici e tecnici.

Come leggere questo manuale

Il manuale può essere letto in modo lineare, dall'Introduzione al Capitolo 4, per chi desidera una formazione progressiva, oppure per singoli capitoli, come strumento di consultazione. Il Capitolo 4 contiene schede di laboratorio immediatamente utilizzabili in classe, ciascuna con una versione estesa (90 minuti) e una versione breve (50 minuti). Il glossario in chiusura può essere consultato in qualsiasi momento per chiarire i termini tecnici o filosofici incontrati nel testo.

1. AI & Ethics Literacy: un framework per l'educazione del futuro

Ines Crispini, Aldo Pisano, Stephen Ferst.

> *Immaginate un'aula del 2030. Gli studenti non usano solo l'intelligenza artificiale per i compiti — la incontrano ovunque: negli algoritmi che selezionano i contenuti sui social, nei sistemi di raccomandazione che orientano le loro scelte, nelle tecnologie che un giorno prenderanno decisioni sulla loro salute, sul loro lavoro, sulla loro libertà. La domanda non è più se l'IA farà parte della loro vita, ma come sapranno abitarla.*

Questo capitolo introduce il framework teorico che guida l'intero manuale: l'AI & Ethics Literacy, un approccio che integra alfabetizzazione tecnica e consapevolezza etica per preparare cittadini capaci di pensiero critico in un mondo pervaso dall'intelligenza artificiale.

1.1 Scopo e framework etico-educativo

Perché un'educazione all'etica dell'IA

Il rapido avanzamento dell'intelligenza artificiale pone sfide educative ed etiche senza precedenti. Non basta insegnare come funziona l'IA: occorre preparare gli studenti a riflettere criticamente sulle sue implicazioni sociali e a sviluppare una mentalità etica verso la tecnologia.

Come evidenziato dalla ricerca internazionale, l'educazione all'IA nelle scuole secondarie richiede non solo alfabetizzazione tecnica ma anche una riflessione etica strutturata, in linea con le linee guida europee per un'IA affidabile.

La vulnerabilità epistemica: una sfida strutturale

Al cuore del nostro framework c'è il concetto di vulnerabilità epistemica — una suscettibilità nella nostra capacità di conoscere, interpretare e dare senso alle informazioni. Non si tratta di una debolezza individuale, ma di una condizione strutturale che riguarda tutti, a ogni età.

> *La vulnerabilità epistemica è l'esposizione degli individui a rappresentazioni distorte o semplificate della realtà, che possono essere inconsapevolmente accettate come verità.*

Questa vulnerabilità nasce dalla convergenza di due vettori: interno (i nostri bias cognitivi) ed esterno (gli ecosistemi narrativi che ci circondano). L'educazione è la risposta più potente a questa sfida.

Vulnerabilità epistemica e altre forme di vulnerabilità

La tabella seguente mette in relazione la vulnerabilità epistemica con le altre forme di vulnerabilità con cui siamo familiari. Essa si distingue perché non riguarda beni materiali o integrità fisica, ma la nostra capacità stessa di interpretare la realtà.

Tipo di vulnerabilità	Descrizione
Epistemica	Suscettibilità a rappresentazioni distorte della realtà. Riguarda la capacità di interpretare il mondo e di costruire giudizi. È la condizione su cui questo manuale si concentra.
Economica	Esposizione a difficoltà finanziarie. Una persona economicamente sicura può comunque essere epistemicamente vulnerabile.
Fisica	Rischio di danni corporei. Una persona fisicamente sana può credere a falsità pericolose sulla salute.
Digitale	Rischi di privacy e sicurezza online. Un esperto di cybersecurity può non riconoscere la disinformazione.

Le conseguenze: l'ingiustizia epistemica

La filosofa Miranda Fricker ha concettualizzato l'ingiustizia epistemica — situazioni in cui le persone subiscono un torto specificamente nella loro capacità di conoscere. Quando le narrazioni dominanti distorcono la comprensione della realtà, le conseguenze investono l'intera società:

- **Attribuzione errata di responsabilità** — si incolpano i bersagli sbagliati per problemi complessi.
- **Impoverimento del dibattito pubblico** — le prospettive diverse vengono escluse o semplificate.
- **Erosione della democrazia** — senza autonomia di pensiero, viene meno la capacità di giudizio critico necessaria alla vita democratica.

1.2 Il Large Ethics Model e la tassonomia PAIA

L'architettura del Large Ethics Model (LEM)

Il Large Ethics Model (LEM) è un framework concettuale sviluppato per contrastare le distorsioni interpretative sistemiche, specialmente nel contesto delle tecnologie emergenti come l'IA. Non si tratta di un approccio puramente deontologico (regole universali) né puramente consequenzialista (focus sugli esiti), ma di un principialismo pratico ispirato al framework bioetico di Beauchamp e Childress.

L'architettura del LEM poggia su alcuni insight fondamentali:

- **L'etica dell'IA non è monolitica** — deve intersecarsi con molteplici domini (bioetica, etica ambientale, etica dei media).
- **I principi devono essere flessibili** — applicati attraverso la *phronesis* (saggezza pratica), non meccanicamente.
- **La differenziazione è cruciale** — le sfide etiche dell'IA in sanità differiscono da quelle in educazione o giustizia.

Il LEM in classe: un esempio concreto

Una scuola sta valutando l'adozione di un tutor IA per l'inglese. Il LEM invita il consiglio di classe a non affidarsi a un principio unico ("è efficace, quindi si usa") ma a intrecciare più piani di analisi: la trasparenza del sistema (come sceglie gli esercizi?), l'autonomia dello studente (quando può decidere di non seguirne i suggerimenti?), la distribuzione equa dell'accesso (chi non ha un PC a casa?), l'impatto ambientale e lavorativo (chi etichetta i dati su cui è addestrato?). Il LEM non fornisce la risposta giusta: fornisce la mappa delle domande che vanno poste prima di decidere.

Il modello PAIA: quattro dimensioni di rischio

Il PAIA è l'acronimo delle quattro proprietà che rendono l'IA un oggetto etico diverso dagli altri: Pervasività, Autonomia, Invisibilità, Adattività. Non è una teoria astratta, ma una griglia di analisi: di fronte a un sistema di IA, un docente o uno studente possono interrogarlo lungo queste quattro dimensioni per individuare dove si annidano le questioni etiche.

Dimensione	Implicazioni etiche
Pervasività (P)	L'IA si integra in ogni aspetto della vita. Comprensioni distorte possono avere conseguenze diffuse in molteplici domini.
Autonomia (A)	L'IA può interrompere processi decisionali umani. Dove finisce la decisione umana e inizia quella della macchina?
Invisibilità (I)	Gli algoritmi sono opachi. Questo favorisce l'illusione che l'IA sia neutrale, quando riflette i bias dei suoi creatori.
Adattività (A)	L'IA apprende e cambia nel tempo. Il suo comportamento può evolversi in modi che gli utenti non anticipano.

Il PAIA applicato a un caso scolastico

Consideriamo un feed social usato ogni giorno dagli studenti. Applichiamo il PAIA.

Pervasività: l'algoritmo non si limita al tempo libero, influenza conversazioni in classe, umore, attenzione. Autonomia: lo studente crede di scegliere cosa guardare, ma la sequenza è selezionata da un sistema ottimizzato per trattenere l'attenzione. Invisibilità: nessuno —

neppure l'utente — vede quali segnali il sistema sta usando per raccomandare i contenuti. Adattività: il feed di oggi non è quello di ieri; il sistema si riaddestra sulle reazioni dell'utente.

Il PAIA non dice se quel sistema è "buono" o "cattivo": mostra quattro punti in cui un docente può costruire domande didattiche operative — per esempio una sessione in cui gli studenti registrano la propria timeline per una settimana e provano a ricostruire i criteri di selezione del sistema.

L'architettura multi-stakeholder della responsabilità

Una delle caratteristiche più distintive del LEM è la mappatura della responsabilità etica su tutti gli stakeholder dell'ecosistema IA:

Stakeholder	Responsabilità
Policymaker (Po)	Creare framework normativi che bilancino innovazione e protezione dell'autonomia umana.
Finanziatori (Fu)	Orientare lo sviluppo dell'IA considerando le implicazioni etiche dei progetti finanziati.
Sviluppatori (De)	Creare sistemi trasparenti, equi, rispettosi dell'autonomia umana — e possedere alfabetizzazione etica.
Utenti (Us)	Impiegare i sistemi IA criticamente, mantenere supervisione umana, richiedere trasparenza.

Il LEM riconosce che gli utenti di oggi possono diventare i policymaker, finanziatori o sviluppatori di domani. Questo crea un «circolo paidetico» — un ciclo in cui l'educazione forma i futuri stakeholder, che a loro volta plasmano l'ecosistema IA che le generazioni future abiteranno.

Due bias da combattere: etico e di automazione

Bias etico (EBi)	Bias di automazione (ABi)
La percezione errata che l'etica sia un insieme di regole restrittive che ostacolano il progresso e l'innovazione.	La tendenza a riporre fiducia eccessiva nei sistemi automatizzati, assumendo che siano oggettivi e superiori al giudizio umano.

Entrambi i bias rappresentano forme di vulnerabilità epistemica. L'educazione deve riformulare l'etica come disciplina costruttiva e pratica, e aiutare gli studenti a comprendere che i sistemi IA, per quanto sofisticati, sono prodotti di scelte umane e possono riflettere errori e pregiudizi umani.

1.3 Educare all'IA: competenze digitali e pensiero critico

L'educazione come risposta alla vulnerabilità epistemica

Confrontare la vulnerabilità epistemica richiede di dotare gli individui delle competenze e disposizioni per navigare criticamente sia i propri bias cognitivi sia il panorama informativo esterno. Questo è fondamentalmente un progetto educativo.

Attraverso l'intero arco della vita — dalla scuola primaria all'apprendimento adulto — l'educazione può coltivare le virtù e le abilità necessarie per resistere a rappresentazioni distorte o semplificate della realtà.

Le competenze chiave dell'AI & Ethics Literacy

Metacognizione: pensare il proprio pensiero. Il primo passo per immunizzare le persone dalle distorsioni epistemiche è aiutarle a riconoscere le proprie abitudini mentali. Questo significa integrare la metacognizione — pensare al proprio modo di pensare — nei processi di apprendimento. Gli studenti beneficiano dall'apprendere esplicitamente i bias cognitivi più comuni: bias di conferma, euristica della disponibilità, bisogno di chiusura cognitiva, schemi culturali e ideologici.

Competenze ermeneutiche: l'arte dell'interpretazione. Se la vulnerabilità epistemica è in parte un'incapacità di interpretare informazioni complesse o nuove, l'educazione deve focalizzarsi sulle competenze ermeneutiche. La metodologia della Philosophical Enquiry (PhiE) colloca gli studenti in una «comunità di ricerca» dove interpretano collaborativamente testi o scenari, pongono domande ed esplorano molteplici punti di vista.

Alfabetizzazione critica: leggere oltre la superficie. L'educazione all'alfabetizzazione critica incoraggia gli studenti non solo a comprendere i contenuti superficiali, ma a esaminare agenda, bias, retorica e contesto. Nel mondo digitale, questo si estende alla comprensione di come gli algoritmi trattano le informazioni e come i trend sui social possono essere orchestrati.

Mappatura dei livelli: dalla tecnologia all'etica. Usando il concetto di Livelli di Astrazione (LdA) di Floridi, gli studenti esaminano le questioni a livello tecnologico, sociale ed etico — separatamente e poi insieme. Questo riduce la suscettibilità alle fallacie della «causa singola» e rivela più punti di intervento dove gli individui possono agire.

ESEMPIO DISCIPLINARE — SCIENZE — BIOLOGIA, CHIMICA

Il PAIA applicato a un sistema diagnostico di imaging medicale è un ottimo ponte con la bioetica: Pervasività (diagnostica diffusa), Autonomia (chi decide, medico o sistema?), Invisibilità (su quali dataset è stato addestrato?), Adattività (il sistema è aggiornato con nuovi casi?). L'allineamento con i quattro principi di Beauchamp e Childress (§ 2.1.6) rende il collegamento con la biologia e la chimica particolarmente naturale.

ESEMPIO DISCIPLINARE — MATEMATICA

Il bias non è solo un concetto etico: è un concetto statistico. Partire da qui permette di collegare naturalmente il Capitolo 1 con la statistica descrittiva e inferenziale, la teoria del campionamento, la propagazione degli errori. Un sistema di raccomandazione addestrato su dati non rappresentativi è, matematicamente, uno stimatore distorto. La stessa logica si estende all'analisi (curve di apprendimento, overfitting) e alla geometria (spazi di rappresentazione vettoriale delle parole).

ESEMPIO DISCIPLINARE — LINGUA E CULTURA INGLESE

Il framework LEM/PAIA si presta naturalmente a lavori CLIL: gli studenti analizzano un articolo originale in inglese (es. un report OECD o un white paper UNESCO), estraggono i quattro attributi PAIA e li rielaborano in una presentazione orale o scritta. Il lessico specialistico (pervasiveness, autonomy, invisibility, adaptivity) diventa parte integrante del vocabolario disciplinare.

ESEMPIO DISCIPLINARE — SOSTEGNO E BISOGNI EDUCATIVI SPECIALI

Il PAIA può essere introdotto con un linguaggio semplificato e con esempi molto concreti e ripetuti (l'app che suggerisce video, l'assistente vocale di casa). Per favorire l'accessibilità si consiglia: schede con icone per le quattro lettere del PAIA, domande guida in formato sì/no, lavori a coppie in cui uno studente osserva e l'altro formula la domanda. Questo trasforma un framework astratto in una check-list operativa.

Le evidenze empiriche: il pilot study italiano

Un recente studio pilota condotto nelle scuole secondarie italiane (Pisano, Crispini, Rossi, 2026) ha dimostrato che interventi educativi ben progettati possono effettivamente migliorare la resilienza epistemica degli studenti. Il progetto, che ha coinvolto 193 studenti tra istituti tecnici e licei, ha utilizzato la metodologia della Philosophical Enquiry in sessioni di 8 ore per classe.

I risultati principali:

- **Consapevolezza etica** — miglioramento significativo ($p < 0.001$, Cohen's $d \approx 0.8$).
- **Conoscenza dell'IA** — guadagno significativo nelle competenze concettuali ($d \approx 0.6$).
- **Riconoscimento dei bias** — shift significativo verso maggiore scetticismo sull'oggettività dell'IA ($d \approx 0.7$).
- **Gradimento metodologico** — alto apprezzamento per co-costruzione della conoscenza e dialogo rispettoso.

Le attività realizzate nelle classi pilota hanno intrecciato momenti di brainstorming iniziale su casi di cronaca, lavori di gruppo sul PAIA, simulazioni di dialogo sul trolley problem, analisi di output di sistemi generativi e debriefing metacognitivi. Questo manuale sistematizza quelle esperienze in un formato trasferibile.

Conclusione: verso una resilienza epistemica

La vulnerabilità epistemica può essere trasformata da una condizione che subiamo passivamente in una sfida che affrontiamo attivamente. Trattandola come oggetto educativo — qualcosa attorno a cui strutturare curricoli e interventi deliberati — affermiamo l'agency umana sulle forze che offuscano la nostra comprensione.

Attraverso riflessione, dialogo e apprendimento guidati da framework comprensivi come il LEM, le persone possono reclamare il potere di interpretare il loro mondo accuratamente e giustamente. Investendo in queste pratiche educative, investiamo nel tessuto stesso della democrazia e della società umana.

Riepilogo — i tre pilastri dell'AI & Ethics Literacy

1. Vulnerabilità epistemica: riconoscerla come condizione strutturale che riguarda tutti.
2. Large Ethics Model: un framework multi-stakeholder per affrontare le sfide etiche dell'IA.
3. Educazione critica: metacognizione, ermeneutica, alfabetizzazione critica e mappatura dei livelli.

Bibliografia essenziale

Beauchamp T.L., Childress J.F., *Principles of Biomedical Ethics*, Oxford University Press, New York 1979.

Cuomo S., Ranieri M., Biagini G., *Scuola e Intelligenza Artificiale*, Carocci, Roma 2024.

Floridi L., *Etica dell'Intelligenza Artificiale*, Raffaello Cortina, Milano 2022.

Fricker M., *Epistemic Injustice: Power and the Ethics of Knowing*, Oxford University Press, 2007.

Holmes W., Porayska-Pomsta K. (eds.), *The Ethics of Artificial Intelligence in Education*, Routledge, London 2023.

Pisano A., "Does the AI Need a Genealogy? The 'PAIA' Model", *Mimesis Journal — Mechané* 9(1), 2025.

Pisano, A., Crispini, I., Rossi, R.A. (2026). AI and Ethics Literacy in Secondary School: A Pilot Study. In: Dipace, A., Limongelli, C., Marras, M., Pagliara, S.M. (eds) Artificial Intelligence with and for Learning Sciences. Past, Present, and Future Horizons. WAILS 2025. Lecture Notes in Computer Science, vol 16438. Springer, Cham.

Worley P., *The If Machine: Philosophical Enquiry in the Classroom*, Bloomsbury, London 2010.

Zuboff S., *Il capitalismo della sorveglianza. Il futuro dell'umanità nell'era dei nuovi poteri*, LUISS University Press, Roma 2019.

2. Etica e IA: i contenuti

Maria Cristina Carrisi, Tiziana Catarci, Colin de la Higuera, Massimiliano Mancini, Giorgia Nieddu, Vojko Strahovnik, Luca Tenneriello.

Nello specifico: Luca Tenneriello e Vojko Strahovnik si sono occupati del paragrafo 2.1; Maria Cristina Carrisi, Tiziana Catarci, Colin de la Higuera e Giorgia Nieddu del 2.2; Massimiliano Mancini, Tiziana Catarci del paragrafo 2.3.

2.1 Le etiche normative: orientarsi tra teoria e pratica

> *Immaginate di essere su un treno. I freni non funzionano. Sul binario ci sono cinque operai che moriranno se non fate nulla. Ma c'è una leva: tirandola, deviate il treno su un binario secondario dove c'è un solo operaio. Che fate?*

Questo scenario, noto come trolley problem, non è solo un rompicapo filosofico. È uno strumento potentissimo per far emergere intuizioni morali che altrimenti resterebbero implicite. E soprattutto: è una porta d'ingresso straordinaria per introdurre le teorie etiche normative in classe.

Questo paragrafo offre una mappa concettuale delle principali teorie etiche — deontologia, consequenzialismo, etica della virtù — insieme a strumenti narrativi e didattici pronti all'uso. Non un trattato accademico, ma una cassetta degli attrezzi: esperimenti mentali, storie, domande provocatorie da portare in aula per il primo framework di AI & Ethics Literacy.

2.1.1 Il consequenzialismo: conta solo il risultato?

Per il consequenzialismo, un'azione è giusta o sbagliata esclusivamente in base alle conseguenze che produce. Non esistono azioni intrinsecamente cattive: tutto dipende dagli effetti. È un'etica che guarda avanti, che valuta ex post.

La forma più nota è l'utilitarismo, formulato da Jeremy Bentham e John Stuart Mill: l'azione giusta è quella che massimizza la felicità (o minimizza la sofferenza) del maggior numero di persone.

In classe — il trolley problem classico. Un carrello ferroviario corre senza controllo verso cinque persone legate ai binari. Puoi tirare una leva per deviarlo su un binario laterale, dove ucciderà una sola persona. Che fai? La maggior parte delle persone tira la leva. Il ragionamento è consequenzialista: meglio una vittima che cinque.

La variante del ponte. Stesso scenario, ma ora sei su un ponte. Accanto a te c'è un uomo corpulento. L'unico modo per fermare il carrello è spingerlo giù: il suo corpo bloccherà il treno, salvando i cinque. Lo spingi? Matematicamente è identico: una vita contro cinque. Eppure quasi nessuno spingerebbe l'uomo. Perché? Questa variante fa emergere qualcosa che il puro calcolo consequenzialista non cattura: l'intuizione che usare una persona come mezzo sia profondamente sbagliato.

Punti di forza	Criticità
Il consequenzialismo è democratico (tutti contano ugualmente), pragmatico (guarda ai risultati concreti), e orientato al benessere collettivo.	Come calcoliamo le conseguenze? Possiamo davvero prevedere tutti gli effetti? E soprattutto: giustifica qualsiasi mezzo per un fine buono?

2.1.2 La deontologia: ci sono azioni sempre sbagliate?

Per la deontologia (dal greco déon, dovere), alcune azioni sono giuste o sbagliate in sé stesse, indipendentemente dalle conseguenze. Esistono doveri e divieti che valgono ex ante, prima di qualsiasi calcolo dei risultati.

Il padre di questa tradizione è Immanuel Kant. Il suo imperativo categorico prescrive: «Agisci soltanto secondo quella massima che puoi volere che divenga una legge universale» e «Tratta l'umanità sempre anche come fine, mai semplicemente come mezzo».

In classe — il violinista di Judith Jarvis Thomson. Ti svegli in ospedale collegato a un violinista famoso, in coma. Ha una rara malattia renale e solo tu hai il gruppo sanguigno compatibile. I medici ti spiegano: «La Società degli Amanti della Musica ti ha rapito e collegato a lui durante la notte. Se ti stacchi, morirà. Ma se resti collegato per nove mesi, guarirà completamente». Hai l'obbligo morale di restare?

Divieti assoluti e doveri prima facie
Una distinzione cruciale, introdotta dal filosofo W.D. Ross, è quella tra doveri assoluti (mai violabili, in nessuna circostanza) e doveri prima facie (validi «a prima vista», ma che possono decadere di fronte a doveri più pressanti). Mentire è sbagliato, ma mentire alle SS per nascondere una famiglia ebrea è giustificato da un dovere superiore.

Punti di forza	Criticità
La deontologia protegge i diritti individuali, impedisce di sacrificare innocenti per il «bene comune», e rispetta la dignità di ogni persona.	Può portare a conseguenze catastrofiche (non mentire mai, neanche ai nazisti?). E come stabiliamo quali principi sono davvero assoluti?

2.1.3 L'etica della virtù: che persona voglio essere?

L'etica della virtù sposta il focus: non «cosa devo fare?» ma «che persona devo essere?». Non si concentra su azioni o conseguenze, ma sul carattere dell'agente. Essere morali significa coltivare disposizioni stabili — le virtù — che ci permettono di percepire e fare la cosa giusta nelle diverse situazioni.

Questa tradizione risale ad Aristotele e alla sua Etica Nicomachea, ed è stata ripresa nel Novecento da filosofe come Elizabeth Anscombe, Philippa Foot e Alasdair MacIntyre.

In classe — il medico virtuoso. Immaginate due medici. Entrambi prescrivono la stessa terapia corretta. Ma il primo lo fa per dovere professionale, quasi meccanicamente. Il secondo lo fa perché si prende genuinamente cura del paziente, ne comprende la sofferenza, cerca di comunicare con empatia. Entrambi hanno fatto la cosa giusta. Ma sono moralmente equivalenti? Per l'etica della virtù, no.

La tradizione distingue quattro virtù cardinali: prudenza (saggezza pratica nel deliberare), giustizia (dare a ciascuno il suo), fortezza (coraggio di fronte alle difficoltà), temperanza (moderazione dei desideri).

Punti di forza	Criticità
Cattura aspetti della vita morale che le altre teorie trascurano (motivazioni, emozioni, formazione del carattere). È concreta, legata alla vita reale.	Quali sono le virtù? Chi lo decide? E quando due persone virtuose sono in disaccordo, come si risolve? L'etica della virtù fatica a fornire guide d'azione precise.

2.1.4 Confronto sinottico: tre sguardi sulla stessa azione

Scenario — il whistleblower. Maria lavora in un'azienda farmaceutica. Scopre che un farmaco in commercio ha effetti collaterali gravi non dichiarati. Se denuncia, salverà molte vite ma perderà il lavoro, l'azienda fallirà, centinaia di colleghi resteranno disoccupati, e la sua famiglia soffrirà economicamente. Che dovrebbe fare?

Teoria	Valutazione
Consequenzialismo	Bisogna calcolare: quante vite salverà la denuncia? Quanto danno causerà? Se il bilancio netto è positivo, Maria deve denunciare.
Deontologia	Maria ha il dovere di non ingannare. Ha anche doveri verso la famiglia. Il dovere di proteggere vite innocenti potrebbe prevalere come dovere prima facie.
Etica della virtù	Cosa farebbe una persona coraggiosa, onesta, giusta? Il coraggio richiede di affrontare rischi per cause giuste. La prudenza suggerisce come e quando agire.

2.1.5 Dall'etica teorica all'etica applicata

L'etica applicata nasce quando le teorie normative incontrano i problemi reali: medicina, ambiente, tecnologia, economia. Non si tratta di «applicare» meccanicamente principi astratti, ma di un dialogo tra teoria e casi concreti — un movimento circolare in cui l'esperienza interroga la teoria e la teoria illumina l'esperienza.

La bioetica è forse il campo più sviluppato dell'etica applicata. Nata negli anni '70, affronta questioni come: quando inizia la vita? Quando finisce? Possiamo manipolare geneticamente gli embrioni? Chi decide se «staccare la spina»?

2.1.6 I quattro principi di Beauchamp e Childress

Nel 1979, Tom Beauchamp e James Childress pubblicano *Principles of Biomedical Ethics*, un testo che diventerà il riferimento mondiale per la bioetica. Propongono quattro principi fondamentali, considerati prima facie, cioè validi "a prima vista" ma bilanciabili tra loro quando entrano in tensione.

Autonomia. Ogni persona ha il diritto di prendere decisioni informate sulla propria vita e sul proprio corpo. Il medico deve rispettare le scelte del paziente, anche quando non le condivide.

Non maleficenza. Primum non nocere — prima di tutto, non nuocere. Il medico deve evitare di causare danni al paziente.

Beneficenza. Il dovere di fare il bene, di agire nell'interesse del paziente, di promuoverne il benessere.

Giustizia. Distribuzione equa di risorse, benefici e oneri. Trattare casi simili in modo simile.

Questi quattro principi non si applicano in modo meccanico: entrano spesso in tensione tra loro (l'autonomia del paziente può confliggere con la sua beneficenza; la giustizia distributiva può limitare alcune richieste individuali). È proprio da questa tensione strutturale che nasce il dibattito bioetico contemporaneo tra etica della sacralità della vita ed etica della qualità della vita, oggetto del paragrafo successivo. L'applicazione in classe di queste distinzioni è poi ripresa nel § 2.1.8 con suggerimenti operativi.

2.1.7 Sacralità e qualità della vita

Etica della sacralità della vita. La vita umana è intangibile, ha valore in sé indipendentemente dalla sua «qualità». Esistono atti intrinsecamente sbagliati che nessuna circostanza può giustificare.

Etica della qualità della vita. Il valore della vita dipende anche dalla sua qualità — benessere, autonomia, dignità. Non esistono divieti assoluti: ogni situazione va valutata nel suo contesto.

2.1.8 Portare l'etica in classe: suggerimenti pratici

- **Partire dalle intuizioni:** gli esperimenti mentali fanno emergere intuizioni pre-teoriche. Lasciate che gli studenti si confrontino con il dilemma prima di introdurre i framework teorici.
- **Valorizzare il disaccordo:** l'etica non è matematica. Il disaccordo non è un fallimento — è il punto di partenza.
- **Collegare al vissuto:** gli studenti hanno già esperienza morale. Fateli riflettere su decisioni che hanno preso.

ESEMPIO DISCIPLINARE — SCIENZE — BIOETICA A SCUOLA

I quattro principi di Beauchamp e Childress funzionano benissimo come ponte con la biologia. Esempi utilizzabili in classe: un caso di screening genetico prenatale (autonomia vs. beneficenza), la donazione di organi (giustizia distributiva), l'obbligo vaccinale (autonomia individuale vs. beneficio collettivo). Ciascuno si presta a una simulazione di comitato etico in cui gli studenti, divisi in ruoli (paziente, medico, familiare, comitato), devono argomentare a partire da principi diversi.

2.1.9 Conclusione: l'etica come pratica del pensiero

Le teorie etiche non sono ricette da applicare meccanicamente. Sono strumenti per pensare meglio — per vedere aspetti che altrimenti sfuggirebbero, per argomentare con maggiore rigore, per comprendere chi la pensa diversamente.

In classe, il nostro compito non è produrre risposte preconfezionate ma coltivare la capacità di porsi domande, di ascoltare posizioni diverse, di argomentare senza aggredire. Formare cittadini capaci di pensiero morale è forse la missione più importante — e più difficile — che abbiamo.

Riepilogo — le tre domande fondamentali

Consequenzialismo: «Quali conseguenze produrrà questa azione?»
Deontologia: «È un'azione intrinsecamente giusta o sbagliata?»
Etica della virtù: «Cosa farebbe una persona virtuosa in questa situazione?»

Bibliografia essenziale

Beauchamp T.L., Childress J.F., *Principi di etica biomedica*, Le Lettere, Firenze 1999.
Foot P., *Virtues and Vices*, Oxford University Press, 2002.
Mori M., *Manuale di bioetica*, Le Lettere, Firenze 2010.
Rachels J., *Gli elementi della filosofia morale*, McGraw-Hill, Milano 2011.
Thomson J.J., "A Defense of Abortion", *Philosophy and Public Affairs*, 1971.

2.2 Fondamenti di Intelligenza Artificiale

> *Chiedete a dieci esperti «cos'è l'intelligenza artificiale?» e otterrete undici risposte diverse. Non è un caso: l'IA è un campo in rapida evoluzione, nato dall'intersezione di discipline diverse — matematica, informatica, filosofia, neuroscienze, linguistica — e il suo stesso oggetto di studio è sfuggente. Cos'è l'«intelligenza»? Quando una macchina può dirsi «intelligente»? Queste domande non hanno risposte univoche, e proprio per questo è fondamentale che studenti e docenti ne siano consapevoli.*

Questo paragrafo non pretende di offrire la definizione di IA, ma di fornire una mappa concettuale per orientarsi: le diverse definizioni storiche, i termini chiave (machine learning, deep learning, IA generativa), le classificazioni utili per l'analisi etica, le tecnologie principali e — crucialmente — i meccanismi attraverso cui i bias si insinuano nei sistemi algoritmici.

2.2.1 Il problema della definizione

Perché è così difficile definire l'IA? L'intelligenza artificiale è un campo dell'Informatica dedicato allo sviluppo di sistemi in grado di svolgere compiti che associamo a capacità cognitive umane: apprendere dai dati, riconoscere schemi, comprendere e utilizzare il linguaggio naturale, prendere decisioni. A causa di questa varietà di applicazioni non esiste una IA monolitica ma piuttosto un insieme di tecnologie adatte allo svolgimento specializzato di compiti diversi. Lo studio dei sistemi di IA è un campo interdisciplinare che integra l'informatica con numerose altre discipline, tra cui matematica, statistica, filosofia, linguistica.

Numerose tecniche sviluppate nell'ambito dell'intelligenza artificiale, in particolare quelle legate al machine learning, sono oggi ampiamente impiegate nella Data Science per l'analisi e la modellazione dei dati.

Nel loro celebre manuale *Artificial Intelligence: A Modern Approach* (2010), Stuart Russell e Peter Norvig hanno proposto una tassonomia che organizza le definizioni di IA — gli obiettivi e i prodotti della ricerca su IA — lungo due assi: pensare vs. agire e umanamente vs. razionalmente. Ne risultano quattro approcci fondamentali:

	Umanamente	**Razionalmente**
Pensare	Costruire macchine che svolgano attività di ragionamento replicando le modalità dell'umano: non solo ottenere il risultato corretto ma farlo nello stesso modo in cui lo farebbe un cervello umano.	Costruire macchine che rappresentino la conoscenza e facciano inferenze logiche secondo schemi formali, arrivando a conclusioni corrette. Spesso troppo rigida e costosa computazionalmente.
Agire	Creare macchine che eseguono attività e producono risultati tali da non consentire di distinguere una macchina da un essere umano (es. Test di Turing), ma seguendo ragionamenti che non replicano necessariamente modalità umane.	Creare sistemi che interagiscono con l'ambiente acquisendo informazioni e costruendo autonomamente una rappresentazione della realtà, per ottenere il risultato nel modo ottimale.

La distinzione è importante: chi cerca di replicare il modo in cui pensano gli umani (approccio cognitivo) fa un lavoro diverso da chi vuole semplicemente creare sistemi che ottengono risultati razionali (approccio ingegneristico). La maggior parte dei sistemi IA odierni — compresi i Large Language Models — rientra nel quadrante «agire razionalmente»: raggiungono obiettivi in modo efficace, senza un sistema di regole rigido e statico, e senza lo scopo di imitare la modalità di funzionamento del pensiero umano.

Una definizione operativa

Data la complessità del campo, è utile adottare una definizione di lavoro. Quella proposta dall'OCSE e ripresa dall'AI Act europeo è oggi un riferimento autorevole:

> **Definizione OCSE / AI Act**
> Un sistema di IA è un sistema basato su macchina che, per obiettivi espliciti o impliciti, deduce dagli input che riceve come generare output quali previsioni, contenuti, raccomandazioni o decisioni che possono influenzare ambienti fisici o virtuali. I diversi sistemi di IA variano nei loro livelli di autonomia e adattività dopo l'implementazione.

In termini più semplici: un'IA è un software creato dall'uomo che ne determina le impostazioni iniziali e gli obiettivi, con la capacità di adattare il proprio comportamento in base ai dati e all'esperienza, per massimizzare il raggiungimento di quegli obiettivi. Non è «intelligente» nel senso umano del termine — a oggi non esistono sistemi capaci di replicare tutte le nostre facoltà cognitive — ma può svolgere compiti specifici spesso meglio di noi.

2.2.2 Una classificazione degli agenti IA

Per orientare l'analisi etica, è utile distinguere i sistemi IA in base al loro livello di autonomia e alla complessità delle funzioni che svolgono. Dumouchel e Damiano, nel loro lavoro sull'empatia artificiale, propongono una tripartizione di agenti che nell'interazione con l'essere umano appaiono intelligenti e/o autonomi:

Tipo di agente	Descrizione	Esempi
Agenti analitici	Funzione computazionale semplice: fare calcoli, eseguire micro-azioni programmate. L'azione è circoscritta con conseguente riduzione dell'ambito della responsabilità.	Calcolatrice; filtri antispam; sistemi di raccomandazione semplici.
Agenti esecutivi	Funzioni più complesse, sempre sulla base di una programmazione, ma il campo di azione — e di responsabilità — si estende. Possono avere conseguenze significative sul mondo fisico.	Robot chirurgici; droni; veicoli a guida autonoma; sistemi diagnostici.
Agenti interattivi	IA con interfacce simili a quelle di animali o persone, con funzione sociale delicata. Sollevano questioni di empatia artificiale e attaccamento.	Robot-sitter (Paro, Pepper); assistenti vocali (Alexa, Siri); chatbot terapeutici.

Questa classificazione è particolarmente utile nel laboratorio didattico 4.2, dove gli studenti sono chiamati a distinguere un filtro antispam (agente analitico), un robot chirurgico (agente esecutivo) e un androide cinematografico (agente interattivo immaginario).

2.2.3 IA debole e IA forte

Narrow AI (IA debole)	AGI (IA forte)
Sistemi progettati per svolgere un compito specifico: riconoscere volti, tradurre testi, giocare a scacchi, generare immagini.	Sistemi ipotetici con capacità cognitive generalizzate, paragonabili o superiori a quelle umane.
Non «comprendono» in senso umano; sono ottimizzati per obiettivi precisi.	Potrebbero apprendere, ragionare, adattarsi a qualsiasi dominio, avere coscienza.
Tutta l'IA attualmente esistente è Narrow AI.	L'AGI non esiste (ancora?). È oggetto di speculazione e dibattito.

Questa distinzione è importante per evitare il «salto fantascientifico»: l'errore di attribuire ai sistemi IA attuali capacità che non hanno. La posizione prevalente nella comunità scientifica, nonché il paradigma adottato in questo volume, è che l'IA non «pensa» come noi, non «capisce», non ha «intenzioni» — anche se a volte sembra che le abbia. I modelli linguistici non possiedono capacità semantiche reali, ma operano attraverso sofisticate correlazioni statistiche. Il dibattito è aperto: esistono scuole di pensiero che inquadrano l'IA in modo diverso, attribuendole forme di comprensione o intenzionalità emergenti. Ai fini didattici, manteniamo un atteggiamento di scetticismo costruttivo: gli LLM sono strumenti potentissimi, ma l'apparenza di intelligenza non è intelligenza.

2.2.4 Le tecnologie chiave: dal Machine Learning ai LLM

Dal programma tradizionale al Machine Learning. Un programma tradizionale è una sequenza finita di istruzioni ben definite, scritte dal programmatore in un linguaggio comprensibile alla macchina, che eseguite in un ordine preciso consentono di risolvere un problema. Un programma tradizionale trasforma i dati iniziali (input) in risultati (output) in modo deterministico: al medesimo input corrisponde sempre lo stesso output. Ciò implica che l'onere di interpretare la realtà (scegliere gli aspetti da considerare, definire le loro relazioni e creare un sistema di regole che la rappresenti) ricada sul programmatore. Più è complessa la realtà che si vuole rappresentare più sarà difficile farlo con metodi tradizionali di programmazione.

Nel Machine Learning (ML), invece, l'interpretazione della realtà è "delegata" alla macchina. Il programmatore decide cosa il modello deve prevedere e sceglie quali aspetti della realtà osservare in quanto funzionali agli obiettivi. Inoltre, prepara i dati etichettati (esempi con la risposta nota) e fornisce le istruzioni per creare e addestrare il modello. Si lascia che l'algoritmo «impari da solo» il modo per riconoscere i pattern (schemi ricorrenti o relazioni tra i dati).

Reti neurali e Deep Learning. Le tecniche di ML sono molteplici. Nella fase di utilizzo alcune sono deterministiche e spiegabili — è possibile capire in modo chiaro e diretto come passare dall'input all'output — come ad esempio gli alberi decisionali. Altre introducono elementi casuali nel sistema e risultano non deterministiche. Tra queste ci sono quelle basate sulle reti neurali

artificiali: strutture non fisiche, ma strati matematici costituiti da centri di calcolo (neuroni artificiali) interconnessi, ispirati (molto liberamente) al cervello umano. Più strati ci sono, più la rete è «profonda».

Il Deep Learning è un sottoinsieme del Machine Learning che utilizza reti neurali artificiali profonde. Ha permesso progressi straordinari, ma ha anche un problema: le reti neurali, in particolare quelle profonde, sono «scatole nere» (black box). Esse funzionano, ma non è possibile descriverne in modo esplicito il funzionamento. È possibile estrarre dalla rete la sequenza di numeri che ha accompagnato il processo decisionale ma non trasformarla in un insieme esplicito di regole comprensibili. Questo solleva questioni etiche di trasparenza e spiegabilità.

IA generativa e Large Language Models. L'IA generativa è una classe di sistemi che non si limitano a classificare o prevedere, ma creano contenuti nuovi: testi, immagini, musica, codice. I Large Language Models (LLM) come GPT-4, Claude o Gemini sono il cuore di questa rivoluzione. Un LLM è, in sostanza, un sistema addestrato su enormi quantità di testo che impara a prevedere la parola successiva in una sequenza. Molti LLM sono oggi incorporati in sistemi più ampi capaci di gestire non solo testo ma anche immagini, audio e video: per questo ci si riferisce spesso ad essi come Large Multimodal Models.

Gli LLM sono «pappagalli stocastici»?
Bender e colleghi (2021) hanno definito gli LLM «pappagalli stocastici»: sistemi che «ricuciono insieme sequenze di forme linguistiche» apprese dai dati, senza comprenderne il significato. Il dibattito è aperto. Per fini didattici, è utile mantenere un atteggiamento di scetticismo costruttivo: gli LLM sono strumenti potentissimi, ma non «pensano» come noi.

2.2.5 Come nascono i bias negli algoritmi

Se il Machine Learning «impara» dai dati, cosa succede quando i dati sono distorti, incompleti o pregiudizievoli? Succede che il sistema impara i pregiudizi. L'IA non è neutrale, perché i dati su cui viene addestrata non sono neutrali.

Le fonti del bias. Il bias può insinuarsi in ogni fase del ciclo di vita di un sistema IA:

Fase	Come si genera il bias
Raccolta dati	Se i dati non sono rappresentativi (es. 80% volti maschili bianchi), il sistema funzionerà peggio per i gruppi sottorappresentati.
Etichettatura	Gli annotatori umani portano i propri pregiudizi. Etichette soggettive incorporano bias culturali.
Dati storici	Se il sistema è addestrato su decisioni passate discriminatorie, replicherà quelle discriminazioni.

Fase	Come si genera il bias
Progettazione	Le scelte dei programmatori (quali variabili considerare, come pesarle) incorporano assunzioni implicite.
Messa in opera	Un modello può sembrare equo in laboratorio ma produrre discriminazioni in contesti reali diversi da quelli di test.

Casi reali emblematici

- **Amazon Recruiting:** un sistema di selezione del personale penalizzava automaticamente i CV contenenti riferimenti al genere femminile, perché addestrato su 10 anni di assunzioni in un'azienda a predominanza maschile.
- **COMPAS:** un algoritmo usato nei tribunali USA per prevedere la recidiva assegnava punteggi di rischio più alti agli imputati afroamericani, anche a parità di altre condizioni.
- **Riconoscimento facciale:** sistemi con tassi di errore fino al 35% per le donne di pelle scura, contro meno dell'1% per gli uomini bianchi — perché addestrati su dataset sbilanciati.

La lezione è chiara: l'IA non è più «oggettiva» dei dati su cui viene addestrata. E poiché i dati sono prodotti da società con lunghe storie di discriminazione, i sistemi IA rischiano di perpetuare e amplificare quelle stesse discriminazioni — spesso in modo invisibile e difficile da contestare. Questo non è sempre dovuto a negligenza dei programmatori o a superficialità nella scelta dei dati di addestramento. A volte i dati di cui si dispone possiedono dei bias strutturali: per esempio, in alcuni casi si dispone di immagini prodotte da sistemi diagnostici quasi esclusivamente riferite a pazienti malati piuttosto che sani. Per questo motivo esistono filoni di ricerca che si occupano di studiare come mitigare questo bias strutturale dei dati.

Riepilogo — i concetti chiave

1. L'IA non è monolitica — esistono sistemi diversi con livelli diversi di autonomia (analitici, esecutivi, interattivi).
2. Tutta l'IA attuale è «debole» — nessun sistema esistente ha intelligenza generale; sono tutti specializzati.
3. Il Machine Learning impara dai dati — e se i dati sono biased, il sistema sarà biased.
4. Gli LLM non «comprendono» — sono potentissimi pattern-matcher, non pensatori.
5. I bias sono sistemici — si insinuano in ogni fase, dalla raccolta dei dati al deployment.

2.2.6 Amplificazione dei bias e cicli di retroazione

Un ulteriore elemento da considerare è la presenza di cicli di retroazione (feedback loop) nei sistemi di machine learning. Gli algoritmi vengono addestrati su dati prodotti da decisioni e comportamenti passati; una volta implementati, però, gli stessi sistemi influenzano le scelte degli utenti attraverso raccomandazioni, classificazioni e suggerimenti. Le interazioni degli utenti con queste

raccomandazioni generano nuovi dati che verranno utilizzati per addestrare nuovamente il modello stesso o altri modelli. Se i dati iniziali contenevano bias — per esempio stereotipi sociali o discriminazioni storiche — tali bias possono essere progressivamente rafforzati. L'algoritmo finisce così per apprendere da dati che contribuisce esso stesso a produrre, creando un processo di auto-rinforzo delle distorsioni.

Inoltre, gli algoritmi di ML producono essi stessi dei dati (per esempio etichette in un sistema di classificazione) che, con la presunzione di verità, possono essere usati come fonte per l'addestramento di un nuovo sistema. Se i dati iniziali erano affetti da bias si produrrà necessariamente un fenomeno di propagazione.

Proprio per questo il design etico di un sistema di IA non può limitarsi al momento del primo addestramento: richiede monitoraggi periodici, audit indipendenti e meccanismi espliciti per individuare e interrompere le spirali di rinforzo del bias. È un tema che riprendiamo nel § 2.3 parlando di accountability e AI Act.

Bibliografia essenziale

Bauer K., Heigl R., Hinz O., Kosfeld M., "Feedback Loops in Machine Learning: A Study on the Interplay of Continuous Updating and Human Discrimination", *Journal of the Association for Information Systems*, 25(4), 804–866, 2024.

Bender, E. M., Gebru, T., McMillan-Major, A., and et al (2021). On the dangers of stochastic parrots: Canlanguage models be too big? . In *Proceedings of FAccT* 2021.

Dumouchel P., Damiano L., *Vivere con i robot. Saggio sull'empatia artificiale*, Raffaello Cortina, Milano 2019.

Noble S.U., *Algorithms of Oppression*, NYU Press, New York, 2018.

O'Neil C., *Weapons of Math Destruction*, Penguin, 2016.

OECD, Recommendation of the Council on Artificial Intelligence, 2024.

Russell S., Norvig P., *Intelligenza artificiale: Un approccio moderno (vol. 1)*. Pearson, Torino, 2010.

ESEMPIO DISCIPLINARE — INFORMATICA

I concetti di questo paragrafo (reti neurali, feedback loop, bias) si prestano a lezioni-laboratorio su dataset aperti: per esempio il dataset MNIST delle cifre scritte a mano o dataset pubblici di volti. Gli studenti possono confrontare accuratezza per diverse sottopopolazioni e verificare empiricamente il concetto di bias strutturale. Un secondo laboratorio utile: far scrivere agli studenti un sistema a regole (if/else) e uno a tabella di apprendimento semplificata, per toccare con mano la differenza tra programmazione tradizionale e ML.

ESEMPIO DISCIPLINARE — MATEMATICA — OLTRE LA STATISTICA

Il bias nei dati è statisticamente un errore sistematico. Ma i concetti di questo capitolo si collegano anche ad algebra lineare (rappresentazioni vettoriali, embedding), analisi (gradiente e discesa del gradiente come ottimizzazione), geometria analitica (iperpiani di separazione, clustering). Un esperimento utile: dato un insieme di punti in R^2, gli studenti tracciano a mano

un "classificatore" lineare e discutono cosa succede quando la distribuzione dei punti è sbilanciata.

2.3 Etica dell'Intelligenza Artificiale

Tre modi di intendere l'educazione e l'IA

Prima di addentrarci nelle questioni etiche sollevate dall'intelligenza artificiale, è fondamentale chiarire quale sia il lavoro che questo manuale intende perseguire. Esistono infatti tre modi diversi di intendere il rapporto tra educazione (non semplice formazione ai contenuti) e IA:

Approccio	Descrizione
Educare l'IA	Riguarda la programmazione e l'addestramento dei sistemi IA. Si occupa di come "insegnare" alla macchina comportamenti etici. È il dominio dell'algoretica e della machine ethics.
Educare con l'IA	Riguarda l'uso dell'IA come strumento didattico: tutor intelligenti, sistemi adattivi, piattaforme di apprendimento personalizzato.
Educare all'IA	Riguarda la formazione delle persone a comprendere, valutare criticamente e usare responsabilmente l'IA. È il dominio dell'AI & Ethics Literacy: il nostro focus.

2.3.1 L'etica dell'IA è un'etica applicata?

L'intelligenza artificiale sta ridefinendo irrimediabilmente il ruolo dell'etica. Come osservano Pisano e colleghi, le teorie etiche sull'IA sono spesso a sfondo principialista fortemente sganciate da scenari concreti e dall'eterogeneità dei contesti in cui l'IA è applicata. Il rischio è duplice: da un lato, una confusione fra riflessioni etiche non sovrapponibili; dall'altro, l'idea di un'"etica dell'IA" che diventa mero slogan per operazioni di bluewashing.

> *Offrire un quadro normativo per l'Etica dell'IA utilizzando principi generali equivale a cercare di adottare la stessa deontologia per giornalisti, medici, insegnanti, psicologi. In alcuni campi si possono trovare principi trasversali, ma non sono sufficienti per rispondere alle sfide etiche delle discipline specifiche.* — Pisano et al., 2025

La proposta è dunque quella di un principialismo pratico: un approccio che attinge dal metodo dell'equilibrio riflessivo di Beauchamp e Childress, ma che si adatta al contesto specifico attraverso la phronesis — la saggezza pratica aristotelica.

Che cos'è la phronesis?

La phronesis (φρόνησις) è la "saggezza pratica" teorizzata da Aristotele nell'Etica Nicomachea: la virtù che ci permette di deliberare bene in situazioni concrete, dove non basta conoscere le regole ma occorre saperle applicare con giudizio. È la capacità del medico che, di fronte a due pazienti con la stessa diagnosi, sceglie terapie diverse perché tiene conto del contesto; del giudice che applica la legge senza ridurla a mera procedura; dell'insegnante che adatta il metodo allo studente reale che ha davanti. Nel nostro framework, la phronesis è ciò che impedisce all'etica dell'IA di diventare una checklist meccanica: i principi (autonomia, giustizia, trasparenza) orientano, ma è la saggezza pratica che li rende operativi nel caso specifico.

L'albero delle tradizioni dell'etica dell'IA

- **Etica sintetica** — si interroga sulla comparazione essere umano-macchina: l'IA può avere stati morali? Include l'algoretica e la machine ethics.
- **Etiche applicate** — analizzano le emergenze etiche sollevate dall'IA nei diversi contesti: educazione, medicina, informazione, ricerca, lavoro, ambiente.
- **Etica e diritto** — si concentra sull'elaborazione di quadri regolatori e linee guida istituzionali (AI Act europeo, linee guida OCSE, raccomandazioni UNESCO).

Sulla natura dell'IA: strumento, arbitro o infrastruttura morale?

Prima di esaminare i principi e le normative che cercano di regolare l'IA, è utile fare un passo indietro e chiedersi: come pensiamo al ruolo che essa svolge nelle nostre scelte?

Esistono due modi di rispondere, entrambi diffusi ma altrettanto problematici. Il primo considera l'IA come uno strumento neutro: un mezzo che ci aiuta a fare cose che faremmo comunque, solo più in fretta o con maggiore precisione. Il secondo la considera come un sistema oggettivo, capace di risolvere problemi complessi, anche morali, in modo più imparziale di quanto riesca a fare un essere umano (vi avevamo accennato prima parlando dell'ipotesi AGI). Come abbiamo visto nel paragrafo sui bias algoritmici, però, nessun sistema di IA è davvero neutro né davvero imparziale: ogni algoritmo riflette le scelte di chi lo ha costruito e i limiti dei dati su cui è stato addestrato.

C'è un terzo modo di guardare all'IA, più utile per chi vuole usarla in modo consapevole: considerarla non come uno strumento passivo né come un arbitro infallibile, ma come una struttura che modifica le condizioni entro cui prendiamo decisioni, rendendo alcune scelte più facili, altre più difficili, certi ragionamenti più visibili e altri meno accessibili. Trasparenza e possibilità di critica diventano così le condizioni minime perché l'IA possa davvero essere al servizio delle persone; in altri termini, la legittimità morale di un sistema di IA dipende dal fatto che le sue scelte possano essere comprese, valutate e contestate da chi ne è coinvolto.

2.3.2 I principi per un'IA affidabile

Nel 2019, il Gruppo di Esperti di Alto Livello sull'IA della Commissione Europea ha pubblicato le Ethics Guidelines for Trustworthy AI. L'IA affidabile deve essere lecita, etica e robusta.

Requisito	Significato
1. Azione e supervisione umana	L'IA deve supportare l'autonomia umana, non sostituirla.
2. Robustezza e sicurezza	I sistemi devono essere affidabili, sicuri, resistenti agli attacchi.
3. Privacy e governance dei dati	I dati personali devono essere protetti e usati in modo trasparente.
4. Trasparenza	I processi decisionali dell'IA devono essere comprensibili e spiegabili.
5. Diversità e non discriminazione	L'IA deve evitare pregiudizi ingiusti e garantire accessibilità universale.
6. Benessere sociale e ambientale	L'IA deve contribuire positivamente alla società e all'ambiente.
7. Accountability	Devono esistere meccanismi per attribuire responsabilità e permettere ricorsi.

L'AI Act europeo: dal principio alla norma

Nel 2024, l'Unione Europea ha approvato l'AI Act (Regolamento UE 2024/1689), il primo quadro normativo al mondo sull'intelligenza artificiale. L'AI Act adotta un approccio basato sul rischio, classificando i sistemi IA in quattro categorie di rischio crescente. La trattazione che segue è introduttiva e orientata alla didattica: per gli aspetti normativi di dettaglio si rimanda al testo del Regolamento.

Livello di rischio	Esempio	Conseguenza normativa
Inaccettabile	Sistemi di social scoring governativo che classificano i cittadini su base comportamentale; manipolazione cognitiva di minori attraverso giocattoli intelligenti.	Divieto assoluto di immissione sul mercato europeo.
Alto	Sistemi di screening dei curriculum vitae per la selezione del personale; sistemi di ammissione o valutazione negli istituti di istruzione; diagnostica medica automatizzata.	Requisiti stringenti: valutazione del rischio, dataset rappresentativi, documentazione tecnica, supervisione umana, accuratezza e robustezza, conformità ex ante.

Livello di rischio	Esempio	Conseguenza normativa
Limitato	Chatbot che interagiscono con gli utenti; sistemi di generazione di contenuti audiovisivi ("deepfake"); assistenti vocali.	Obblighi di trasparenza: l'utente deve sapere di stare interagendo con un sistema di IA e che il contenuto è generato o manipolato.
Minimo o nullo	Filtri antispam della posta elettronica; motori di raccomandazione di videogiochi; traduttori automatici generalisti.	Nessun obbligo specifico oltre alla disciplina ordinaria. È incoraggiata l'adesione volontaria a codici di condotta.

In classe, il quadro dei quattro livelli si presta molto bene a un'attività in cui gli studenti ricevono un mazzo di "carte" con sistemi di IA reali e devono collocare ciascuna nella colonna giusta, argomentando la scelta. È un modo efficace per far scoprire che la classificazione non è ovvia: un assistente scolastico che valuta i compiti può essere rischio limitato o rischio alto a seconda dell'uso.

2.3.3 Casi concreti: le sfide etiche dell'IA oggi

La "schiavitù del clic": il lavoro invisibile dietro l'IA. Il sociologo Antonio Casilli, nel suo *Schiavi del clic* (2020), svela il "lato oscuro" dell'intelligenza artificiale: dietro i sistemi apparentemente automatici c'è un enorme lavoro umano, spesso invisibile, mal pagato e precario. Sono le persone che etichettano immagini, moderano contenuti, verificano traduzioni — le quali rendono possibile l'addestramento dei grandi modelli. In classe, questo punto si presta a una riflessione congiunta con storia, diritto ed economia: chi sono, dove si trovano, quali condizioni di lavoro hanno?

I costi ambientali dell'IA. I grandi modelli linguistici richiedono enormi risorse computazionali. Come osserva Kate Crawford in *Atlas of AI*, l'intelligenza artificiale non è "nel cloud": è ancorata a infrastrutture materiali enormi — data center, acqua per il raffreddamento, terre rare per i chip — con costi ambientali che ricadono spesso su comunità già vulnerabili.

Bias e discriminazioni: quando l'algoritmo è ingiusto. Gli algoritmi non sono neutrali. Riflettono i pregiudizi presenti nei dati di addestramento e nelle scelte di chi li progetta. Casi emblematici li abbiamo già visti nel § 2.2.5.

2.3.4 Conclusione: la sfida educativa

L'etica dell'IA non è un lusso accademico: è una competenza di cittadinanza. Il compito educativo è duplice: da un lato, fornire agli studenti gli strumenti concettuali per analizzare criticamente l'IA; dall'altro, coltivare le disposizioni — curiosità, scetticismo costruttivo, responsabilità — che permettono di abitare un mondo tecnologico senza esserne dominati.

C'è però una domanda che vale la pena porsi prima di concludere: può uno strumento che tende a dare risposte preconfezionate aiutarci davvero a pensare in modo autonomo?

Come abbiamo visto nella parte dedicata ai bias algoritmici, i sistemi di IA non sono neutri: incorporano valori, priorità e assunti che derivano dai dati su cui sono stati addestrati e dalle scelte di chi li ha progettati. Questo vale anche quando l'IA viene usata in classe. Studi recenti mostrano che le persone tendono ad adeguarsi alle indicazioni di un sistema di IA anche in situazioni che richiederebbero una valutazione personale; non perché abbiano ragionato e concordato, ma perché il sistema trasmette una forma di autorità difficile da mettere in discussione (Fabre et al., 2024). Un sistema di IA che fornisce spiegazioni personalizzate, valutazioni automatizzate e percorsi già tracciati può riprodurre proprio questa dinamica, in modo ancora più invisibile.

Questo significa che il valore dell'IA dipende da come viene usata: se invita a ragionare o sostituisce il ragionamento, se mostra i propri limiti o si presenta come infallibile, se lascia spazio al dubbio o chiude le domande con una risposta ottimale. Formare cittadine e cittadini capaci di pensiero critico, oggi, significa anche formarli a usare queste tecnologie senza delegare ad esse il compito di pensare.

Riepilogo — le tre domande per l'etica dell'IA

1. Chi beneficia e chi paga? (lavoro invisibile, costi ambientali, distribuzione dei vantaggi).
2. Chi decide e come? (trasparenza, accountability, supervisione umana).
3. Chi viene escluso? (bias, discriminazioni, accessibilità).

Bibliografia essenziale

Casilli A.A., *Schiavi del clic: Perché lavoriamo tutti per il nuovo capitalismo*, Feltrinelli, Milano 2020.

Crawford K., *Atlas of AI*, Yale University Press, New Haven, 2021.

Fabre, C., et al., "Making moral decisions with artificial agents as advisors. A fNIRS study", Computers in Human Behavior: Artificial Humans 2/2024.

Floridi L. et al., "AI4People — An Ethical Framework for a Good AI Society", Minds & Machines 28, 2018.

High-Level Expert Group on AI, Ethics Guidelines for Trustworthy AI, European Commission, 2019.

Regolamento (UE) 2024/1689 del Parlamento Europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (AI Act).

Tiribelli S., *Identità personale e algoritmi,* Carocci, Roma 2024.

3. Metodologia, strumenti e risorse didattiche

Rossana Adele Rossi, Antonella Valenti, Nunziatina Sanfilippo, Maria Concetta Carruba, Dario Lombardi

3.1 La Philosophical Enquiry[1] come metodologia per l'AI & Ethics Literacy

> *Un docente entra in classe con un dilemma: «Un'auto a guida autonoma, in una situazione di emergenza, deve scegliere se investire un pedone anziano o deviare causando la morte del suo passeggero. Cosa dovrebbe fare?». Invece di dare una risposta, invita gli studenti a formulare le loro domande. In venti minuti, l'aula si trasforma: non c'è più un insegnante che spiega e studenti che ascoltano, ma una comunità che indaga insieme.*

Le origini: da Matthew Lipman alla Philosophical Enquiry

Negli anni Settanta, il filosofo americano Matthew Lipman fondò il movimento Philosophy for Children (P4C), una metodologia didattica che trasforma la classe in una «comunità di ricerca». Dopo la sua rielaborazione da parte di Peter Worley in ambito britannico, la P4C è conosciuta oggi come Philosophical Enquiry (PhiE): una pratica dialogica che non riguarda esclusivamente i più piccoli, ma adolescenti e adulti.

Il pilot study condotto nelle scuole secondarie italiane (Pisano, Crispini, Rossi, 2026) ha confermato l'efficacia della PhiE per l'AI & Ethics Literacy. La metodologia non trasmette conoscenze dall'esterno: guida gli studenti a costruirle collettivamente attraverso il dialogo, un processo che si rivela particolarmente potente per affrontare questioni etiche complesse.

PhiE in 60 secondi

1. Si offre uno stimolo (testo, video, caso, immagine).
2. Gli studenti formulano le proprie domande.
3. Il gruppo sceglie una domanda da esplorare.
4. Il dialogo è facilitato — non guidato — dal docente.
5. Si chiude con un debriefing metacognitivo: cosa abbiamo scoperto? cosa resta aperto?

[1] Un ringraziamento per l'attività di diffusione e formazione sul tema al progetto *Inventio:* https://site.unibo.it/aion/it/ricerca/inventio.

I principi pedagogici della PhiE

1. La comunità di ricerca. Il cambiamento di paradigma non è banale: richiede ai docenti di rinunciare al ruolo di «depositari del sapere» e di diventare facilitatori del pensiero critico. Gli studenti imparano a pensare insieme, non l'uno contro l'altro.

2. Il caring thinking. Il «pensiero che si prende cura» è una dimensione trasformativa: pensare bene richiede non solo rigore logico ma anche attenzione, rispetto, empatia per le posizioni altrui.

3. Le 4C del pensiero. Pensiero critico (analisi, valutazione, argomentazione), creativo (immaginazione, innovazione, ipotesi), collaborativo (ascolto, costruzione comune, dialogo) e caring (empatia, responsabilità, cura).

Le pratiche dialogiche: tre metodologie a confronto

Approccio	Descrizione
Debate	Due o più team sostengono posizioni contrapposte, spesso predeterminate. L'obiettivo è vincere. Sviluppa competenze argomentative ma può ridursi a un esercizio di retorica senza trasformazione del pensiero.
Dialogo socratico	Ispirato al metodo di Socrate, procede per domande che destrutturano le false certezze (la maieutica). Richiede un facilitatore esperto, capace di porre domande che stimolino la riflessione senza manipolare le risposte.
Philosophical Enquiry	La metodologia sviluppata da Peter Worley (2010), erede del lavoro di Lipman, Sharp e Splitter. Combina elementi dell'approccio socratico con una più precisa scaffoldatura che la rende adatta a contesti scolastici diversi.

Gli strumenti operativi della PhiE

La macchina dell'«If-ing». Una tecnica di facilitazione che usa domande ipotetiche per aprire spazi di indagine: «Se tutti usassero sempre l'IA per decidere, cosa succederebbe?», «Se l'IA prendesse decisioni migliori di noi, saremmo obbligati a seguirle?».

Il cameriere silenzioso (silent waiter). La strategia di non-interventismo del facilitatore. Come un bravo cameriere che serve senza farsi notare, il docente facilita senza dominare: pone domande, riformula, protegge lo spazio del dialogo, ma non impone risposte.

Il mito in classe: Arianna, il Colosso, Medusa

Un elemento distintivo del nostro approccio è l'uso del mito come dispositivo generativo di domande. I miti portano «fuori» dallo schema abituale e permettono di affrontare l'IA con uno sguardo obliquo.

Arianna e il filo. Il filo che permette a Teseo di uscire dal labirinto. Un LLM — un assistente virtuale — può essere il nostro filo di Arianna nei problemi complessi?

Il Colosso di Rodi. La meraviglia tecnica che cade al primo terremoto. Metafora della fragilità delle infrastrutture digitali. Quanto siamo dipendenti?

Medusa e lo specchio. Perseo sconfigge Medusa guardandola attraverso uno specchio. L'IA è uno specchio dei nostri bias sociali: vedendola, vediamo noi stessi.

3.2 Un curriculum verticale per il biennio e il triennio

La proposta curricolare che segue articola l'AI & Ethics Literacy su tutti e cinque gli anni della scuola secondaria di secondo grado. Non è un programma rigido: è una mappa di obiettivi progressivi e di connessioni disciplinari da adattare al contesto di ogni istituto.

Rispetto alla bozza originaria, la tabella è stata riarticolata specificando, per ogni classe, gli obiettivi di apprendimento distinti in tre colonne: saperi (cosa conoscere), saper-fare (cosa saper svolgere operativamente), consapevolezze (quale disposizione riflessiva coltivare). L'obiettivo è permettere al docente di collegare ciascuna attività del Capitolo 4 a un obiettivo preciso del curricolo.

Classe / Focus	Saperi (conoscere)	Saper-fare	Consapevolezze
I — Introduzione all'IA	Cos'è l'IA: definizioni, storia breve, esempi quotidiani. Concetto di algoritmo. Differenza tra programma tradizionale e sistema che apprende.	Riconoscere la presenza dell'IA nella propria giornata; descrivere un algoritmo con parole proprie; formulare domande di fronte a un sistema sconosciuto.	L'IA non è magia; non tutto ciò che è automatico è intelligente; è utile sospendere il giudizio prima di decidere se fidarsi.
II — Dati, bias, rappresentazione	Concetti di dato, dataset, rappresentazione statistica di una popolazione, bias di campionamento e di etichettatura.	Leggere una tabella di dati; individuare sotto/sovra-rappresentazioni; progettare un piccolo sondaggio rappresentativo.	I numeri non sono neutri; chi produce il dato orienta la conoscenza; la trasparenza è una questione etica oltre che tecnica.
III — Teorie etiche e dilemmi	Consequenzialismo, deontologia, etica della virtù; quattro principi bioetici; dilemmi	Argomentare una posizione etica nei termini di più teorie; mappare un dilemma	Disagio di fronte alla pluralità; riconoscere che il disaccordo etico

Classe / Focus	Saperi (conoscere)	Saper-fare	Consapevolezze
	classici (trolley problem, violinista).	concreto su un caso scolastico; scrivere un breve saggio comparativo.	non è confusione ma ricchezza; valore della deliberazione.
IV — Etica dell'IA applicata	AI Act (quattro livelli di rischio) con esempi; principi dell'IA affidabile; casi di bias algoritmici reali; lavoro invisibile dietro l'IA.	Classificare sistemi di IA per livello di rischio; smontare un caso di cronaca distinguendo piano tecnico, sociale ed etico; condurre un audit etico scolastico.	La legge è un tentativo storicamente situato; la tecnologia ha costi materiali e umani che spesso non vediamo; responsabilità distribuita.
V — Cittadinanza digitale e progetto	Governance dell'IA, GDPR, ruolo dei diversi stakeholder (policy-maker, sviluppatori, utenti); philosophia come metodo.	Progettare e realizzare un prodotto di AI & Ethics Literacy (video, podcast, sito, intervento pubblico) destinato a un'audience non specialistica.	Sentirsi stakeholder; saper comunicare con chi non possiede il vocabolario tecnico; disposizione a continuare a formarsi.

Suggerimenti per il docente di sostegno (BES)

ESEMPIO DISCIPLINARE — SOSTEGNO E INCLUSIONE

Ciascuno dei cinque livelli si presta a essere articolato con strumenti compensativi: schede sintetiche con icone, organizer grafici, mappe concettuali preformate, tempi dilatati per le attività dialogiche. La PhiE in particolare, con i suoi turni di parola strutturati e le sue domande aperte, offre naturalmente opportunità a studenti con difficoltà linguistiche o relazionali. Nei laboratori del Capitolo 4 è indicata, dove rilevante, la variante breve (50 min) che può essere ulteriormente semplificata per classi con ampia presenza di studenti BES.

3.3 Philosophical practices e cultura della domanda

La PhiE non è un metodo isolato: si inserisce nella più ampia tradizione delle pratiche filosofiche (Philosophical Practices), un movimento che dagli anni Ottanta rivendica per la filosofia un ruolo operativo nella vita quotidiana — consulenza filosofica, café philosophique, filosofia in azienda,

filosofia clinica. Queste pratiche condividono un presupposto: la filosofia non è solo una disciplina accademica, ma un esercizio del pensiero che può — e deve — uscire dalle università.

Per l'insegnante, adottare uno sguardo "filosofico-pratico" significa: valorizzare le domande degli studenti più delle risposte già pronte; rendere esplicito il metodo (domanda, argomentazione, contro-esempio, revisione); coltivare la cultura della domanda come forma di cittadinanza attiva. In questo orizzonte, la PhiE descritta in 3.1 e il curriculum in 3.2 non sono un corpo estraneo al programma disciplinare ma un suo moltiplicatore.

Questa cultura della domanda diventa decisiva proprio quando nella classe entra l'IA generativa. La diffusione degli LLM rende strutturalmente fragile una didattica centrata sulla performance — ricevere, memorizzare e restituire contenuti — che un prompt ben formulato può cortocircuitare in pochi secondi: occorre spostare il baricentro dalla didattica della performance a quella del processo. Le evidenze empiriche confermano la posta in gioco. Uno studio del MIT (Kosmyna et al., 2025) ha rilevato che chi scriveva saggi delegando a ChatGPT mostrava il minor coinvolgimento cognitivo e la minore ritenzione dei contenuti prodotti; al contrario, i chatbot progettati con domande socratiche stimolano una riflessione critica più profonda e una maggiore autonomia metacognitiva (Favero et al., 2024). La differenza non risiede nello strumento, ma nelle procedure pedagogiche che ne governano l'impiego.

In questa prospettiva l'IA non è il delegato a cui affidare passivamente il lavoro cognitivo, ma l'espansione algoritmica di capacità tipicamente umane — connettere, generare, immaginare. È quanto mostra il modello *IdeA* (Lombardi, 2026), che reinterpreta le quattro fasi della retorica classica posizionando gli LLM come partner cognitivi anziché generatori passivi di contenuto e preservando l'agency dello studente attraverso dispositivi metacognitivi: pilotato su 105 studenti universitari, ha documentato percezioni favorevoli soprattutto sulla consapevolezza critica nell'uso dell'IA. Qui si salda il senso più profondo della PhiE (§ 3.1): il valore pedagogico non risiede nelle risposte che si ottengono, ma nella qualità delle domande che si impara a porre. Un LLM usato in chiave socratica — progettato non per rispondere, ma per restituire contro-esempi, domande e formulazioni alternative — non delega il pensiero alla macchina, ma usa la macchina per creare le condizioni in cui il pensiero degli studenti è messo alla prova.

Tutto ciò chiede una formazione docente che non si riduca alla tecnica del prompting. Richiede una postura critica e filosofica, la capacità di integrare gli LLM in una progettazione inclusiva — secondo i principi dello *Universal Design for Learning* (CAST, 2018) — e lungo l'intero ciclo di progettazione didattica. Il docente formato a questo sguardo non è colui che usa l'IA, ma colui che la governa: ne conosce i limiti, ne valuta criticamente gli output, la piega agli obiettivi formativi invece di lasciarsi piegare da essa. *La sfida, in altri termini, non è adattare la didattica all'IA: è formare docenti capaci di adattare l'IA alla didattica.*

Bibliografia essenziale

CAST, *Universal Design for Learning Guidelines, Version 2.2*, CAST, Wakefield, 2018.

Favero L., Pérez-Ortiz J., Käser T., Oliver N., *Enhancing Critical Thinking in Education by Means of a Socratic Chatbot*, arXiv:2409.05511, 2024.

Kosmyna N. et al., *Your Brain on ChatGPT: Accumulation of Cognitive Debt When Using an AI Assistant for Essay Writing Task*, arXiv:2506.08872, 2025.

Lipman M., *Thinking in Education*, Cambridge University Press, Cambridge, 2003.

Lombardi D., *Artificial Intelligence for Territorial Innovation: An Empirical Evaluation of the IdeA Model*, ICS Exchange, 2026.

Sanfilippo N., *Esistere è interrogare: la filosofia come possibilità. Condizione umana e pratiche filosofiche per il nostro tempo*, Diogene Multimedia, Bologna, 2026.

Splitter L.J., Sharp A.M., *Teaching for Better Thinking: The Classroom Community of Enquiry*, Acer, Melbourne 1995.

Worley P., *The If Machine: Philosophical Enquiry in the Classroom*, Bloomsbury, London 2010.

4. Schede operative: esempi didattici

Antonella Valenti, Rossana Adele Rossi, Emiliana Murgia, Nunziatina Sanfilippo, Maria Concetta Carruba, Giorgia Nieddu, Vojko Strahovnik, Luca Tenneriello, Stephen Ferst.

Questa sezione raccoglie otto schede operative pronte per essere utilizzate in classe. Ciascuna scheda è pensata come un laboratorio autonomo, ma l'insieme disegna un percorso coerente: dai dilemmi morali classici (4.1) all'uso consapevole dell'IA generativa (4.5), dalla rappresentazione statistica corretta dei dati (4.6) alla governance democratica delle tecnologie (4.8).

Come usare le schede: tempi flessibili

Ciascuna scheda è progettata per una sessione estesa da 90 minuti, ideale per un modulo di potenziamento, un'ora di Educazione civica o una giornata tematica. Per chi opera nella scansione ordinaria dell'ora di lezione, ciascuna scheda riporta anche una Variante breve (50 minuti) che mantiene il focus concettuale essenziale rinunciando ad alcune fasi di approfondimento. Le due versioni non sono gerarchizzate: anche la variante breve produce apprendimenti significativi se il docente è disposto a chiudere su ciò che è emerso piuttosto che completare la scheda integrale.

4.1 Il Trolley Problem: etica normativa e intuizioni morali

Voce	Descrizione
Obiettivo generale	Introdurre le teorie etiche normative (consequenzialismo, deontologia, etica della virtù) attraverso un esperimento mentale classico.
Classi di riferimento	III, IV, V — licei e istituti tecnici.
Discipline coinvolte	Filosofia, Storia, Educazione civica, Letteratura italiana.
Durata standard	90 minuti (1 ora e mezza).
Variante breve	50 minuti (v. più avanti).
Prerequisiti	Nessuno specifico; utile una introduzione alla differenza tra "bene", "giusto", "utile".
Modalità	Frontale breve + lavoro in piccoli gruppi + plenaria.

Fasi — versione standard (90 min)

Fase	Tempo	Attività
1. Presentazione scenario	10 min	Il docente racconta il caso classico del tram in corsa. Nessuna risposta del docente alle domande dei ragazzi.
2. Intuizione individuale	10 min	Ogni studente scrive anonimamente la propria risposta e una breve giustificazione (3-4 righe).
3. Raccolta e discussione	20 min	Si mostrano le distribuzioni delle risposte (tipicamente 80-90% "tiro la leva"). Si raggruppano le giustificazioni in insiemi.
4. Variante del ponte	15 min	Il docente introduce la variante (spingere l'uomo corpulento). Nuova votazione. Lo shift è sempre netto.
5. Introduzione teorie	20 min	Sintesi delle tre famiglie (consequenzialismo, deontologia, virtù). Gli studenti ri-etichettano le proprie giustificazioni.
6. Debriefing metacognitivo	15 min	Cosa abbiamo scoperto del nostro modo di decidere? Quale teoria ci somiglia di più? Scriviamo un post-it conclusivo.

Variante breve (50 min)

Fase	Tempo	Attività
1. Scenario + intuizione	15 min	Presentazione dilemma, scrittura anonima individuale della risposta e di una riga di giustificazione.
2. Discussione e variante	15 min	Plenaria rapida sulle risposte; introduzione della variante del ponte; nuova posizione.
3. Micro-teoria	10 min	Presentazione sintetica (mini-lezione di 10 minuti) delle tre famiglie etiche con un solo esempio ciascuna.
4. Chiusura riflessiva	10 min	Ogni studente scrive su un post-it quale teoria sente più vicina alla propria intuizione e perché.

Materiali

Slide con lo scenario del tram e la variante del ponte · post-it anonimi · una sintesi di una pagina delle tre teorie (se si adotta la variante breve).

Valutazione

Valutazione formativa: qualità delle giustificazioni scritte (chiarezza, presenza di un appiglio teorico, apertura a obiezioni). Nessuna risposta è "giusta" in sé.

Note per il docente

Attenzione a non suggerire la "risposta giusta": l'efficacia della scheda sta nella sospensione del giudizio del docente. Se qualche studente è in difficoltà emotiva di fronte al dilemma, offrire sempre la possibilità di non rispondere.

4.2 Agenti IA: analitici, esecutivi, interattivi

Voce	Descrizione
Obiettivo generale	Classificare sistemi di IA in tre tipologie (analitici, esecutivi, interattivi) e discutere le implicazioni etiche di ciascuna.
Classi di riferimento	II, III, IV — licei e istituti tecnici.
Discipline coinvolte	Informatica, Filosofia, Scienze.
Durata standard	90 minuti.
Variante breve	50 minuti.
Prerequisiti	§ 2.2.2 del manuale.
Modalità	Piccoli gruppi + gallery walk + plenaria.

Fasi — versione standard (90 min)

Fase	Tempo	Attività
1. Brainstorming	10 min	Ogni gruppo elenca su un foglio tutti i sistemi di IA che usa in una giornata.
2. Mini-lezione	15 min	Il docente introduce la tripartizione Dumouchel-Damiano con esempi.
3. Classificazione	25 min	I gruppi ricevono un mazzo di "carte-sistema" (filtro antispam, traduttore, robot chirurgico, Alexa, Waymo, ChatGPT, Paro, ecc.) e le dispongono in tre colonne, argomentando.

Fase	Tempo	Attività
4. Gallery walk	15 min	I gruppi girano tra i cartelloni; annotazioni di disaccordo e domande.
5. Plenaria	15 min	Confronto sulle classificazioni dubbie: ChatGPT è analitico o interattivo? Un drone militare autonomo?
6. Mini-saggio	10 min	Ciascuno sceglie un agente e scrive quali questioni etiche specifiche solleva.

Variante breve (50 min)

Fase	Tempo	Attività
1. Mini-lezione + esempi	15 min	Presentazione della tripartizione con esempi; il brainstorming iniziale è sostituito da una domanda rapida collettiva.
2. Classificazione a gruppi	20 min	Stessa attività di classificazione con le carte, ma su un sottoinsieme ridotto (6 carte).
3. Plenaria focalizzata	15 min	Discussione concentrata su 2-3 casi dubbi scelti dal docente; chiusura con una domanda generativa affidata per casa.

Materiali

Mazzo di carte-sistema (predisposto dal docente, circa 12 per la versione standard e 6 per la breve); fogli A3 con tre colonne; post-it colorati; pennarelli.

4.3 PAIA all'opera: analisi di un sistema reale

Voce	Descrizione
Obiettivo generale	Applicare il framework PAIA a un sistema di IA realmente utilizzato dagli studenti.
Classi di riferimento	III, IV, V.
Discipline coinvolte	Informatica, Filosofia, Educazione civica, Italiano.
Durata standard	90 minuti.
Variante breve	50 minuti.

Voce	Descrizione
Prerequisiti	§ 1.2 del manuale.
Modalità	Piccoli gruppi + produzione di un poster + plenaria.

Fasi — versione standard (90 min)

Fase	Tempo	Attività
1. Richiamo PAIA	10 min	Breve ripasso delle quattro dimensioni; un esempio comune (es. un feed social).
2. Scelta del sistema	10 min	Ogni gruppo sceglie un sistema che usa nella vita quotidiana.
3. Scheda PAIA	30 min	Il gruppo compila la scheda con le quattro domande (una per dimensione) e raccoglie evidenze.
4. Poster	20 min	Ogni gruppo produce un poster A3 con la mappa PAIA del proprio sistema.
5. Galleria e feedback	15 min	Gallery walk con post-it di feedback.
6. Chiusura	5 min	Sintesi del docente e domanda generativa: "Cosa cambierei in questo sistema?".

Variante breve (50 min)

Fase	Tempo	Attività
1. Richiamo + scelta	10 min	Ripasso del PAIA in 5 minuti; ogni gruppo sceglie velocemente un sistema.
2. Scheda PAIA rapida	25 min	Compilazione della scheda a quattro domande con tempi stringenti (6 minuti per dimensione).
3. Micro-presentazione	15 min	Ogni gruppo riferisce in 90 secondi una sola dimensione su cui ha avuto l'insight più forte.

Materiali

Scheda PAIA (una griglia A4 con le quattro domande: dove è pervasivo? dove riduce la mia autonomia? cosa non vedo? come cambia nel tempo?); fogli A3; pennarelli; post-it.

4.4 Costruire un'IA affidabile: casi dal mondo reale

Voce	Descrizione
Obiettivo generale	Classificare sistemi IA secondo i quattro livelli di rischio dell'AI Act europeo e discutere criticamente la classificazione.
Classi di riferimento	IV, V.
Discipline coinvolte	Filosofia, Diritto/Economia, Educazione civica, Informatica.
Durata standard	90 minuti.
Variante breve	50 minuti.
Prerequisiti	§ 2.3.2 del manuale.
Modalità	Piccoli gruppi + role-play di comitato etico.

Fasi — versione standard (90 min)

Fase	Tempo	Attività
1. Brief normativo	15 min	Presentazione dei quattro livelli di rischio con un esempio per livello.
2. Distribuzione casi	10 min	Ogni gruppo riceve 3 casi reali (tratti da cronaca) e deve classificarli.
3. Classificazione	25 min	Discussione interna; giustificazione scritta della classificazione scelta.
4. Ruoli in comitato	25 min	Role-play: ogni gruppo si trasforma in comitato etico aziendale chiamato a decidere se mettere sul mercato europeo un sistema controverso.
5. Plenaria e voto	15 min	Confronto tra comitati; votazione finale; analisi dei disaccordi.

Variante breve (50 min)

Fase	Tempo	Attività
1. Brief normativo	15 min	Introduzione dei quattro livelli con un solo esempio per livello; consegna della tabella sintetica.

Fase	Tempo	Attività
2. Classificazione di gruppo	20 min	Ciascun gruppo classifica 2 casi (invece di 3) e prepara la giustificazione.
3. Confronto rapido	15 min	Plenaria focalizzata su un caso controverso a scelta del docente; chiusura con domanda a casa.

Materiali

Tabella AI Act a 4 livelli; schede-caso (una per ogni caso reale scelto: social scoring, screening CV, chatbot, filtro antispam, diagnosi medica automatica, deepfake politico); schede-ruolo per il comitato etico.

4.5 Dialogare con un LLM: prompt, bias e attendibilità

Voce	Descrizione
Obiettivo generale	Osservare empiricamente i limiti di un LLM (bias, allucinazioni, sensibilità al prompt) e costruire un vocabolario critico di base.
Classi di riferimento	III, IV, V.
Discipline coinvolte	Italiano, Filosofia, Informatica.
Durata standard	90 minuti.
Variante breve	50 minuti.
Prerequisiti	§ 2.2.4 del manuale.
Modalità	Laboratorio in aula informatica; lavoro a coppie; debriefing plenaria.

Fasi — versione standard (90 min)

Fase	Tempo	Attività
1. Brief	10 min	Che cos'è un LLM. Spiegazione dell'idea di predizione statistica di token.
2. Esperimento 1 — domanda fattuale	20 min	Le coppie pongono al modello una domanda fattuale specifica (es. su un personaggio locale poco noto). Verificano su fonti indipendenti.

Fase	Tempo	Attività
3. Esperimento 2 — bias	20 min	Le coppie formulano tre varianti dello stesso prompt cambiando solo nome/genere/provenienza. Confronto delle risposte.
4. Esperimento 3 — prompt	20 min	Stessa domanda con prompt "neutro" e con prompt che esplicita un ruolo o un contesto. Confronto.
5. Debriefing	20 min	Ogni coppia sintetizza in tre frasi: di cosa mi fido, di cosa non mi fido e cosa chiederei al sistema per fidarmi di più.

Variante breve (50 min)

Fase	Tempo	Attività
1. Brief	10 min	Spiegazione rapida di cosa sia un LLM.
2. Esperimento unico	25 min	Solo esperimento (test del bias): tre varianti di uno stesso prompt.
3. Debriefing	15 min	Ogni coppia riporta una sola osservazione sorprendente; chiusura con domanda generativa.

Materiali

Accesso a uno o più LLM in versione gratuita; schede di registrazione esiti; fonti di verifica predisposte dal docente; un piccolo glossario operativo (prompt, token, allucinazione, temperatura del modello).

Note etiche

Evitare che gli studenti inseriscano dati personali nei prompt; scegliere strumenti compatibili con il GDPR; non mostrare al modello contenuti protetti dal diritto d'autore o dati sensibili di terzi.

4.6 "Di chi ci fidiamo?" Statistiche, IA e disinformazione

Voce	Descrizione
Obiettivo generale	Riconoscere l'uso fuorviante di grafici e statistiche nella comunicazione mediatica; comprendere come i sistemi di IA possono amplificare o contrastare la disinformazione.
Classi di riferimento	III, IV, V.

Voce	Descrizione
Discipline coinvolte	Matematica, Italiano, Informatica, Educazione civica.
Durata standard	90 minuti.
Variante breve	50 minuti.
Prerequisiti	Statistica descrittiva di base.
Modalità	Gruppi + analisi di grafici + discussione plenaria.

Fasi — versione standard (90 min)

Fase	Tempo	Attività
1. Innesco	15 min	Vengono diffusi da diverse testate giornalistiche alcuni grafici a torta, diversi tra loro, per rappresentare una ripartizione. Sono tutti veritieri? I gruppi osservano e annotano discrepanze.
2. Smontaggio tecnico	25 min	Ogni gruppo riceve i grafici e li interpreta, confrontandosi con l'IA e individuando errori o scelte che orientano la percezione: settori leggermente diversi tra un grafico e l'altro, scelta dei colori, ecc.
3. Ricostruzione onesta	20 min	Ogni gruppo ridisegna a mano (o in foglio elettronico) un grafico che rappresenti i dati di partenza.
4. Confronto	15 min	Gallery walk: i gruppi confrontano le versioni autoprodotte e le versioni giornalistiche.
5. Discussione finale	15 min	Quale responsabilità delle redazioni nella creazione di grafici errati o fuorvianti? Quale ruolo l'IA svolge nell' interpretazione di grafici? Quale ruolo potrebbe svolgere nel verificarli e trovare gli errori?

Variante breve (50 min)

Fase	Tempo	Attività
1. Innesco	10 min	Presentazione di un solo grafico controverso tratto da stampa recente.

Fase	Tempo	Attività
2. Smontaggio e ricostruzione	30 min	Lavoro unico: smontaggio delle scelte grafiche e ridisegno (a mano o in foglio).
3. Plenaria focalizzata	10 min	Confronto rapido tra ricostruzioni, confrontandosi con l'IA; una sola domanda di chiusura: "Cosa cambia nella mia percezione?".

Materiali

Set di 4-6 grafici (il docente seleziona in funzione del proprio contesto; alcuni di questi devono essere errati); fogli quadrettati e/o foglio elettronico; schede di analisi con domande guida (asse, scala, unità, fonte, periodo); strumenti di IA (chatbot).

4.7 Dualità etiche: i "tre coltelli" della tecnologia

Voce	Descrizione
Obiettivo generale	Riconoscere che le tecnologie — IA inclusa — non sono né buone né cattive in sé, ma strumenti "a doppia faccia" il cui uso etico dipende dal contesto e dagli scopi.
Classi di riferimento	III, IV, V.
Discipline coinvolte	Filosofia, Storia, Scienze, Italiano.
Durata standard	90 minuti.
Variante breve	50 minuti.
Prerequisiti	§ 1.1 del manuale.
Modalità	Piccoli gruppi + dibattito strutturato.

Fasi — versione standard (90 min)

Fase	Tempo	Attività
1. Immagine-stimolo	10 min	Il coltello come simbolo: strumento del cuoco, del chirurgo, dell'omicida.
2. Brainstorming	15 min	Ogni gruppo elenca cinque tecnologie "a doppia faccia" (nucleare, internet, crispr, droni, riconoscimento facciale...).

Fase	Tempo	Attività
3. Scelta e analisi	25 min	Ogni gruppo sceglie una tecnologia e costruisce una mappa con gli usi buoni, gli usi cattivi, gli usi ambigui.
4. Dibattito	25 min	Due gruppi presentano; gli altri fanno da comitato etico e formulano 3 domande critiche.
5. Chiusura metacognitiva	15 min	Cosa ho imparato sul giudizio etico nei confronti di uno strumento?

Variante breve (50 min)

Fase	Tempo	Attività
1. Immagine-stimolo + tecnologia comune	10 min	Tutti i gruppi lavorano sulla stessa tecnologia (es. riconoscimento facciale) per velocizzare il confronto.
2. Mappa buono/cattivo/ambiguo	25 min	Lavoro di gruppo in parallelo.
3. Micro-dibattito	15 min	Uno o due gruppi presentano; chiusura con una domanda affidata a casa.

4.8 Governance dell'IA: chi decide?

Voce	Descrizione
Obiettivo generale	Comprendere l'architettura multi-stakeholder della governance dell'IA e le tensioni tra gli attori coinvolti.
Classi di riferimento	IV, V.
Discipline coinvolte	Diritto/Economia, Filosofia, Educazione civica, Storia.
Durata standard	90 minuti.
Variante breve	50 minuti.
Prerequisiti	§ 2.3 del manuale.
Modalità	Simulazione: ogni gruppo rappresenta uno stakeholder.

Fasi — versione standard (90 min)

Fase	Tempo	Attività
1. Scenario	10 min	Un governo europeo sta decidendo se introdurre il riconoscimento facciale nelle stazioni ferroviarie.
2. Distribuzione ruoli	10 min	Gruppi: governo, industria, cittadinanza, garante privacy, forze dell'ordine, ONG, giornalisti.
3. Preparazione	25 min	Ogni gruppo prepara la propria posizione (2 argomenti principali, 1 richiesta concreta).
4. Audizione	30 min	Il governo convoca le audizioni. Ogni stakeholder interviene (3 minuti) e risponde a una domanda.
5. Decisione e debriefing	15 min	Il governo comunica la decisione. Dibattito finale: chi ha vinto? chi è stato escluso?

Variante breve (50 min)

Fase	Tempo	Attività
1. Scenario + ruoli	10 min	Presentazione rapida; riduzione dei ruoli a 4 gruppi (governo, industria, cittadinanza, garante).
2. Preparazione	15 min	Ogni gruppo prepara un solo argomento.
3. Audizione compressa	15 min	Ogni gruppo ha 2 minuti di intervento; una sola domanda per tutti.
4. Decisione	10 min	Il "governo" decide e motiva; debriefing lampo.

Nota finale — condivisione dei materiali

I materiali di queste otto schede sono concepiti come open: ogni docente è invitato ad adattarli, tradurli, modificarli in funzione della propria classe e a restituire alla comunità AI&Ethics Literacy le versione prodotte, inviandole a questo indirizzo email: aiethicsliteracy@gmail.com.

Glossario

Il glossario raccoglie i termini tecnici (informatici) e filosofici che ricorrono nel manuale. Le definizioni sono pensate per essere direttamente utilizzabili in classe, senza rinviare a fonti esterne. Ogni voce è accompagnata da un rimando al paragrafo in cui il concetto è sviluppato.

Voce	Definizione
AI Act	Regolamento dell'Unione Europea (2024/1689) che stabilisce regole armonizzate sull'intelligenza artificiale, adottando un approccio basato sul rischio. Classifica i sistemi di IA in quattro livelli (inaccettabile, alto, limitato, minimo). Vedi § 2.3.2.
Algoritmo	Sequenza finita e ordinata di istruzioni che porta, a partire da un input, a un output determinato. In informatica classica è esplicito e ripetibile; nei sistemi di machine learning diventa il risultato di un processo e produce un modello probabilistico.
Bias cognitivo	Distorsione sistematica del giudizio umano, derivante dal modo in cui la mente elabora informazioni. Esempi: bias di conferma, euristica della disponibilità. Vedi § 1.3.
Bias algoritmico	Distorsione sistematica delle risposte di un sistema di IA. Non nasce dalla "cattiva volontà" del sistema, ma dai dati di addestramento, dall'etichettatura, dalla progettazione. Vedi § 2.2.5.
Deep Learning (DL)	Sottoinsieme del Machine Learning che utilizza reti neurali artificiali profonde (con molti strati). Ha permesso progressi straordinari ma è intrinsecamente opaco ("scatola nera"). Vedi § 2.2.4.
Ecosistema di IA	L'insieme di attori (policy-maker, finanziatori, sviluppatori, utenti), infrastrutture, dati e norme che rendono possibile lo sviluppo e l'uso dei sistemi di IA. Nel LEM, la responsabilità etica è distribuita su tutto l'ecosistema. Vedi § 1.2.
Ermeneutica	Disciplina filosofica che studia i processi di interpretazione. In chiave educativa, designa la capacità di leggere e comprendere un testo, un dato o un fenomeno cogliendone il significato nel contesto. Vedi § 1.3.
Feedback loop	Ciclo di retroazione tra un sistema e il suo ambiente. In ML, le interazioni degli utenti con le raccomandazioni di un sistema

Voce	Definizione
	generano nuovi dati che lo ri-addestrano: i bias iniziali possono essere rinforzati anziché corretti. Vedi § 2.2.6.
Ingiustizia epistemica	Concetto introdotto da Miranda Fricker (2007): torto specifico subito dalle persone nella loro capacità di conoscere e di essere riconosciute come fonti attendibili di sapere. Vedi § 1.1.
IA generativa	Sistemi di IA che non si limitano a classificare o prevedere, ma generano contenuti nuovi (testo, immagine, audio, codice). I Large Language Models e i Large Multimodal Models sono gli esempi più diffusi. Vedi § 2.2.4.
Large Ethics Model (LEM)	Framework concettuale multi-stakeholder, sviluppato da Pisano e colleghi, per orientare la riflessione etica sull'IA. Fonda un principialismo pratico ispirato a Beauchamp e Childress. Vedi § 1.2.
Large Language Model (LLM)	Modello statistico addestrato su enormi quantità di testo per predire la parola successiva in una sequenza. Non "comprende" in senso umano: è un pattern-matcher sofisticato. Vedi § 2.2.4.
Machine Learning (ML)	Approccio all'IA in cui l'interpretazione dei dati è "delegata" alla macchina: l'algoritmo impara pattern dai dati anziché seguire regole scritte esplicitamente dal programmatore. Vedi § 2.2.4.
PAIA	Acronimo delle quattro proprietà che rendono l'IA un oggetto etico peculiare: Pervasività, Autonomia, Invisibilità, Adattività. Non è una teoria ma una griglia di analisi. Vedi § 1.2.
Philosophical Enquiry (PhiE)	Metodologia dialogica erede del Philosophy for Children di Lipman e sviluppata da Peter Worley. Trasforma la classe in una comunità di ricerca dove il docente è facilitatore, non trasmettitore di risposte. Vedi § 3.1.
Phronesis	Saggezza pratica, virtù aristotelica che orienta il giudizio nelle situazioni concrete, laddove le regole generali non bastano. È il ponte tra principi etici astratti e decisione situata. Vedi § 2.3.1.
Prompt	Istruzione data dall'utente a un modello generativo. Piccole modifiche nel prompt possono portare a risposte molto diverse: è una fonte nascosta di variabilità e di bias. Vedi laboratorio 4.5.

Voce	Definizione
Schiavitù del clic	Espressione (Casilli, 2020) che designa il lavoro umano invisibile che sostiene l'IA: etichettatura dei dati, moderazione dei contenuti, verifica delle traduzioni. Spesso precario e mal retribuito. Vedi § 2.3.3.
Stakeholder	Parti interessate coinvolte in un sistema di IA: policy-maker, finanziatori, sviluppatori, utenti. Nel LEM, ciascuna è portatrice di responsabilità etiche specifiche. Vedi § 1.2.
Trolley problem	Esperimento mentale ideato da Philippa Foot e sviluppato da Judith Jarvis Thomson: un carrello in corsa può essere deviato per salvare cinque persone a costo di uccidere una sola. Strumento classico per introdurre le teorie etiche. Vedi § 2.1.1 e laboratorio 4.1.
Vulnerabilità epistemica	Condizione strutturale di esposizione a rappresentazioni distorte o semplificate della realtà. Non è una debolezza individuale ma una caratteristica dell'essere umano nella sua relazione con l'ecosistema informativo. Vedi § 1.1.

Bibliografia generale

La bibliografia raccoglie, in ordine alfabetico, i riferimenti principali citati nei quattro capitoli del manuale. Per le letture ulteriori si rinvia alle bibliografie in coda a ciascun capitolo.

Anscombe G.E.M., "Modern Moral Philosophy", *Philosophy* 33(124), 1958.

Aristotele, *Etica Nicomachea*, a cura di C. Natali, Laterza, Roma-Bari 1999.

Bauer K., Heigl R., Hinz O., Kosfeld M., "Feedback Loops in Machine Learning: A Study on the Interplay of Continuous Updating and Human Discrimination", *Journal of the Association for Information Systems* 25(4), 2024, pp. 804-866.

Beauchamp T.L., Childress J.F., *Principles of Biomedical Ethics*, 7th ed., Oxford University Press, New York 2013.

Bender, E. M., Gebru, T., McMillan-Major, A., and et al (2021). On the dangers of stochastic parrots: Canlanguage models be too big? . In *Proceedings of FAccT* 2021

Bostrom N., *Superintelligenza. Tendenze, pericoli, strategie*, Bollati Boringhieri, Torino 2018.

Casilli A.A., *Schiavi del clic. Perché lavoriamo tutti per il nuovo capitalismo*, Feltrinelli, Milano 2020.

Crawford K., *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*, Yale University Press, New Haven 2021.

Cuomo S., Ranieri M., Biagini G., *Scuola e Intelligenza Artificiale: pervotrsi di alfabetizzazione critica*, Carocci, Roma 2024.

De la Higuera C., Iyer N., *AI for Teachers: An Open Textbook*, UNESCO Chair RELIA, Nantes 2024.

Dumouchel P., Damiano L., *Vivere con i robot. Saggio sull'empatia artificiale*, Raffaello Cortina, Milano 2019.

Fabre C. et al., "Making moral decisions with artificial agents as advisors. A fNIRS study", Computers in Human Behavior: Artificial Humans 2, 2024.

Floridi L., *Etica dell'Intelligenza Artificiale*, Raffaello Cortina, Milano 2022.

Floridi L. et al., "AI4People - An Ethical Framework for a Good AI Society", Minds & Machines 28, 2018.

Foot P., *Virtues and Vices*, Oxford University Press, Oxford 2002.

Fricker M., *Epistemic Injustice: Power and the Ethics of Knowing*, Oxford University Press, Oxford 2007.

High-Level Expert Group on AI, Ethics Guidelines for Trustworthy AI, European Commission, Brussels 2019.

Holmes W., Porayska-Pomsta K. (eds.), *The Ethics of Artificial Intelligence in Education*, Routledge, London 2023.

Kant I., *Fondazione della metafisica dei costumi* (1785), trad. it. a cura di V. Mathieu, Bompiani, Milano 2003.

Lipman M., *Thinking in Education*, Cambridge University Press, Cambridge 2003.

MacIntyre A., *Dopo la virtù*, Armando Editore, Roma 2007.

Mori M., *Manuale di bioetica*, Le Lettere, Firenze 2010.

Noble S.U., *Algorithms of Oppression: How Search Engines Reinforce Racism*, NYU Press, New York 2018.

OECD, Recommendation of the Council on Artificial Intelligence, OECD Publishing, Paris 2024.

O'Neil C., *Weapons of Math Destruction*, Penguin, London 2016.

Pisano A., "Does the AI Need a Genealogy? The 'PAIA' Model", *Mimesis Journal - Mechané* 9(1), 2025.

Pisano, A., Crispini, I., Rossi, R.A. (2026). AI and Ethics Literacy in Secondary School: A Pilot Study. In: Dipace, A., Limongelli, C., Marras, M., Pagliara, S.M. (eds) Artificial Intelligence with and for Learning Sciences. Past, Present, and Future Horizons. WAILS 2025. Lecture Notes in Computer Science, vol 16438. Springer, Cham.

Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio del 13 giugno 2024 che stabilisce regole armonizzate sull'intelligenza artificiale (AI Act).

Ross W.D., *Il giusto e il bene*, Bompiani, Milano 2004.

Russell S., Norvig P., *Intelligenza artificiale: Un approccio moderno (vol. 1)*. Pearson, Torino, 2010.

Splitter L.J., Sharp A.M., *Teaching for Better Thinking: The Classroom Community of Inquiry*, ACER, Melbourne 1995.

Tiribelli S., *Identità personale e algoritmi*, Carocci, Roma 2024.

UNESCO, Recommendation on the Ethics of Artificial Intelligence, UNESCO, Paris 2021.

Worley P., *The If Machine: Philosophical Enquiry in the Classroom*, Bloomsbury, London 2010.

Zuboff S., *Il capitalismo della sorveglianza. Il futuro dell'umanità nell'era dei nuovi poteri*, LUISS University Press, Roma 2019.

Ringraziamenti

Questo manuale è il risultato di un lavoro collettivo durato oltre diciotto mesi, che ha attraversato tre paesi e coinvolto decine di persone.

Ringraziamo in primo luogo i 28 docenti della scuola secondaria di secondo grado che hanno validato il testo nella primavera 2026: con le loro osservazioni precise, le loro perplessità costruttive e i loro suggerimenti concreti hanno contribuito in modo determinante a rendere questo manuale quello che è oggi. Le loro firme in apertura del volume non sono una formalità: sono una responsabilità condivisa per le pagine di questo manuale.

Un grazie particolare alla SIpEIA - Società Italiana per l'Etica dell'Intelligenza Artificiale - per aver ospitato e promosso il progetto; al Dipartimento di Culture, Educazione e Società dell'Università della Calabria, che ha offerto la sua infrastruttura scientifica; alla Consulta di Bioetica, che ha contribuito al dialogo tra etica dell'IA e bioetica; alla UNESCO Chair RELIA di Nantes, con Colin de la Higuera, per l'apertura internazionale; al Center for Human-Centred AI and New Ethics of Technologies dell'Università di Ljubljana; al College of Staten Island - CUNY.

Ringraziamo tutti gli studenti del progetto pilota condotto nella scuola secondaria di secondo grado: senza la loro disponibilità a mettersi in gioco nel dialogo, questo manuale non sarebbe stato possibile. Le loro domande, i loro silenzi e le loro risate ci hanno insegnato, più di ogni bibliografia, cosa significhi "educare all'etica dell'IA".

Infine, un ringraziamento alle colleghe e ai colleghi delle associazioni, dipartimenti e istituzioni che hanno creduto nel progetto fin dall'inizio, e a quelle\i che lo faranno vivere nei prossimi anni.

Profili curatori e collaboratori

In questa sezione, in ordine alfabetico, le presentazioni dei curatori e dei collaboratori del manuale.

Maria Cristina Carrisi

Professoressa associata presso il Dipartimento di Matematica e Informatica dell'Università degli Studi di Cagliari (SSD MATH-01/B, Didattica e Storia della Matematica) e direttrice dei percorsi di formazione e abilitazione per docenti delle scuole secondarie (PF60). È co-curatrice del manuale. La sua ricerca si concentra sull'impatto delle tecnologie digitali e dell'intelligenza artificiale nei processi di insegnamento e apprendimento della matematica, sull'interazione tra didattica della matematica e dell'informatica. È co-curatrice di questo manuale.

Tiziana Catarci

Professoressa ordinaria di Ingegneria Informatica presso Sapienza Università di Roma. Direttrice dell'Istituto di Scienze e Tecnologie della Cognizione del CNR. Presidente di SIpEIA e co-curatrice di questo manuale. La sua ricerca si concentra su interazione uomo-computer, basi di dati e qualità dei dati, con sviluppi recenti nell'etica dell'intelligenza artificiale.

Loredana Rosa Cardullo

Università degli Studi di Catania. Professoressa ordinaria di Storia della filosofia antica e direttrice del Dipartimento di Scienze della Formazione. La sua ricerca si concentra sul pensiero aristotelico e il neoplatonismo, con particolare attenzione alla tradizione esegetica antica e ai rapporti tra filosofia, educazione e formazione.

Maria Concetta Carruba

Professoressa Associata di Didattica e Pedagogia Speciale presso l'Università Pegaso. Da sempre interessata alla tecnologia per l'apprendimento e la promozione della compossibilità, l'equità sociale e l'inclusione. Tra i temi di interesse anche la VR come strumento inclusivo, l'AI a supporto dei docenti e degli studenti con Bisogni Educativi Speciali, il Micro Learning e l'Informal micro learning.

Ines Crispini

Professoressa ordinaria di Filosofia morale presso l'Università della Calabria e Presidente del CUG di Atene. Tesoriera e socia fondatrice della Società italiana per l'etica dell'IA. I suoi interessi di ricerca includono: l'antropologia filosofica ed etica kantiana; filosofia ed questioni di genere; bioetica ed etica del digitale. È co-curatrice di questo manuale.

Colin de la Higuera

Professore Ordinario di Computer Science e Direttore della UNESCO Chair RELIA. Il suo lavoro combina machine learning, elaborazione del linguaggio naturale e pedagogia dell'IA per educatori. È il curatore della versione internazionale di questo manuale.

Piergiorgio Donatelli

Professore ordinario di Filosofia morale presso la Sapienza Università di Roma e presidente della Consulta di Bioetica. Dal 2012 al 2018 ha coordinato il Dottorato di ricerca in Filosofia e, dal 2011 al 2015, ha diretto il Master in Etica pratica e bioetica. È stato visiting professor presso la University

of Chicago, ha insegnato presso la Facoltà di Scienze politiche della LUISS Guido Carli e presso l'Université Paris 1 Panthéon-Sorbonne, e ha tenuto lezioni e conferenze in numerose università italiane e straniere. È inoltre coautore e curatore di diversi volumi dedicati all'etica e alla bioetica, e co-curatore di questo manuale.

Stephen Ferst

Executive Director del Center for Global Engagement presso il College of Staten Island (CUNY), con oltre 35 anni di esperienza nell'istruzione internazionale. È specializzato in programmi di studio all'estero, servizi per studenti/docenti internazionali e internazionalizzazione del campus, avendo in precedenza lavorato come Direttore dello Study Abroad alla Rutgers University.

Dario Lombardi

Dottore di ricerca in Scienze dello Sviluppo, della Formazione e dell'Apprendimento presso l'Università di Foggia, dove svolge attualmente attività di ricerca post-doc in Pedagogia Speciale. La sua attività scientifica si concentra sull'applicazione dell'Intelligenza Artificiale, dei Serious Game e della realtà virtuale ai contesti formativi. È docente di sostegno nella scuola secondaria di secondo grado e tutor del tirocinio indiretto presso il TFA dell'Università di Foggia."

Massimiliano Mancini

Dottorando in Matematica, Informatica e Statistica, specializzato nell'equità nei sistemi di machine learning e di classificazione. Contribuisce alle basi tecniche di questo manuale con particolare attenzione ai bias, alla valutazione e ai limiti degli attuali modelli di IA.

Emiliana Murgia

Dottoressa di ricerca in Learning Sciences in Digital Technologies (Università degli Studi di Genova). Insegnante, formatrice e ricercatrice sull'integrazione delle tecnologie digitali nell'istruzione primaria e secondaria. Il suo lavoro si concentra sulla progettazione didattica di attività che combinano alfabetizzazione tecnica e pensiero critico.

Giorgia Nieddu

Dottoranda Università degli Studi di Macerata e Università degli Studi di Cagliari. Il suo interesse di ricerca verte sull'impatto dell'intelligenza artificiale sulla trasposizione didattica dei contenuti di matematica e sui processi di apprendimento degli studenti.

Aldo Pisano

Dottore di ricerca in Filosofia morale, assistente alla didattica in Filosofia morale presso l'UNICAL, membro del Comitato Esecutivo di SIpEIA e del Direttivo della Consulta di Bioetica. La sua ricerca si concentra sull'etica dell'IA, sulla filosofia dell'educazione e sulla progettazione di interventi educativi nell'ambito dell'alfabetizzazione IA & Etica. È co-curatore e coordinatore principale di questo manuale.

Rossana Adele Rossi

Professoressa associata di Pedagogia generale e sociale presso l'Università della Calabria, dove ricopre anche la carica di coordinatice dei corsi di laurea in Scienze dell'Educazione e Scienza Pedagogiche. Nell'arco della sua attività scientifica, ha coltivato alcune direzioni di ricerca tendenti a chiarire il nesso teoria-pratica sia in alcuni momenti della storia del pensiero pedagogico italiano sia in particolari luoghi centrali della più recente riflessione pedagogica.

Nunziatina Sanfilippo

Dottoressa di ricerca in Storia della filosofia, assistente alla didattica in Filosofia morale e Storia della filosofia presso l'Università di Catania, docente di Filosofia e Scienze umane nella scuola secondaria. Studiosa di storia e teoria dell'educazione, con lavori recenti dedicati all'adattamento didattico delle pratiche filosofiche al contesto scolastico.

Vojko Strahovnik

Professore di filosofia e direttore del *Centre for Human-Centred Artificial Intelligence and the Ethics of New Technologies*, Università di Lubiana, Facoltà di Arti e Teologia. I suoi interessi di ricerca comprendono etica, epistemologia e teoria morale, con particolare attenzione all'etica dell'intelligenza artificiale, alle virtù epistemiche e alle implicazioni normative delle tecnologie digitali. È coinvolto in progetti internazionali dedicati allo studio interdisciplinare delle sfide etiche e sociali dell'IA e contribuisce al dialogo tra filosofia, educazione e sviluppo tecnologico.

Luca Tenneriello

Dottore di ricerca in filosofia, svolge attività di ricerca presso la Sapienza Università di Roma. È stato Visiting Researcher presso la University of Kent, nel Regno Unito. I suoi interessi spaziano tra filosofia morale e politica, etica pubblica ed etica dell'intelligenza artificiale. Tra il 2024 e il 2025 ha lavorato al Progetto per Avvio alla Ricerca intitolato *IA, autonomia, responsabilità: radici storico-filosofiche e linee normative* (nell'ambito dei Bandi di Ateneo per la Ricerca 2024) di cui è stato proponente. È autore, tra l'altro, di *Thomas Hobbes. La religione e la coscienza* (ETS, Pisa 2023). È membro del Comitato Esecutivo di SIpEIA – Società Italiana per l'Etica dell'Intelligenza Artificiale. Contribuisce al lavoro del manuale sull'etica normativa.

Antonella Valenti

Professoressa Ordinaria di Didattica e Pedagogia Speciale presso l'Università della Calabria. È coordinatrice del Corso di laurea magistrale quinquennale a ciclo unico in Scienze della formazione primaria, abilitante all'insegnamento nella scuola primaria e dell'infanzia e Direttrice dei corsi di specializzazione per gli insegnanti di sostegno. La sua attività di ricerca, basata sull'evidenza scientifica, è rivolta all'individuazione di proposte e modelli didattici che possano facilitare gli apprendimenti e l'inclusione di studenti con Bisogni Educativi Speciali (BES).

Edizione open access · Prima edizione, UNICApress 2026
AI & Ethics Literacy for Teachers

e-ISBN 978-88-3312-248-9
DOI https://doi.org/10.13125/unicapress.978-88-3312-248-9