



Gruppo di lavoro sulle tecniche
di insegnamento e di apprendimento

<https://riviste.unige.it/index.php/glia/index>
2975-0075

N° 5 - Anno 2026
pp. 289-309

L'intelligenza artificiale nella valutazione dell'apprendimento in università Fondamenti etici, quadro normativo e prospettive di governance

**Flavio MANGANELLO¹, Alberto NICO², Martina RAGUSA³,
Giannangelo BOCCUZZI^{4,5}**

1 Istituto per le Tecnologie Didattiche, CNR, Genova (GE), flavio.manganello@cnr.it

2 Università degli Studi di Bari, Bari (BA), alberto.nico@uniba.it

3 Istituto per le Tecnologie Didattiche, CNR, Genova (GE), martinaragusa@cnr.it

4 Istituto per le Tecnologie Didattiche, CNR, Genova (GE), giannangeloboccuzzi@cnr.it

5 Università di Bologna, Bologna (BO), giannangelo.boccuzz2@unibo.it

Open Access article distributed under CC BY-NC-ND 4.0
Copyright © Genova University Press



Abstract

Il contributo esamina le implicazioni etiche, normative e operative dell'intelligenza artificiale applicata alla valutazione dell'apprendimento in ambito universitario, attraverso un approccio a studi multipli che integra analisi bibliometrica, revisioni della letteratura e sviluppo teorico di framework. L'analisi bibliometrica condotta su 334 documenti Scopus (2019-2025) documenta una crescita sostenuta dell'interesse scientifico sull'intersezione tra IA spiegabile e valutazione universitaria ($R^2 = 0,73$), evidenziando la persistente frammentazione dei framework etici di riferimento. A partire da una revisione narrativa di 121 studi (2015-2025) vengono identificate dieci dimensioni etiche ricorrenti, raggruppabili attorno a tre aree principali: principi fondamentali di sistema, dimensioni centrate sullo studente e governance. Il quadro normativo europeo, in particolare l'EU AI Act e il GDPR, viene analizzato nelle sue implicazioni per i sistemi ibridi umano-algoritmo, mostrando l'assenza di linee guida operative adeguate. Sul piano teorico-operativo, il contributo ricostruisce il percorso che ha condotto da una prima formalizzazione operativa, il Framework XAI-ED, a un modello di governance teoricamente più solido, il XAI-ED Consequential Assessment Framework (XAI-ED CAF), che integra i sette requisiti ALTAI con validità consequenziale di Messick, del modello di Kirkpatrick e del framework CIPP di Stufflebeam. Vengono discusse le ricadute pratiche per chi opera nella formazione e nella ricerca sulla didattica universitaria.

Keywords

intelligenza artificiale; valutazione dell'apprendimento; etica dell'IA; EU AI Act; XAI-ED CAF

1. Introduzione

I sistemi basati sull'intelligenza artificiale sono ormai presenti, in forme e gradi diversi, all'interno dei processi valutativi universitari. Non si tratta soltanto di strumenti di supporto marginali: modelli computazionali vengono oggi impiegati per valutare elaborati scritti, predire l'esito di prove d'esame, sorvegliare sessioni di testing a distanza e generare feedback individualizzato. Questa progressiva espansione ha spostato il baricentro della discussione scientifica e normativa: non si discute più soltanto se sia opportuno introdurre l'IA nella valutazione, ma come governare sistemi già operativi che producono effetti diretti sulle carriere accademiche degli studenti.

Il Regolamento europeo sull'intelligenza artificiale, adottato nel 2024, ha dato una risposta normativa parziale a questo problema classificando la valutazione dell'apprendimento basata su IA come applicazione ad alto rischio (Allegato III, punto 3b), con conseguenti obblighi di trasparenza, supervisione umana e gestione del rischio (European Commission, 2024). Prima ancora, l'articolo 22 del GDPR aveva introdotto limiti alle decisioni interamente automatizzate con effetti significativi sulle persone, riconoscendo implicitamente un diritto alla spiegazione dei procedimenti algoritmici. Eppure la letteratura ha mostrato con chiarezza che l'intersezione operativa tra questi due strumenti normativi rimane problematica, soprattutto nei sistemi ibridi nei quali la valutazione è il risultato di una combinazione tra elaborazione algoritmica e giudizio umano. Nessuno dei due impianti regolativi fornisce, da solo, orientamenti pratici sufficienti per questa configurazione (Boccuzzi, Nico & Manganello, 2025b).

Da un lato, l'analisi bibliometrica della produzione scientifica sull'IA spiegabile nella valutazione universitaria mostra una crescita sostenuta delle pubblicazioni tra il 2019 e il 2025, con quattro macro-cluster tematici (etica e tecnologia educativa; deep learning e analytics predittivo; sistemi di feedback e learning analytics; valutazione e monitoraggio educativo) che procedono in parallelo senza convergere verso framework etici condivisi; l'analisi di co-citazione mostra inoltre che il lavoro fondazionale del campo è una survey informatica sull'XAI (eXplainable Artificial Intelligence), confermando che il campo si è sviluppato adattando strumenti tecnici al dominio educativo piuttosto che costruendo dalla teoria valutativa (Manganello, Nico & Boccuzzi,

2025a). Dall'altro, le revisioni narrative della letteratura registrano che concetti fondamentali come "assessment", "spiegabilità" e "trasparenza" sono utilizzati con accezioni divergenti tra informatica, scienze dell'educazione e diritto, rendendo difficile qualsiasi confronto sistematico tra studi e, in assenza di definizioni condivise, lo sviluppo di standard istituzionali coerenti (Boccuzzi, Nico & Manganello, 2025a).

La convergenza di queste due evidenze segnala una lacuna che non è solo di ordine bibliografico. Le università che intendano adottare sistemi di IA nella valutazione non dispongono, allo stato attuale, di strumenti interpretativi che integrino i requisiti normativi europei con una teoria valutativa pedagogicamente fondata: i quadri normativi impongono obblighi senza fornire i criteri per verificarne il rispetto sul piano degli esiti formativi, e i framework proposti dalla letteratura sull'etica dell'IA derivano più spesso dall'informatica che dalla tradizione della valutazione educativa.

Il presente contributo affronta questo doppio nodo, normativo e concettuale, ponendosi tre obiettivi tra loro connessi. Il primo è ricostruire il paesaggio della ricerca sull'etica dell'IA nella valutazione universitaria attraverso evidenze bibliometriche e narrative, identificando le dimensioni etiche ricorrenti e le tensioni strutturali che le attraversano. Il secondo è analizzare il quadro regolativo europeo vigente nelle sue implicazioni operative per le istituzioni universitarie italiane, con particolare attenzione ai sistemi ibridi. Il terzo è presentare il percorso teorico che ha condotto dal Framework XAI-ED, prima formalizzazione degli autori, al XAI-ED Consequential Assessment Framework (XAI-ED CAF), sua evoluzione fondata sulla teoria valutativa, discutendone le implicazioni per la didattica universitaria.

2. Il quadro normativo europeo e le sue implicazioni per la valutazione universitaria

L'adozione di sistemi di intelligenza artificiale nella valutazione accademica non avviene in un vuoto regolativo. Due strumenti normativi europei definiscono oggi il perimetro entro cui le istituzioni universitarie devono operare: il Regolamento generale sulla protezione dei dati (GDPR, 2016) e il Regolamento sull'intelligenza artificiale (EU AI

Act, 2024). Sebbene i due testi abbiano origini, finalità e strutture diverse, la loro sovrapposizione nell'ambito della valutazione automatizzata produce un insieme di obblighi che le università italiane sono chiamate a rispettare, spesso senza disporre di strumenti interpretativi adeguati.

2.1 L'EU AI Act e la classificazione ad alto rischio

Il Regolamento (UE) 2024/1689, meglio noto come AI Act, introduce un sistema di classificazione dei sistemi di IA basato sul livello di rischio che tali sistemi possono generare per diritti fondamentali, salute e sicurezza. Al vertice di questa classificazione si trovano le applicazioni "ad alto rischio", elencate nell'Allegato III del Regolamento. L'Allegato III, punto 3b, ricomprende espressamente i sistemi di IA destinati a essere utilizzati per la valutazione degli studenti negli istituti di istruzione e formazione professionale, inclusa la valutazione degli apprendimenti. Questa classificazione comporta obblighi specifici per fornitori e deployer, che spaziano dalla gestione del rischio e dalla documentazione tecnica alla trasparenza verso gli utenti e alla garanzia di supervisione umana.

Quest'ultimo requisito merita una riflessione particolare. L'AI Act non vieta l'uso di sistemi automatizzati nella valutazione, ma impone che essi non sostituiscano integralmente il giudizio umano in decisioni che incidono significativamente sulla vita delle persone. La supervisione umana non è intesa come una semplice procedura formale di ratifica, ma come un intervento sostanziale capace di individuare e correggere eventuali errori o distorsioni del sistema. Ciò implica che il personale docente e amministrativo chiamato a utilizzare tali strumenti debba possedere una sufficiente comprensione del loro funzionamento, dei loro limiti e delle loro potenziali fonti di errore: una condizione che chiama direttamente in causa le politiche di formazione continua degli atenei (Manganello, Nico & Boccuzzi, 2025b).

2.2 Il GDPR e il diritto alla spiegazione

Il GDPR, con il suo articolo 22, aveva già anticipato parte di queste problematiche ponendo limitazioni alle decisioni basate esclusivamente sul trattamento automatizzato di dati personali che producano effetti giuridici o che incidano in modo analogo significativamente sugli

interessati. In ambito universitario, una valutazione prodotta da un sistema algoritmico che determini il superamento o il mancato superamento di un esame rientra senza dubbio in questa categoria. L'articolo 22 individua le condizioni disgiunte che possono legittimare tale trattamento (consenso esplicito, necessità contrattuale o norma di legge) garantendo in ogni caso all'interessato almeno il diritto all'intervento umano e alla contestazione della decisione.

Al di là dei casi in cui il trattamento automatizzato sia esplicitamente autorizzato, il GDPR riconosce implicitamente il diritto alla spiegazione delle decisioni algoritmiche attraverso le disposizioni degli articoli 13 e 14 sull'informativa e dell'articolo 15 sul diritto di accesso. In contesti educativi, questo si traduce nell'obbligo di informare lo studente circa l'esistenza di processi decisionali automatizzati, la logica sottostante e le conseguenze previste. La ricerca ha peraltro evidenziato che la nozione di "spiegazione" rimane terminologicamente ambigua: le garanzie di trasparenza previste dal diritto amministrativo per le decisioni umane motivate non si sovrappongono automaticamente ai requisiti di spiegabilità tecnica propri dei sistemi di IA, e il raccordo tra queste due tradizioni rimane un problema aperto sia in dottrina sia nella prassi istituzionale (Boccuzzi, Nico & Manganello, 2025b).

2.3 Il problema dei sistemi ibridi

La difficoltà più rilevante non riguarda, tuttavia, i sistemi pienamente automatizzati, la cui disciplina è relativamente più chiara, ma quelli ibridi, nei quali la valutazione finale è il prodotto di una combinazione tra elaborazione algoritmica e giudizio umano. Queste configurazioni sono oggi le più diffuse nella pratica universitaria: un sistema di IA genera una valutazione preliminare o un insieme di indicatori che il docente poi integra, modifica o conferma. In tale schema, non è facile stabilire dove termini la responsabilità del sistema e dove inizi quella del valutatore umano, né quale peso abbia effettivamente avuto ciascuna componente nella decisione finale.

L'AI Act non affronta in modo diretto questa questione, concentrandosi sulle responsabilità dei fornitori e dei deployer piuttosto che sulle modalità di integrazione del giudizio umano nei processi decisionali. Il GDPR, dal canto suo, si applica alle decisioni "basate esclusivamente" sul trattamento automatizzato, lasciando un margine di incertezza interpretativa proprio per i casi ibridi, dove l'intervento

umano è presente ma potrebbe essere meramente formale. Un'indagine empirica sulla governance dei sistemi di IA nelle università italiane ha rilevato che solo il 13% degli atenei aveva adottato politiche formali in materia entro il 2025, e che le istituzioni analizzate documentano con maggiore sistematicità gli impegni valoriali di carattere aspirazionale piuttosto che le salvaguardie tecniche operative (Manganello, Nico, Ragusa & Boccuzzi, 2025). Il problema è dunque anche organizzativo: rispettare gli obblighi dell'AI Act richiede investimenti in competenze e strutture di governance che la maggior parte degli atenei italiani non ha ancora avviato.

3. Il paesaggio della ricerca: analisi bibliometrica e narrativa

La comprensione delle implicazioni etiche dell'IA nella valutazione universitaria richiede, come premessa, una ricostruzione sistematica dello stato dell'arte. Il programma di ricerca qui presentato ha affrontato questa esigenza attraverso due approcci complementari: un'analisi bibliometrica sull'IA spiegabile, cioè i metodi che rendono interpretabili i sistemi di IA (Adadi & Berrada, 2018), nella valutazione universitaria, e una serie di revisioni della letteratura, narrativa e rapida, sulle dimensioni etiche della valutazione automatizzata.

3.1 L'analisi bibliometrica: crescita, struttura e lacune del campo

La prima componente ha interrogato Scopus con la stringa ALL("right to explanation" OR "right to explainability" OR "explainable AI" OR "Explainable Artificial Intelligence") AND TITLE-ABS-KEY("higher education") AND (assessment OR evaluation), da cui un corpus iniziale di 334 documenti analizzati con performance analysis e science mapping in VOSviewer.

Il dato più evidente riguarda la dinamica temporale delle pubblicazioni. Il campo emerge con contributi isolati nel 2019 (2 documenti) e nel 2020 (3 documenti), per poi espandersi in modo sostenuto: 9 documenti nel 2021, 38 nel 2022, 60 nel 2023, 140 nel

2024. I dati parziali del 2025, che contano già 82 documenti a metà anno, proiettano una stima di circa 164 pubblicazioni annue, confermando una traiettoria di crescita non casuale. La regressione lineare applicata alla serie storica restituisce un coefficiente $R^2 = 0,7343$ ($p < 0,05$), indicando che circa il 73% della varianza nel numero di pubblicazioni è spiegato dalla progressione temporale. Questa crescita coincide, non sorprendentemente, con l'inasprimento del quadro regolativo europeo: l'accelerazione più marcata si registra tra il 2022 e il 2024, negli anni immediatamente successivi all'intensificazione del dibattito sull'AI Act.

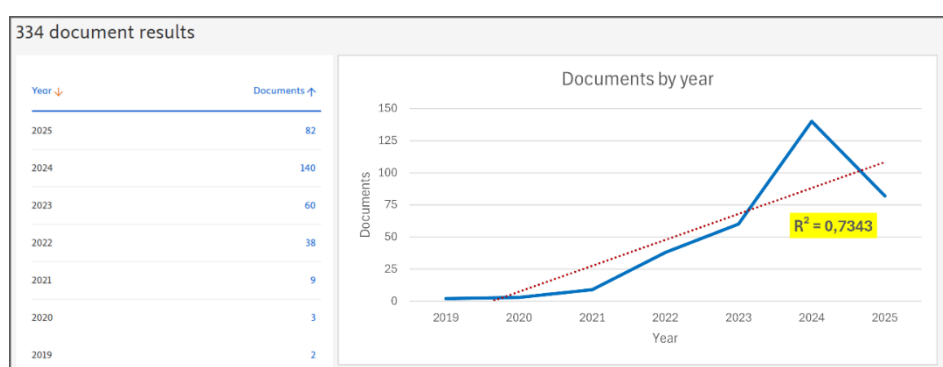


Figura 1. Distribuzione annuale delle pubblicazioni sul tema IA spiegabile e valutazione universitaria (Scopus, 2019-2025), con retta di regressione lineare ($R^2 = 0,7343$)

La mappa di co-occorrenza (Figura 2) evidenzia un ecosistema articolato in tre aree: un nucleo tecnologico denso attorno a machine learning e data mining, un cluster educativo che raccoglie higher education e learning analytics, e applicazioni più periferiche come i large language models. Il tema dell'etica e della trasparenza appare in posizione decentrata, con legami ancora deboli verso entrambi i nuclei principali, indicando un campo frammentato sulle questioni normative.

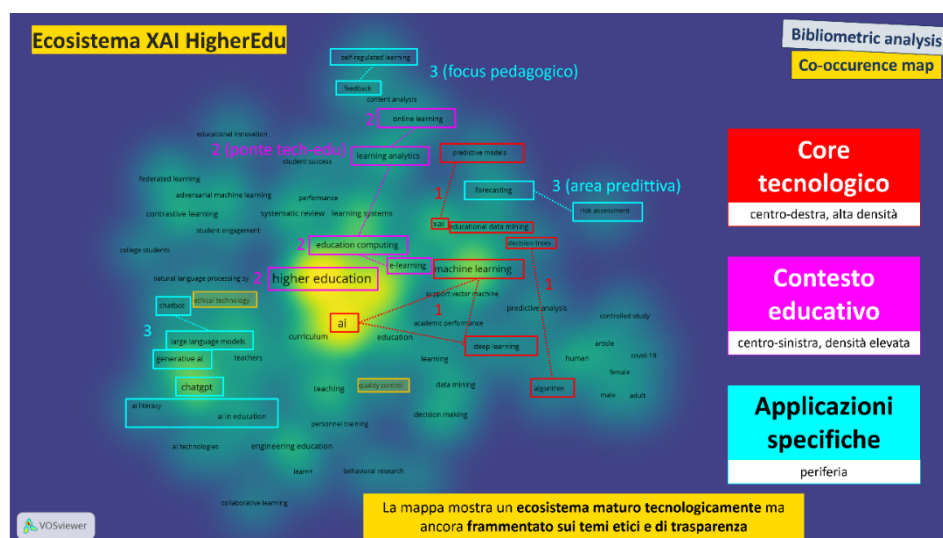


Figura 2. Mappa di co-occorrenza delle parole chiave della letteratura su IA spiegabile e valutazione universitaria (VOSviewer, corpus Scopus 2019-2025)

Una seconda indagine bibliometrica, condotta con criteri più restrittivi seguendo le raccomandazioni metodologiche GLOBAL (Manganello, Nico & Boccuzzi, 2025a) tramite una stringa booleana a tre componenti su IA spiegabile, istruzione superiore e valutazione, ha esaminato 72 pubblicazioni peer-reviewed estratte da un corpus iniziale di 141 documenti. Anche su questo corpus la crescita risulta statisticamente significativa ($R^2 = 0,7666$, $p = 0,0098$), con l'81,9% delle pubblicazioni concentrate nel triennio 2023-2025. Dal punto di vista geografico, l'Europa emerge come il contribuente regionale principale con il 43,8% delle pubblicazioni geografiche rilevate (49 documenti da 19 Paesi, su un conteggio per Paese che eccede il numero di pubblicazioni), dato che riflette la pressione normativa del GDPR prima e del dibattito sull'AI Act poi.

L'analisi della struttura tematica tramite co-occorrenza delle parole chiave, con standardizzazione di 220 termini in 147 concetti core, ha identificato quattro macro-cluster: "AI Ethics and Educational Technology", che aggrega ricerca sui framework etici, sui modelli generativi e sui principi human-centred; "Deep Learning and Predictive Analytics", orientato alle metodologie tecniche di predizione e interpretabilità; "Learning Analytics and Feedback Systems", focalizzato sull'implementazione pratica nei contesti educativi; "Assessment and Educational Evaluation", dedicato ai requisiti di trasparenza per le decisioni ad alto rischio. L'analisi di co-citazione

conferma che il lavoro fondazionale del campo è la survey di Adadi e Berrada (2018) sull'XAI, con 11 citazioni e un total link strength di 24 (la forza complessiva dei suoi legami di co-citazione con gli altri documenti, indicatore VOSviewer): il campo si è sviluppato a partire dall'informatica, adattando strumenti tecnici al dominio educativo piuttosto che dalla teoria valutativa. È questa asimmetria che il XAI-ED Consequential Assessment Framework (XAI-ED CAF), presentato al par. 4, intende correggere.

3.2 Le revisioni della letteratura: dimensioni etiche e tensioni strutturali

In parallelo all'analisi bibliometrica, il programma di ricerca ha condotto una revisione narrativa della letteratura per ricostruire il panorama concettuale delle dimensioni etiche associate all'uso dell'IA nella valutazione educativa. Il corpus esaminato comprende 121 studi pubblicati tra il 2015 e il 2025, con un approccio interdisciplinare che integra prospettive dell'informatica, della pedagogia e del diritto (Boccuzzi, Nico & Manganello, 2025a).

Una prima evidenza riguarda la terminologia. Il termine “assessment” è impiegato con accezioni significativamente diverse nei tre ambiti disciplinari: nelle scienze dell'educazione indica un processo formativo orientato allo sviluppo delle competenze; nel diritto si riferisce a una decisione con effetti giuridicamente rilevanti; nell'informatica designa spesso una funzione di classificazione o scoring. Analogamente, “spiegabilità” e “trasparenza” assumono contenuti tecnici distinti a seconda del contesto, rendendo problematica qualsiasi comparazione cross-disciplinare (Boccuzzi, Nico & Manganello, 2025a). Questa eterogeneità terminologica non è meramente accademica: produce conseguenze dirette sulla possibilità di costruire framework regolativi coerenti e di sviluppare standard istituzionali comparabili tra atenei. Ha inoltre implicazioni metodologiche immediate per chiunque si accinga a costruire strumenti di governance in questo dominio. Un framework che ambisca a disciplinare l'uso dell'IA nella valutazione universitaria deve preliminarmente esplicitare in quale accezione disciplinare utilizza ciascun termine fondamentale, altrimenti il rischio concreto è che documenti di policy formalmente condivisi vengano interpretati da comunità professionali diverse in modi tra loro inconciliabili,

producendo un'apparenza di consenso che non regge alla prima applicazione operativa.

Accanto alla revisione narrativa, il programma di ricerca ha condotto anche una revisione rapida della letteratura, su un corpus più selettivo di 44 studi pubblicati nello stesso arco temporale (Boccuzzi, Manganello & Nico, 2025). Da questa seconda revisione emergono dieci dimensioni etiche ricorrenti nei sistemi di valutazione automatizzata in educazione. Queste dimensioni non si presentano in modo isolato, ma possono essere raggruppate in tre aree funzionalmente distinte, come sintetizzato nella Tabella 1.

La stessa revisione rapida ha inoltre rilevato l'esistenza di tre tensioni critiche strutturali che attraversano trasversalmente le dimensioni etiche identificate. La prima contrappone trasparenza e complessità: i modelli predittivi più accurati sono spesso quelli meno interpretabili, e le tecniche di spiegazione post-hoc comunemente impiegate possono trasmettere una comprensione fuorviante dei processi decisionali effettivi. La seconda oppone efficienza e autonomia dello studente: i sistemi di feedback automatizzato, pur aumentando la frequenza e la tempestività delle risposte, tendono a favorire un apprendimento passivo e a ridurre la capacità di autoregolazione. La terza mette in tensione personalizzazione e privacy: la granularità dei dati necessaria per personalizzare l'esperienza valutativa aumenta contestualmente i rischi di profilazione e sorveglianza.

Queste tensioni non sono risolvibili semplicemente scegliendo un polo a scapito dell'altro. Richiedono soluzioni di design che le affrontino in modo esplicito, riconoscendo i trade-off implicati e distribuendo responsabilità chiare tra i diversi attori del sistema. È su questa esigenza che si fonda il percorso teorico descritto nella sezione successiva, che muove da una prima formalizzazione operativa, il Framework XAI-ED, verso un modello di governance più solidamente fondato sul piano pedagogico, il XAI-ED Consequential Assessment Framework (XAI-ED CAF).

Tabella 1. Le dieci dimensioni etiche dei sistemi di valutazione basati su IA: una classificazione tematica (elaborazione da Boccuzzi, Manganello & Nico, 2025)

Area	Dimensione	Contenuto
Principi	Trasparenza e	Chiarezza nei processi decisionali; spiegazioni comprensibili per studenti e

Area	Dimensione	Contenuto
fondamentali	spiegabilità	docenti
	Equità e non-discriminazione	Trattamento equo tra gruppi; prevenzione dei bias algoritmici
	Integrità	Affidabilità dei risultati; sicurezza contro frodi e manipolazioni
Dimensioni centrate sullo studente	Agency e autonomia	Mantenimento del controllo studentesco; prevenzione della dipendenza dall'IA
	Motivazione e coinvolgimento	Preservazione della motivazione intrinseca; feedback orientato all'apprendimento
	Accessibilità e inclusione	Adattamento a diverse esigenze; supporto per studenti con disabilità
Governance e supervisione	Privacy e protezione dei dati	Salvaguardia delle informazioni sensibili; conformità GDPR
	Responsabilità e accountability	Ruoli chiari tra stakeholder; meccanismi di controllo e audit
	Conformità legale	Allineamento EU AI Act; rispetto normative nazionali
	Equità sociale	Riduzione disparità educative; accesso democratico all'educazione

4. Dalla revisione alla governance: il Framework XAI-ED e il suo sviluppo nel modello XAI-ED CAF

Il percorso teorico che ha condotto alla proposta di un framework operativo per la governance dell'IA nella valutazione universitaria non si è sviluppato in modo lineare. È passato attraverso una prima formalizzazione concettuale, il Framework XAI-ED, che ha svolto una

funzione euristica importante ma che presentava limiti strutturali riconosciuti, e attraverso una successiva rielaborazione teoricamente più solida, il XAI-ED Consequential Assessment Framework (XAI-ED CAF), che ne supera le insufficienze ancorando il modello a tradizioni consolidate della teoria della valutazione educativa.

4.1 Il Framework XAI-ED: una prima formalizzazione e i suoi limiti

Il Framework XAI-ED è emerso direttamente dalle revisioni della letteratura come tentativo di organizzare in un modello operativo le dieci dimensioni etiche identificate. Il framework articola la governance dei sistemi di valutazione basati su IA attorno a sei assi: gli attori coinvolti, i benefici attesi, gli approcci esplicativi disponibili, i modelli computazionali raccomandati, i principi di design e le criticità da presidiare. È stato applicato esemplarmente al caso della valutazione automatizzata di tesi di laurea magistrale, mostrando come un sistema possa fornire feedback strutturato su metodologia e argomentazione preservando l'autonomia intellettuale e il rigore accademico.

Tuttavia, una critica legittima può essere mossa al suo impianto: le dieci dimensioni etiche da cui esso è derivato sono state identificate attraverso un processo induttivo di revisione narrativa, con un passaggio alla formalizzazione prescrittiva non sufficientemente ancorato a fondamenti teorici indipendenti. Il framework descrive con coerenza interna le componenti rilevanti della governance, ma non fornisce i criteri per stabilire con quale evidenza e rispetto a quali standard educativi esse debbano essere valutate. Questa lacuna è stata riconosciuta come punto di partenza per uno sviluppo ulteriore (Manganello, Nico & Boccuzzi, 2025b).

4.2 Il XAI-ED CAF: fondamenti teorici e architettura del modello

Il XAI-ED Consequential Assessment Framework risponde a queste insufficienze attraverso un'operazione teorica più rigorosa. Anziché derivare il framework esclusivamente dalla revisione della letteratura sull'etica dell'IA, il XAI-ED CAF parte dalla riformulazione dei sette requisiti dell'Assessment List for Trustworthy AI (ALTAI) della Commissione Europea attraverso tre tradizioni consolidate della teoria della valutazione educativa: la validità consequenziale di Messick

(1995), secondo cui la validità di una valutazione dipende anche dalle conseguenze sociali del suo uso; il modello a quattro livelli di Kirkpatrick e Kirkpatrick (1998), che distingue reazione, apprendimento, comportamento e risultati; e il framework Context-Input-Process-Product (CIPP) di Stufflebeam (2003), che struttura la valutazione istituzionale in quelle quattro fasi. Questa integrazione sposta il fulcro della governance dalla conformità tecnico-normativa alla validità pedagogica degli esiti, in coerenza con i requisiti dell’AI Act per i sistemi ad alto rischio (Manganello, Nico & Boccuzzi, 2025b).

La scelta non è arbitraria: Messick orienta la valutazione verso le conseguenze sociali ed etiche degli esiti, Kirkpatrick consente di misurarne gli effetti su piani temporali distinti, Stufflebeam introduce la prospettiva processuale per un’autovalutazione istituzionale sistematica.

Il risultato è una rilettura dei sette requisiti ALTAI che ne sposta il contenuto dall’astratto al pedagogicamente fondato. La dimensione “human agency and oversight” viene reinterpretata non come presenza formale di un operatore umano, ma come preservazione dell’autonomia di studenti ed educatori nei processi valutativi. La “transparency” non riguarda più soltanto la disponibilità di spiegazioni tecniche, ma la loro comprensibilità e utilità per gli specifici stakeholder educativi. La mappatura sistematica tra i sette requisiti ALTAI, le fondamenta pedagogiche e il fuoco dell’assessment è presentata nella Tabella 2.

Questa architettura consente al XAI-ED CAF di svolgere una funzione che il framework XAI-ED non poteva assolvere: fornire ai responsabili istituzionali criteri teorici espliciti per stabilire quale tipo di evidenza dimostri che un sistema di valutazione produce esiti validi, equi e pedagogicamente fondati. Il passaggio è, in questo senso, dalla catalogazione delle dimensioni rilevanti alla giustificazione teorica delle scelte di governance.

Tabella 2. Mappatura tra requisiti ALTAI, fondamenta pedagogiche e fuoco dell’assessment nel XAI-ED CAF (da Manganello, Nico & Boccuzzi, 2025b)

Dimensione ALTAI	Fondamento pedagogico	Fuoco dell’assessment
Human agency e oversight	Messick + Kirkpatrick (livelli 1-4)	Autonomia studente, autorità docente, intervento umano

Dimensione ALTAI	Fondamento pedagogico	Fuoco dell'assessment significativo
Robustezza tecnica e sicurezza	Messick (validità di costruito e predittiva)	Allineamento tra output algoritmici e reali esiti di apprendimento
Privacy e governance dei dati	Stufflebeam (CIPP)	Stewardship istituzionale, consenso, conformità
Trasparenza	Messick + Kirkpatrick (livello 2)	Interpretabilità e utilità delle spiegazioni per stakeholder
Diversità, equità, non-discriminazione	Messick (validità consequenziale)	Monitoraggio equità e impatti differenziali per gruppi
Benessere sociale e ambientale	Kirkpatrick + Stufflebeam (prodotto)	Contributo alla cultura istituzionale di apprendimento
Accountability	Stufflebeam (governance dei processi)	Strutture di governance, responsività istituzionale

Un test empirico preliminare dell'applicabilità del modello, condotto su 14 università italiane con una checklist a 27 voci e scoring ordinale a quattro livelli (Fleiss's kappa = 0,626; Krippendorff's alpha = 0,838), è riportato in Manganello, Nico, Ragusa e Boccuzzi (2025). I risultati hanno mostrato sostanziale eterogeneità, con un Institutional Index tra 0,00 e 60,32. Le dimensioni di robustezza tecnica presentavano la copertura più bassa ($M = 19,64$), mentre il benessere sociale risultava l'area più documentata ($M = 52,38$). Le istituzioni tendono a formalizzare impegni aspirazionali piuttosto che salvaguardie operative, con scarsa capacità di verifica della validità e dell'equità delle valutazioni algoritmiche.

5. Discussione: ricadute pratiche per la formazione e la ricerca sulla didattica universitaria

I risultati presentati nelle sezioni precedenti configurano un quadro sufficientemente articolato da consentire alcune considerazioni di ordine pratico, rivolte specificamente a chi si occupa di formazione alle metodologie didattiche e di ricerca sulla didattica nell'istruzione superiore. La comunità dei formatori universitari si trova oggi in una posizione peculiare rispetto all'IA valutativa: non è né la progettista dei sistemi né la sola responsabile della loro governance istituzionale, ma è l'attore che più direttamente media tra le logiche tecniche e normative dei sistemi e le pratiche quotidiane di insegnamento e valutazione.

5.1 La formazione docente come condizione di governance

Il primo ordine di implicazioni riguarda la formazione. Il requisito di supervisione umana sostanziale imposto dall'AI Act non può essere soddisfatto da docenti che non abbiano una comprensione funzionale del funzionamento dei sistemi di IA con cui interagiscono, dei loro limiti epistemici e delle loro potenziali fonti di errore sistematico. Questo non significa che ogni docente debba acquisire competenze tecniche approfondite, ma che i programmi di formazione alla didattica universitaria debbano integrare, con urgenza crescente, moduli dedicati all'IA applicata alla valutazione: cosa fanno questi sistemi, cosa non possono fare, come si riconoscono i bias algoritmici, come si interpreta una spiegazione generata da un modello di IA e come si contestualizza rispetto al giudizio pedagogico.

I dati empirici raccolti sulle università italiane suggeriscono che questa consapevolezza è ancora largamente assente a livello istituzionale: le politiche formali di governance dell'IA sono adottate da poco più di un ateneo su otto, e i documenti esistenti privilegiano dichiarazioni di principio piuttosto che procedure operative (Manganello, Nico, Ragusa & Boccuzzi, 2025). I centri e i gruppi di lavoro che si occupano di formazione didattica nelle università italiane si trovano quindi in una posizione strategica: possono anticipare l'adeguamento istituzionale progettando e sperimentando percorsi formativi che mettano i docenti in condizione di esercitare una supervisione reale, non meramente formale. In questo quadro, il XAI-ED CAF può svolgere anche una funzione formativa, non solo di governance istituzionale. Le sue sette dimensioni ALTAI, reinterpretate in chiave pedagogica attraverso Messick, Kirkpatrick e Stufflebeam, forniscono una mappa concettuale che i docenti possono utilizzare per interrogarsi

sistematicamente sulle scelte valutative: un sistema produce esiti validi nel senso di Messick? I suoi effetti cambiano ai quattro livelli di Kirkpatrick? Il contesto istituzionale censito da Stufflebeam è compatibile con un'integrazione responsabile? Formulare queste domande in modo strutturato è già, di per sé, un atto di supervisione sostanziale.

5.2 Il design della valutazione come scelta etica

Un secondo ordine di implicazioni riguarda il design delle pratiche valutative. Le tre tensioni critiche identificate dalla ricerca, quella tra trasparenza e complessità, quella tra efficienza e autonomia dello studente e quella tra personalizzazione e privacy, non sono problemi tecnici risolvibili dal fornitore del sistema: sono problemi pedagogici che richiedono scelte esplicite da parte del docente e dell'istituzione. Adottare un sistema di feedback automatizzato senza domandarsi in che misura esso riduca la capacità degli studenti di autoregolare il proprio apprendimento è una scelta pedagogica, anche quando viene effettuata per omissione.

Il modello ibrido proposto dalla ricerca sull'ibridazione della valutazione educativa a partire dalla scienza giuridica offre una chiave di lettura utile: come nei sistemi giudiziari sperimentali nei quali l'algoritmo genera una bozza di giudizio che il giudice umano rivede e riconfigura, anche nella valutazione universitaria l'IA può svolgere una funzione di supporto alla decisione senza sostituirsi al giudizio pedagogico (Boccuzzi, Nico & Manganello, 2025c). Ciò presuppone però che il docente disponga degli strumenti interpretativi necessari per esercitare tale giudizio in modo effettivo.

5.3 La governance istituzionale come responsabilità condivisa

Il terzo ordine di implicazioni coinvolge il livello istituzionale. Il XAI-ED CAF, con la sua articolazione in sette dimensioni ancorate a teorie valutative consolidate, può essere utilizzato non solo come strumento di ricerca ma come base per audit istituzionali periodici della qualità etica e pedagogica dei sistemi di IA in uso. In questo senso, i dipartimenti universitari e i senati accademici sono chiamati a sviluppare strutture di governance partecipativa, che includano docenti, studenti e personale tecnico, capaci di valutare in modo continuo se i

sistemi adottati producano esiti validi ed equi rispetto agli obiettivi formativi dichiarati.

Un'implicazione concreta riguarda l'integrazione degli indicatori di governance dell'IA nei sistemi nazionali di assicurazione della qualità. Se la verifica della validità e dell'equità delle valutazioni algoritmiche rimane esterna ai processi di accreditamento, il rischio è che la conformità formale all'AI Act si traduca in adempimento burocratico senza ricadute sulla qualità dell'esperienza valutativa degli studenti. La ricerca suggerisce che questo rischio è reale e già documentabile nel contesto italiano (Manganello, Nico, Ragusa & Boccuzzi, 2025). Una risposta concreta potrebbe consistere nell'integrare un sottoinsieme selezionato degli indicatori XAI-ED CAF nelle procedure di autovalutazione richieste dagli organismi nazionali di assicurazione della qualità, trasformando la governance dell'IA valutativa da adempimento autonomo dei singoli atenei a dimensione sistematicamente verificabile del sistema universitario. Ciò consentirebbe di raccogliere nel tempo dati comparabili tra istituzioni diverse, colmando progressivamente quella lacuna empirica che il campione esplorativo di 14 atenei analizzato in Manganello, Nico, Ragusa e Boccuzzi (2025) non può, per definizione, colmare da solo.

6. Conclusioni

Il presente contributo ha ricostruito un percorso di ricerca che muove dalla mappatura bibliometrica di un campo emergente verso la proposta di un framework di governance teoricamente fondato. I risultati principali si sintetizzano in tre punti. Primo, la ricerca sull'IA spiegabile nella valutazione universitaria è in crescita sostenuta ma rimane frammentata sul piano etico e normativo, con cluster tematici che procedono in parallelo senza convergere verso strumenti operativi condivisi. Secondo, il quadro normativo europeo impone obblighi cogenti alle università che adottano l'IA nella valutazione, ma non orienta a sufficienza la configurazione dei sistemi ibridi, oggi i più diffusi nella prassi. Terzo, il XAI-ED CAF offre una risposta teoricamente più solida rispetto al Framework XAI-ED, integrando i requisiti ALTAI con le tradizioni della teoria valutativa educativa e fornendo criteri di

evidenza per l'autovalutazione istituzionale. Rispetto ai lavori precedenti del medesimo programma di ricerca, questo contributo svolge una funzione di sintesi che nessuno dei singoli studi componenti era nella posizione di assolvere: mostrare come analisi bibliometrica, revisione narrativa, riflessione normativa e sviluppo teorico siano fasi di un unico percorso argomentativo che approda a un criterio condiviso per valutare la qualità etica e pedagogica dei sistemi di IA nella valutazione accademica.

Il contributo presenta limiti che meritano di essere esplicitati. Il dato empirico sulla governance degli atenei italiani non deriva da una rilevazione condotta per questo lavoro, ma riprende in sintesi i risultati già pubblicati in Manganello, Nico, Ragusa e Boccuzzi (2025), cui si rimanda per la discussione dei limiti di generalizzabilità del campione. Le revisioni della letteratura, pur sistematiche nel metodo, operano su un campo in rapida evoluzione, con nuove pubblicazioni rilevanti a cadenza mensile. Il XAI-ED CAF è, per autodichiarazione degli autori, uno strumento teoricamente derivato che attende validazione empirica attraverso studi Delphi e applicazioni istituzionali estese.

Le prospettive di sviluppo riguardano la validazione del framework attraverso consultazioni con esperti, l'integrazione dei suoi indicatori nei sistemi di assicurazione della qualità e la traduzione dei requisiti normativi in percorsi formativi per i docenti. Su quest'ultimo versante, la collaborazione tra centri didattici e gruppi di ricerca sulla formazione universitaria appare la strada più percorribile.

Riferimenti bibliografici

- Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE Access*, 6, 52138-52160. <https://doi.org/10.1109/ACCESS.2018.2870052>
- Boccuzzi, G., Manganello, F., & Nico, A. (2025, settembre). Delegated Authority and Algorithmic Power. A Rapid Review of Ethical Issues in AI-Based Educational Assessment. In *Proceedings of the 2025 International Conference on Information Technology for Social Good* (pp. 1-5). <https://doi.org/10.1145/3748699.3749766>
- Boccuzzi, G., Nico, A., & Manganello, F. (2025a). Educational assessment in the age of AI: A narrative review on definitions and ethical-legal principles for trustworthy automated systems. In M. B. Nunes, P. Isaías, T. Issa, & T. Issa (Eds.), *Proceedings of the 19th International Conference on e-Learning and Digital Learning and the Sustainability, Technology and Education Conference (ELDL + STE 2025)* (pp. 129-136). IADIS. ISBN 978-989-8704-68-9
- Boccuzzi, G., Nico, A., & Manganello, F. (2025b). Harmonizing human and algorithmic assessment: Legal reflections on the right to explainability in education. In *EDULEARN25 Proceedings* (pp. 9489-9493). <https://doi.org/10.21125/edulearn.2025.2446>
- Boccuzzi, G., Nico, A., & Manganello, F. (2025c). Hybridizing educational assessment: A theoretical model from judicial science. In *HAI-WS 2025: Workshops at the Fourth International Conference on Hybrid Human-Artificial Intelligence (HAI)* (Vol. 4074). CEUR-WS. <https://ceur-ws.org/Vol-4074/short6-14.pdf>
- European Commission (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*. Official Journal of the European Union.
- Kirkpatrick, D. L., & Kirkpatrick, J. D. (1998). *Evaluating Training Programs: The Four Levels* (2nd ed.). Berrett-Koehler.

- Manganello, F., Nico, A., & Boccuzzi, G. (2025a, dicembre). Mapping the research landscape of transparent AI in university assessment: A bibliometric investigation. In *Proceedings of the 5th International Conference on AI Research*. Academic Conferences and Publishing Limited. <https://doi.org/10.34190/icaire.5.1.4117>
- Manganello, F., Nico, A., & Boccuzzi, G. (2025b). Theoretical foundations for governing AI-based learning outcome assessment in high-risk educational contexts. *Information*, 16(9), 814. <https://doi.org/10.3390/info16090814>
- Manganello, F., Nico, A., Ragusa, M., & Boccuzzi, G. (2025). Testing the applicability of a governance checklist for high-risk AI-based learning outcome assessment in Italian universities under the EU AI Act annex III. *Frontiers in Artificial Intelligence*, 8, 1718613. <https://doi.org/10.3389/frai.2025.1718613>
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741-749.
- Stufflebeam, D. L. (2003). The CIPP model for evaluation. In T. Kellaghan & D. L. Stufflebeam (Eds.), *International Handbook of Educational Evaluation* (pp. 31-62). Kluwer.