



30
December 2024

Special Issue on

The Contribution of Artificial Intelligence
to the Qualification of Educational Processes

*Il contributo dell'intelligenza artificiale
alla qualificazione dei processi di istruzione*

Edited by
Gaetano Domenici

Gaetano Domenici

Editoriale / *Editorial*

L'intelligenza artificiale generativa per l'innalzamento
della qualità dell'istruzione e la fioritura del pensiero critico. 11
Quale contributo?

*(Generative Artificial Intelligence for Increasing the Quality of Education
and the Flourishing of Critical Thinking. What Kind of Contribution?)*

STUDI E CONTRIBUTI DI RICERCA

STUDIES AND RESEARCH CONTRIBUTIONS

Giancarlo Fortino - Fabrizio Mangione - Francesco Pupo
Intersezione tra intelligenza artificiale generativa e educazione: 25
un'ipotesi

*(Intersection between Generative Artificial Intelligence and Education:
A Hypothesis)*

<i>Stefano Moriggi - Mario Pireddu</i> Apprendere (con) l'intelligenza artificiale. Un approccio media-archeologico <i>(Learning (with) Artificial Intelligence. A Media-Archaeological Approach)</i>	53
<i>Roberto Trinchero</i> Usi intelligenti dell'intelligenza artificiale. Il <i>man-with-the-machine</i> learning <i>(Intelligent Uses of Artificial Intelligence. The Man-with-the-Machine Learning)</i>	65
<i>Giovanna Di Rosario - Matteo Ciastellardi</i> The Integration of Artificial Intelligence in Communication Design. Case Studies from the Polytechnic of Milan: from Digital Culture to Sociology of Media <i>(L'integrazione dell'intelligenza artificiale nel design della comunicazione. Casi di studio del Politecnico di Milano: dalla cultura digitale alla sociologia dei media)</i>	83
<i>Massimo Marcuccio - Maria Elena Tassinari - Vanessa Lo Turco</i> Progettare e valutare con il supporto dell'intelligenza artificiale: elementi per un approccio critico all'uso dei chatbot <i>(Designing and Assessing with the Support of Artificial Intelligence: Elements for a Critical Approach to the Use of Chatbots)</i>	105
<i>Maria Luongo - Michela Ponticorvo - Maria Beatrice Ligorio</i> <i>Pietro Crescenzo - Giuseppe Ritella</i> Artificial Intelligence to Enhance Qualitative Research: Methodological Reflections on a Pilot Study <i>(L'intelligenza artificiale per potenziare la ricerca qualitativa: riflessioni metodologiche su uno studio pilota)</i>	119
<i>Daniele Dragoni - Massimo Margottini</i> L'intelligenza artificiale generativa: rischi e opportunità in ambito educativo. Il progetto «CounselorBot» per il supporto tutoriale <i>(Generative Artificial Intelligence: Risks and Opportunities in Education. The «CounselorBot» Project for Tutorial Support)</i>	137
<i>Stefania Nirchi - Giuseppina Rita Jose Mangione</i> <i>Conny De Vincenzo - Maria Chiara Pettenati</i> Indagine esplorativa sulla percezione dei docenti neoassunti circa l'impiego dell'intelligenza artificiale nella didattica: punti di forza, ostacoli e prospettive	151

(Exploratory Survey on Newly Recruited Teachers' Perceptions of the Use of Artificial Intelligence in Teaching: Strong Points, Obstacles and Perspectives)

Donatella Padua

Artificial intelligence and Quality Education: The Need for Digital Culture in Teaching 181

(Intelligenza artificiale e istruzione di qualità: la necessità della cultura digitale nell'insegnamento)

NOTE DI RICERCA

RESEARCH NOTES

Cristiano Corsini

Una valutazione col pilota automatico? Una riflessione sulle cose che possiamo guadagnare e quelle che rischiamo di perdere impiegando l'intelligenza artificiale nei processi valutativi 197

(Evaluation on Autopilot? A Reflection on the Things We Can Gain and Those We Risk Losing by Using Artificial Intelligence in Evaluation Processes)

Alessio Fabiano

Per un nuovo paradigma educativo tra intelligenza artificiale, curriculum e cittadinanza digitale. Una prima riflessione 209

(For a New Educational Paradigm between Artificial Intelligence, Curriculum and Digital Citizenship. A First Reflection)

Nazarena Patrizi - Angelo Girolami - Claudia Crescenzi

Il contributo dell'intelligenza artificiale per la qualificazione dei processi di istruzione 225

(The Contribution of Artificial Intelligence to the Qualification of Education Processes)

Fiorella D'Ambrosio

Intelligenza artificiale e istruzione: tra sperimentazione e prospettive evolutive 243

(Artificial Intelligence and Education: Between Experimentation and Evolutionary Perspectives)

COMMENTI, RIFLESSIONI, PRESENTAZIONI,
RESOCONTI, DIBATTITI, INTERVISTE

COMMENTS, REFLECTIONS, PRESENTATIONS,
REPORTS, DEBATES, INTERVIEWS

Giuseppe Spadafora

L'esperienza e il metodo dell'intelligenza nel pensiero di John Dewey 259

(Experience and the Method of the Intelligence in John Dewey's Thought)

Teodora Pezzano

La teoria dell'Arco Riflesso e l'educazione. L'esperienza come questione didattica nel pensiero di John Dewey 269

(The Reflex Arc Theory and Education. Experience as Didactic Issue in John Dewey's Thought)

Author Guidelines 281

Una valutazione col pilota automatico?

Una riflessione sulle cose che possiamo guadagnare e quelle che rischiamo di perdere impiegando l'intelligenza artificiale nei processi valutativi

Cristiano Corsini

Università degli Studi Roma Tre - Department of Education (Italy)

DOI: <https://doi.org/10.7358/ecps-2024-030-corc>

cristiano.corsini@uniroma3.it

EVALUATION ON AUTOPILOT? A REFLECTION ON THE THINGS WE CAN GAIN AND THOSE WE RISK LOSING BY USING ARTIFICIAL INTELLIGENCE IN EVALUATION PROCESSES

ABSTRACT

The paper analyses pros and cons of using generative artificial intelligence in the students' learning assessment process. In particular, the article focuses on the possibility offered by programs such as ChatGPT to reduce the time devoted to complex operations such as construction of multiple-choice questions or classification of open-ended answers. Thus, it is possible to use the generated content to support a phase of the evaluation process that normally dilutes time and makes the workload difficult to sustain. Although the benefits from this side are significant, the contribution emphasises that it is quite inappropriate to claim that generative artificial intelligence can take over the responsibility for judgement processing. Firstly, since there are no objective evaluations, such a choice would merely conceal the necessarily subjective criteria used to make well-founded value judgements. Secondly, such a choice would imply a deresponsibility with respect to judgements that have relevant practical consequences and a renunciation of the possibility of actively and consciously employing the evaluation process for the enrichment of future experiences.

Keywords: Artificial intelligence; Evaluation; Learning; Objectivity; Responsibility.

*Potrei ma non voglio fidarmi di te
Io non ti conosco e in fondo non c'è
In quello che dici qualcosa che pensi
Sei solo la copia di mille riassunti.*

Samuele Bersani, *Giudizi universali*

Quali sono i rischi e quali gli scenari connessi all'impiego dell'intelligenza artificiale nei processi valutativi? Proverò a ragionare su questa domanda partendo dal mio lavoro. Nel corso dell'ultimo anno ho iniziato a impiegare programmi di intelligenza artificiale generativa a supporto dei processi valutativi. In particolare, ho usato ChatGPT nella fase di formulazione di domande per il test a scelta multipla che precede il colloquio d'esame nei miei appelli di Docimologia e Pedagogia sperimentale e valutazione scolastica. Il colloquio costituisce l'elemento decisivo ai fini della valutazione degli apprendimenti di studentesse e studenti. Tuttavia, la prova a scelta multipla ha una sua rilevante funzione. Infatti, essa, sebbene da sola non sia sufficiente a valutare competenze (ma – da sola – nessuna prova lo è), fornisce una stima utile a individuare le studentesse e gli studenti che mostrano di possedere le conoscenze adatte per affrontare il colloquio d'esame.

La formulazione di domande per i test a scelta multipla è un processo delicato e richiede tempo, attenzione e consapevolezza. Troppo spesso, a scuola e in università, chi costruisce queste domande non è cosciente di quanto sia difficile scegliere gli stimoli giusti (tipicamente, le domande, ovvero la parte iniziale dell'item) e formulare le adeguate alternative di risposta (sia quella corretta sia quelle sbagliate, ovvero i distrattori).

Come può accertare chiunque svolga attività di formazione in ambito valutativo, le «regole» che dovrebbero guidare la formulazione di domande a scelta multipla (*Tabb. 1 e 2*) non sono universalmente note né nelle scuole né nelle università. Per esempio, sono sin troppo diffuse prassi come quella di inserire un distrattore del tutto assurdo tra le alternative di risposta inesatte o quella di proporre alternative di risposta che rimandano ad altre risposte (del tipo «nessuna delle precedenti», «tutte le precedenti», «sia a sia b» ecc.). Si tratta di prassi che rappresentano una costante minaccia alla validità e all'affidabilità dell'accertamento basato sulle prove a scelta multipla. Inserendo un distrattore assurdo non facciamo altro che aumentare la possibilità di rispondere correttamente scegliendo a caso tra le alternative plausibili superstiti, mentre la scelta di proporre astrusi rimandi ad altre alternative compromette la chiarezza e la leggibilità della domanda.

Va però considerato che se queste prassi sono diffuse è perché presentano un vantaggio fondamentale: consentono di risparmiare tempo nella costruzione delle prove. Infatti, un'attenta costruzione delle domande

richiede molto tempo (possiamo dire che in un certo senso il tempo risparmiato nella «correzione» della prova è inversamente proporzionale a quello investito nella sua costruzione) e questo porta a prediligere soluzioni affrettate basate su luoghi comuni e saperi impliciti. La scelta di inserire un'alternativa assurda come distrattore non ha dunque alcun fondamento dal punto di vista della validità dell'accertamento e deve la sua diffusione all'abitudine all'esposizione a simili domande e al fatto che il mancato sviluppo di competenze metodologico-didattiche rende processi come quelli valutativi preda delle prime affrettate e apparenti soluzioni a disposizione.

*Tabella 1. – Alcuni suggerimenti per costruire prove oggettive
(riadattato da Gattullo & Giovannini, 1989 e da Benvenuto, 2003).*

AVVERTENZE	GIUSTIFICAZIONI
Il linguaggio (termini e struttura delle proposizioni) non sia inutilmente complicato, ma adeguato ai destinatari	La misurazione degli obiettivi sarebbe accompagnata o filtrata dalla misurazione delle abilità linguistiche riguardanti tali complicazioni
Gli stimoli siano brevi ed essenziali, nella misura del possibile	Si perde meno tempo, si provoca minore stanchezza
Non fare tranelli	Si misurerebbe l'abilità di sfuggire ai tranelli
Non fare domande alle quali si possa rispondere solo in base al buon senso o alla cultura generale	Si misurerebbe il buon senso e la cultura generale
Le risposte sbagliate non siano ingenuie né raffinate, ma adeguate alla preparazione dei destinatari	Si misurerebbero acquisizioni estranee agli obiettivi perseguiti. Taluni soggetti potrebbero trovarsi in difficoltà

Possiamo considerare la valutazione educativa un processo che porta a formulare un giudizio di valore sulla distanza tra obiettivi e realtà in modo da informare scelte volte a ridurre tale distanza. Il processo di misurazione (o, se si preferisce, di accertamento, rilevazione, verifica, osservazione ecc.) della distanza tra realtà e obiettivi è inserito in una posizione intermedia. Infatti, esso sta tra due atti valutativi (prima di misurare facciamo delle scelte valoriali sul perché, sul come, sul cosa ecc., dopo aver misurato facciamo delle scelte valoriali sul giudizio e sulle sue forme comunicative). Come ricordava Aldo Visalberghi (1955), la misurazione nasce dalla valutazione e nella valutazione confluisce. La scelta di usare strumenti «oggettivi» di misurazione non rende oggettiva la valutazione ma consente di ridurre l'esposizione del processo di accertamento agli elementi di distorsione che normalmente lo caratterizzano (Domenici, 1993; Benvenuto, 2003; Corsini, 2023).

Tabella 2. – Alcuni suggerimenti per la formulazione dei quesiti (riadattato da Benvenuto, 2003).

In generale: comunicare con chiarezza gli obiettivi della prova e fornire esempi di risposta.

VERO / FALSO	
Costruire molti item	La possibilità di rispondere correttamente a caso è elevata (50%)
Evitare negazioni semplici o doppie	Rendono ambiguo lo stimolo
Usare affermazioni indipendenti tra loro	Il rispondere correttamente a un'affermazione non deve pregiudicare o facilitare la risposta ad altre
COMPLETAMENTO	
Usare spazi (buchi) identici per ampiezza	La diversa ampiezza potrebbe orientare le risposte
Scegliere le parole da cancellare in base all'obiettivo	Se vengono cancellate parole in maniera casuale, la prova tenderà a rilevare la comprensione del testo. Se vengono cancellati in maniera mirata elementi come nella tab. 16, la prova rileverà abilità e conoscenze grammaticali
CORRISPONDENZA	
Evitare di costruire una corrispondenza «perfetta»	Le due liste dovrebbero essere disomogenee per numero, in modo tale da forzare la scelta delle risposte rimanenti
Almeno uno dei due elenchi dovrà contenere elementi semplici (date, singole parole, ecc.)	Per evitare ambiguità nella risposta
SCELTA MULTIPLA	
La domanda deve focalizzare un solo problema o concetto, esprimendolo con precisione	Per evitare incertezze nella scelta della risposta e nell'interpretazione della stessa
Tutti i distrattori (risposte sbagliate) devono essere plausibili rispetto al problema considerato	L'introduzione di distrattori non plausibili renderebbe più facile rispondere correttamente senza conoscere la risposta esatta
I distrattori e la risposta esatta devono essere formulati in modo omogeneo, per struttura sintattica e per lunghezza	La presenza di alternative molto più corte o molto più lunghe potrebbe orientare le risposte
Evitare alternative del tipo: nessuna di queste, sia a) che c), tutte le precedenti	Non sono omogenee rispetto alle altre, inseriscono elementi di complicazione nella comprensione della prova che sono generalmente indipendenti dalle conoscenze e dalle abilità oggetto di rilevazione
Evitare che elementi grammaticali o la struttura della frase favoriscano l'individuazione della risposta esatta	Abilità e conoscenze di natura grammaticale, non necessariamente oggetto di misura, inciderebbero sulla prestazione
Le domande dovrebbero evitare di contenere negazioni semplici o doppie	Rendono ambiguo lo stimolo. Se è proprio indispensabile usare una negazione semplice, è consigliabile evidenziarla (per es. in grassetto)

Se è vero che non esistono valutazioni oggettive, viene da chiedersi se e quanto una macchina possa supportare il processo valutativo. A tale scopo, nelle pagine che seguono descriverò il contributo apportato da ChatGPT alle mie valutazioni in università. Prima è però necessario tornare sull'assenza di valutazioni oggettive e proporre alcune considerazioni sul legame tra tale assenza e la dimensione politica della valutazione (Macdonald, 1977; House & Howe, 2000).

Non esistono valutazioni oggettive perché ogni processo valutativo comporta decisioni soggettive su finalità, obiettivi, criteri, tempi, contenuti, forme ecc. che governano il processo valutativo. Tuttavia, se è vero che non esistono valutazioni oggettive, non è affatto detto che ogni valutazione sia arbitraria. Infatti, nelle faccende umane l'alternativa all'oggettività non è necessariamente l'arbitrarietà: fortunatamente, ci è data anche la possibilità di essere intersoggettivi. L'intersoggettività nella valutazione è basata su due dinamiche. La prima è la trasparenza rispetto a finalità, obiettivi, criteri, tempi, contenuti, forme ecc. che governano il processo. Chi valuta sceglie se, quando e quanto esplicitare le proprie scelte rispetto a queste fondamentali questioni. La seconda dimensione è la condivisione, ovvero il grado di partecipazione attiva al processo valutativo da parte di tutti gli individui coinvolti. Chi valuta sceglie se gli altri soggetti coinvolti nel processo possono assumere decisioni rilevanti in relazione a elementi come i contenuti o i tempi oppure – nei casi di autovalutazione – alla formulazione del giudizio. Trasparenza e condivisione assumono una funzione fondamentale se intendiamo usare la valutazione per promuovere apprendimenti significativi (Boud, 2000; Earl, 2013; Gentile, 2019; Trincherò, 2021).

Le scelte su trasparenza e condivisione ci ricordano inoltre che la valutazione ha una sua dimensione politica. Chi valuta gestisce un potere e presentare le proprie valutazioni come oggettive significa mistificare la realtà, legittimando come naturali le proprie scelte senza sottoporle al vaglio del confronto con altri esseri umani. È proprio su questo tema che ho chiesto a ChatGPT di aiutarmi a formulare delle domande per la prova a scelta multipla.

Ho dunque chiesto alla macchina di formulare delle domande a scelta multipla sul testo che segue, relativo alla gestione autocratica del potere valutativo.

La dimensione autocratica della valutazione

La dimensione politica della valutazione impone di tenere in debita considerazione l'incidenza dei rapporti di potere sulla conduzione dell'accertamento degli apprendimenti. Chi valuta gestisce un potere. Sempre. Negare che la valutazione sia una forma di gestione del potere è un atto di pura mistificazione. Considerata la ritrosia a respirare la stessa aria d'un secondino è bene sapere che la valutazione ha il potere di generare conseguenze che possono

essere annichilenti o emancipanti. Chi valuta sceglie perché, cosa, come valutare. E, più di ogni artificio retorico, la valutazione rivela la funzione che chi insegna si attribuisce. Di questo erano pienamente consapevoli Aldo Visalberghi, don Lorenzo Milani, Mario Lodi e altre figure di spicco della nostra storia educativa. E così, se è vero che si insegna per asservire o per liberare, è vero anche che valutiamo per riprodurre o per trasformare. Si tratta di decisioni che troppo spesso non vengono percepite come tali e rimangono a uno stadio tacito, sommerse dal quotidiano ossequio all'abitudine sintetizzata da frasi come «abbiamo sempre fatto così». Eppure, sono scelte fondamentali: incidono sulle conseguenze generate dal processo e orientano in maniera decisiva gli apprendimenti.

Il potere gestito nei processi valutativi può essere amministrato autocraticamente o democraticamente, più e meno condiviso, finalizzato al controllo o allo sviluppo di altri esseri umani. Con il beneficio d'inventario, considerando che ogni schematizzazione sacrifica sfumature intermedie e provando a non cullare speranze nell'esistenza di un potere buono in sé, è possibile proporre quattro livelli nella condivisione del potere da parte dell'insegnante che valuta: la monarchia assoluta, la monarchia costituzionale, la democrazia rappresentativa, la democrazia diretta.

Un primo livello, sin troppo diffuso e per nulla educativo, è quello della gestione autocratica. In tale scenario, chi valuta tende a escludere altri soggetti dal processo valutativo, negando loro non solo la partecipazione attiva alla fase di elaborazione del giudizio, ma anche l'accesso al perché, al cosa, al come della valutazione. Tipicamente, in questo caso abbiamo docenti che valutano senza esplicitare il ragionamento seguito nella formulazione del giudizio. «Buono, otto, a posto» e la faccenda si chiude lì, al netto di vaghi enunciati moralistici basati su aggettivi e avverbi che tendono a fare riferimento ai tratti personali di chi apprende anche quando sembrerebbero rivolti alla sua prestazione o, peggio ancora, a incentrarsi su una meschina comparazione con compagne e compagni.

Il perché, il come, il cosa della valutazione non sono comunicati. Questo significa in primo luogo che la valutazione coincide col voto (numerico o meno che sia). Un'identificazione che viene presto introiettata e naturalizzata al punto che, fuori e dentro le scuole, per gran parte delle persone valutare senza voto equivale a non valutare. Non è un caso che la scuola dell'Infanzia o la primaria finlandese, dove si valuta facendo a meno dei voti, vengano erroneamente descritte come scuole «in cui non si valuta».

In secondo luogo, in questo primo scenario è assente o sottaciuto l'obiettivo dell'attività svolta, espressione del legame profondo esistente tra il perché si insegnano e si apprendono certe cose e le specifiche esperienze di manipolazione e trasformazione della realtà a esse associate. D'altra parte, il potere valutativo qui dominante ritiene che a motivare l'esperienza di apprendimento basti il voto. E, dal canto suo, il voto fa quel che può, quel che non può non fa: agisce come bastone e carota, con quel che ne consegue. Ne consegue che, paradossalmente, le discipline – ovvero quelle sbalorditive formalizzazioni di

idee e prassi che hanno consentito all'umanità di conferire senso alla realtà per trasformarla – vengono troppo spesso percepite e liquidate da chi apprende come faccende tediose, astratte, separate dalla propria esperienza di vita. Così, chi è attrezzato per apprendere apprende, chi non lo è no.

Sta di fatto che questa modalità di gestione del potere valutativo, nella sua autoreferenzialità, tende a inibire il posizionamento attivo di studentesse e studenti rispetto alla valutazione del proprio apprendimento. Inoltre, le distorsioni che, come vedremo, esercitano generalmente un certo peso nell'accertamento degli apprendimenti, agiscono particolarmente indisturbate soprattutto laddove le modalità di elaborazione del giudizio vengono sottratte allo sguardo altrui. L'opacità tende a incidere negativamente sulla validità e sull'affidabilità dell'operazione ma, questioni più tecnicamente docimologiche a parte, quello che preme sottolineare è il fatto che in questo caso il soggetto sottoposto a valutazione, spogliato di ogni diritto deliberativo e alienato rispetto al proprio apprendimento, è oggetto di una valutazione che è autocratica perché arbitraria ed è arbitraria perché autocratica. Una valutazione che si pone come termine e come scopo (sta alla fine, dunque è il fine) di una routine che non avverte il bisogno di esplicitare la propria ragion d'essere né di mettere alla prova dell'esperienza la propria efficacia. In tal modo, il mantra domanda/risposta/voto viene assunto in maniera irriflessa e perpetuato come elemento che incontestabilmente, necessariamente e – soprattutto – naturalmente caratterizza ogni appropriata attività didattica. E così, troppo spesso, chi insegna, avendo passato tra i banchi più tempo del resto della popolazione, introietta siffatta tiritera come se fosse l'unica modalità d'azione possibile. Non importa che la litania funzioni o meno, quel che conta è che fuori da essa non v'è salvezza. Il totem di questa valutazione, ovvero il voto, considerato fine (nella duplice accezione di termine e scopo) dell'apprendimento, agisce in regime di totale opacità e consente a chi insegna di rifugiarsi nella propria routine, evitando ogni confronto con l'esperienza e rifuggendo lo sguardo di chi dovrebbe imparare qualcosa da questa esperienza. Ed è inevitabile che studentesse e studenti qualcosa apprendano. Considerato che è impossibile istruire senza educare, tendenzialmente imparano a esercitare una sudditanza strumentale e a evitare di correre meraviglioso rischio di reclamare potere di controllo sul come, sul cosa e sul perché del proprio apprendimento.

Di seguito, riporto la domanda formulata da ChatGPT.

Qual è il principale problema della gestione autocratica della valutazione?

- A) La partecipazione attiva degli studenti è incoraggiata.
- B) Gli studenti hanno pieno accesso al ragionamento dietro la valutazione.
- C) *Gli studenti sono esclusi dal processo valutativo.
- D) La valutazione è sempre trasparente e chiara.
- E) Sia A sia B.

* Risposta esatta.

Come si può notare, lo stimolo (Qual è il principale problema della gestione autocratica della valutazione?) è nel complesso ben costruito, così come la chiave (Gli studenti sono esclusi dal processo valutativo). Tuttavia, ChatGPT, lavorando sulla costruzione di risposte che appaiano sensate per gli esseri umani, ha optato per scelte popolari ma del tutto stereotipate, inserendo distrattori confusivi e non plausibili rispetto al problema considerato.

Ho dunque scelto di lavorare su questo aspetto, modificando lo stimolo rivolto alla macchina:

Trovo molto pertinenti le domande. Hai lavorato bene anche sulla risposta corretta. Ora ti chiedo la cortesia di migliorare le risposte sbagliate per renderle più plausibili. Ricorda anche di evitare alternative come «sia A sia B», «nessuna delle precedenti» ecc.

Attraverso interazioni progressive, ho potuto accertare un certo miglioramento nelle successive domande. Le ho trovate ancora in parte simili a domande formulate da persone senza solide competenze nel testing e questo significa che per capire che sono generate automaticamente serve uno sguardo competente. Tuttavia, le domande sono risultate utili per costruire una buona prova. In breve tempo, esse mi hanno fornito moltissimi suggerimenti (stimoli, chiavi e distrattori) da modificare e riassemblare. Ne ho concluso che, sebbene non abbia ovviamente senso chiedere a ChatGPT di assumersi responsabilità valutative, possiamo usare il materiale generato per supportare una fase del processo valutativo che altrimenti richiederebbe un tempo eccessivo.

COSA ABBIAMO DA PERDERE?

Una ventina d'anni fa iniziai a lavorare nella valutazione delle risposte alle domande aperte e complesse poste dall'OCSE-PISA. Si trattava di centinaia di migliaia di risposte aperte fornite da quindicenni italiani a quesiti di comprensione della lettura. Per la valutazione di ciascuna risposta come membri del gruppo di valutatori e valutatrici eravamo forniti di linee guida ed esempi che avrebbero dovuto consentirci di attribuire il codice corrispondente a ciascuna di esse. Tuttavia, come è ovvio, ogni giorno ci imbattevamo in una risicata ma costante minoranza di risposte difficili da inquadrare in un codice. Ne discutevamo, ci confrontavamo e poi assumevamo le nostre decisioni in merito. Su alcune risposte (una minoranza di una minoranza) non raggiungevamo però un accordo. In tal caso, chiedevamo al consorzio internazionale che gestiva l'OCSE-PISA di assumere una decisione. Lo stesso facevano tutti i paesi partecipanti. Il consorzio

proponeva la sua codifica che e ogni giorno le decisioni assunte sulle risposte anomale erano condivise e divenivano a loro volta *esempi di risposta* da inserire nelle linee guida alla valutazione.

Ha senso automatizzare questa operazione? Probabilmente se lo scopo è quello di ottenere informazioni su un campione della popolazione ha senso automatizzare una parte rilevante di questa operazione. Questa scelta ci consentirebbe di formulare molte domande a risposta aperta e complessa e rendere più valide le prove, dato che le domande a scelta multipla non consentono che di verificare la capacità di riconoscere le risposte esatte e non di costruirne.

Tuttavia, la scelta ha meno senso se lo scopo è valutare la comprensione del testo di singoli individui. In primo luogo, perché il margine d'errore è statisticamente sostenibile su gruppi ampi di individui, non su singoli individui. Inoltre, perché se vogliamo valutare la competenza nella comprensione del testo di un singolo individuo sappiamo che non è sufficiente sottoporre questo individuo a un'unica prova, ma è necessario documentare la sua comprensione di testi diversi in momenti diversi, attraverso diverse attività di lettura e diversi dispositivi di accertamento.

Se scegliessimo di incentrare la valutazione delle competenze di singoli individui sulle prove standardizzate commetteremmo un errore molto grave sia dal punto di vista statistico sia dal punto di vista docimologico. Se poi usassimo simili prove per rilasciare etichette di «eccellenza» o «fragilità» ecco che l'errore assumerebbe una terza dimensione, legata alle nefaste conseguenze che una siffatta valutazione avrebbe nel successivo percorso scolastico, lavorativo o universitario dei singoli individui.

Il problema è che questa scelta l'abbiamo già fatta.

Infatti, somministrando le prove INVALSI per valutare quelle che vengono definite le competenze di ogni singolo studente noi abbiamo già commesso questo errore. C'è anche un'aggravante rispetto all'ipotesi prospettata partendo dall'OCSE-PISA. Infatti, l'INVALSI non somministra che un numero irrilevante di domande a risposta aperta e complessa, dato che agendo su tutta la popolazione e non su un campione non può fare affidamento, per motivi economici e di tempo, a gruppi umani di valutatori e valutatrici al lavoro sulle risposte aperte. Non avendo (o, almeno, non ancora) approntato un sistema di intelligenza artificiale a supporto della valutazione di risposte aperte e complesse, l'INVALSI somministra in larga parte domande a risposta chiusa (come quelle a scelta multipla) e dunque non è in grado di fornire informazioni adeguate sulla capacità di costruire attivamente delle risposte. Tuttavia, anche se l'istituto si dotasse di un sistema di valutazione basato sull'intelligenza artificiale, rimarrebbero intatti i problemi di natura statistica, docimologica e pedagogica già segnalati. Va

poi aggiunto che l'INVALSI, con la scelta della somministrazione computer-based, ha da tempo anche scelto di non mettere a disposizione di scuole, docenti, studentesse e studenti i contenuti concreti, ovvero le prove sostenute. Questa scelta consente all'INVALSI di riutilizzare gli stessi quesiti nei diversi anni, risparmiando dunque tempo e risorse nella loro costruzione. Tuttavia, essa rappresenta un enorme ostacolo alla possibilità di impiegare le prove in funzione formativa.

In sostanza, il processo di automatizzazione e digitalizzazione delle prove consente di risparmiare tempo e denaro e al tempo stesso compromette sia la validità dell'accertamento sia la trasparenza e la condivisione della valutazione. Questo significa che l'eventuale scelta di ricorrere all'intelligenza artificiale per supportare il processo valutativo non comporterebbe alcuna perdita. Con l'intelligenza artificiale sacrifichiamo cose alle quali abbiamo già deciso di rinunciare. Nello specifico, abbiamo già scelto di rinunciare alla possibilità di impiegare la valutazione come occasione di partecipazione attiva volta all'arricchimento delle esperienze successive di scuole, docenti, studentesse e studenti e abbiamo scelto di usarla come strumento di burocratica rendicontazione finalizzata alla classificazione di individui e contesti.

Ragionando su quello che abbiamo da guadagnare o da perdere con l'uso dell'intelligenza artificiale (Moriggi & Pireddu, 2024), piuttosto che prospettare scenari apocalittici o utopistiche integrazioni, vale forse la pena fare un inventario delle cose che abbiamo, delle cose alle quali non siamo disposti a rinunciare e di quelle che intendiamo riscattare. Altrimenti, come suggerisce Douglas Adams, rischiamo di perdere di vista ciò che è sotto i nostri occhi.

Il bagno non era grande. Le pareti erano rivestite con una vecchia tappezzeria di quercia che, data l'età e la natura dell'edificio, era probabilmente di valore inestimabile, ma per il resto l'arredamento era spoglio e istituzionale. Sul pavimento c'era un vecchio linoleum scrostato a quadretti bianchi e neri, una piccola vasca da bagno standard, ben pulita ma con macchie e schegge di smalto molto vecchie, e anche un piccolo lavabo dozzinale con uno spazzolino da denti e un dentifricio in un bicchiere di vetro accanto ai rubinetti. Sopra il lavabo, avvitato a una boiserie probabilmente di valore inestimabile, c'era un mobiletto da bagno di latta con il rivestimento a specchio. Sembrava che fosse stato ridipinto molte volte e lo specchio era macchiato sui bordi dalla condensa. Il gabinetto aveva una vecchia cassetta in ghisa con chiusura a catena. Nell'angolo c'era un vecchio armadietto di legno dipinto di crema, con accanto una vecchia sedia di legno curvo marrone, sulla quale erano appoggiati alcuni piccoli asciugamani ben piegati ma logori. C'era anche un grande cavallo che occupava la maggior parte della stanza. (Douglas Adams, *L'ispettore olistico Dirk Gently*)

RIFERIMENTI BIBLIOGRAFICI

- Benvenuto, G. (2003). *Mettere i voti a scuola*. Roma: Carocci.
- Boud, D. (2000). Sustainable assessment: Rethinking assessment for the learning society. *Studies in Continuing Education*, 22(2), 151-167.
- Corsini, C. (2023). *La valutazione che educa*. Milano: FrancoAngeli.
- Domenici, G. (1993). *Manuale della valutazione scolastica*. Roma - Bari: Laterza.
- Earl, L.M. (2013). *Assessment as learning: Using classroom assessment to maximize student learning*. Thousand Oaks, CA: Corwin Press.
- Gattullo, M., & Giovannini, M.L. (1989). *Misurare e valutare l'apprendimento nella scuola media*. Milano: B. Mondadori.
- Gentile, M. (2019). *Valutare per apprendere*. Novara: Utet - De Agostini.
- House, E., & Howe, K. (2000). Deliberative democratic evaluation. *New Directions for Evaluation*, 85, 3-12.
- MacDonald, B. (1977). A political classification of evaluation studies. In D. Hamilton, D. Jenkins, C. King, B. Mac Donald, & M. Parlett (Eds.), *Beyond the numbers game: A reader in educational evaluation*. London: Macmillan Education.
- Moriggi, S., & Pireddu, M. (2024). *L'intelligenza artificiale e i suoi fantasmi. Vivere e pensare con le reti generative*. Trento: Il Margine.
- Trincherò, R. (2021). *Valutare gli apprendimenti nella didattica universitaria*. Verona: QuiEdit.
- Visalberghi, A. (1955). *Misurazione e valutazione nel processo educativo*. Ivrea: Edizioni di Comunità.

RIASSUNTO

L'articolo analizza i vantaggi e i rischi del ricorso all'intelligenza artificiale generativa nel processo di valutazione degli apprendimenti di studentesse e studenti. In particolare, l'articolo si focalizza sulla possibilità offerta da programmi come ChatGPT di ridurre i tempi dedicati a operazioni complesse come la costruzione di domande a scelta multipla o la valutazione di risposte aperte. È dunque possibile usare il materiale generato per supportare una fase del processo valutativo che normalmente dilata i tempi e rende difficilmente sostenibile il carico di lavoro. Sebbene i vantaggi da questo punto di vista siano significativi, il contributo evidenzia come sia del tutto inappropriato affermare che l'intelligenza artificiale generativa possa assumersi responsabilità nell'elaborazione del giudizio. In primo luogo, non esistendo valutazioni oggettive, una simile scelta non farebbe altro che occultare i criteri inevitabilmente soggettivi impiegati per pervenire a fondati giudizi di valore. In secondo luogo, tale scelta implicherebbe una deresponsabi-

lizzazione rispetto a giudizi che hanno conseguenze pratiche rilevanti e una rinuncia alla possibilità di impiegare attivamente e consapevolmente il processo di valutazione ai fini dell'arricchimento di esperienze future.

Parole chiave: Apprendimento; Intelligenza artificiale; Oggettività; Responsabilità; Valutazione.

Copyright (©) 2024 Cristiano Corsini

Editorial format and graphical layout: copyright (©) LED Edizioni Universitarie



This work is licensed under a Creative Commons

Attribution-NonCommercial-NoDerivatives 4.0 International License.

How to cite this paper: Corsini, C. (2024). Una valutazione col pilota automatico? Una riflessione sulle cose che possiamo guadagnare e quelle che rischiamo di perdere impiegando l'intelligenza artificiale nei processi valutativi [Evaluation on autopilot? A reflection on the things we can gain and those we risk losing by using artificial intelligence in evaluation processes]. *Journal of Educational, Cultural and Psychological Studies (ECPS)*, 30, 197-208. <https://doi.org/10.7358/ecps-2024-030-corc>