tutor

VOL. 24 N. 1, 2024: 78 - 79
DOI: 10.13135/1971-8551/11030

Considerazioni sul bias nella pedagogia medica basata su AI

Chiara Rabbito[1], Carlo Maria Petrini[2], Antonio Vittorino Gaddi[1], Pierangelo Veltri[3], Mario Ettore Giardini[4]

[1] Società Italiana di Telemedicina
[2] Istituto Superiore di Sanità
[3] Dipartimento di Ingegneria Informatica, Modellistica, Elettronica e Sistemistica, Università della Calabria, Cosenza
[4] School of Science and Engineering, University of Dundee, UK

Contatto per la corrispondenza: mgiardini001@dundee.ac.uk

Nella presente lettera si riportano alcune riflessioni sulla progettazione di sistemi di *Medical Education* che prevedano l'uso di strumenti basati sull' Intelligenza Artificiale (AI) sia in fase di progettazione che in fase applicativa.

Al fine dell'effettivo conseguimento degli obiettivi formativi e didattici previsti, siano essi raggiunti con sistemi di tutoraggio virtuale (ad esempio, agenti conversazionali basati su AI per un supporto personalizzato agli studenti) o siano basati sulla generazione e/o erogazione e/o valutazione di contenuti formativi, è necessario procedere ad un'analisi approfondita dei requisiti di progettazione della didattica.

Questi requisiti devono includere in modo sostanziale, e dunque non con funzione meramente estetica o superficiale, quelli richiesti dall'etica e dalla legge. Per questa ragione è opportuna una analisi preventiva dedicata ad individuare e considerare *ex ante* alle regole fondamentali della progettazione, quelle dell'etica e quelle giuridiche quali essenziali al pari di quelle tecnologiche.

A titolo esemplificativo si presentano di seguito due casi particolarmente rilevanti.

Il primo concerne l'applicazione del principio etico e giuridico di non discriminazione, affermato, tra gli altri atti, dalle Linee guida della Unione Europea del 2019 e dall'*Artificial Intelligence Act* approvato in data 14 giugno 2023.

L'applicazione del principio di non discriminazione comporta la necessità di evitare nel processo di formazione i cosiddetti bias, ovvero pregiudizi nel funzionamento dei modelli

di AI causati, a loro volta, da pregiudizi impliciti contenuti nei dati che a loro volto sono stati usati per generare i modelli di AI stessi (Mehrabi et al. 2021). È di tutta evidenza come l'applicazione di tale principio sia molto rilevante in pedagogia medica sotto due profili: l'evitare qualsiasi forma di discriminazione ai danni dello studente nella fase di insegnamento, e la necessità di non trasmettere contenuti trasmissivi di attitudini discriminatorie (Zhang et al. 2024).

La corretta formazione, il corretto e lecito addestramento dei modelli diAI utili alla progettazione e realizzazione di sistemi didattici, è dunque essenziale affinché essa possa assurgere - nell'ipotesi che ciò venga proposto e attuato - al rango di "formatore" effettivo e positivo e di conseguenza affinché il discente non venga discriminato e/o apprenda nozioni discriminanti.

La seconda esemplificazione riguarda l'applicazione dell'AI ai fini della analisi del comportamento degli studenti. Anche in questo caso è necessario che tale analisi sia effettuata nel rispetto dei principi dell'etica e giuridici enucleati dagli organismi internazionali e che, di conseguenza, la progettazione dei software didattici tenga conto delle disposizioni normative, nonché e in modo particolare all' applicazione dei principi di libertà e tutela della privacy della persona.

Sono noti, infine, dalla letteratura scientifica casi recenti in cui sistemi didattici basati su AI abbiano esposto categorie di studenti ad ingiustificati rischi di marginalizzazione (Liang et al. 2023).

BIBLIOGRAFIA

Liang W., et al. (2023) GPT detectors are biased against non-native English writers. *Patterns* 4(7) n. 100779. https://doi.org/10.1016/j.patter.2023.100779

Mehrabi, N., et al. (2021). A Survey on Bias and Fairness in Machine Learning. *ACM Computing Surveys 54*(6), 115:1-115:35. https://doi.org/10.1145/3457607

Zhang, W., et al. (2024) AI in Medical Education: Global situation, effects and challenges. *Education and Information Technologies* 29, 4611–4633. https://doi.org/10.1007/s10639-023-12009-8