

Sistemi interattivi *AI-based* interpretabili e capaci di generare fiducia negli utenti - le attività del nodo AIIS del Dipartimento di Informatica dell'Università di Salerno

Pietro Battistoni, Bernardo Breve, Loredana Caruccio, Stefano Cirillo, Vincenzo Deufemia, Marianna Di Gregorio, Giuseppe Polese, Marco Romano, Monica Sebillio, Genoveffa Tortora, Giuliana Vitiello
Università di Salerno

Abstract

In questo contributo sono state sintetizzate le attività svolte dal gruppo interdisciplinare di HCI-UsE Lab¹, Lab GIS² e CAIS Lab³ presso l'Università di Salerno negli ultimi anni relativamente alla progettazione di sistemi interattivi *AI-based* e alla loro capacità di generare fiducia negli utenti.

Sono stati sviluppati metodi, tool e servizi informatici finalizzati a migliorare l'aspetto della fiducia operando in due direzioni principali: primo, tramite una progettazione innovativa del sistema orientata al potenziamento della *user experience*, e secondo, tramite la creazione di tool e tecniche di visualizzazione e di *data mining* per supportare l'interpretazione dell'AI.

1 Temi della ricerca

L'evoluzione dell'intelligenza artificiale (AI) e la sua rapida diffusione nella vita quotidiana delle persone ha aperto nuovi orizzonti e al contempo nuove sfide e problematiche legate proprio al rapporto tra gli utenti umani e la stessa AI. Tra i temi principali del gruppo interdisciplinare del Laboratorio di HCI-UsE Lab, Lab GIS e CASA Lab del Dipartimento di Informatica dell'Università di Salerno ci sono appunto quelli legati alla progettazione e sviluppo di sistemi *AI-based* capaci, da un lato, di essere facilmente interpretabili e, dall'altro, generare fiducia nei loro utenti. In questo documento mostreremo il nostro approccio tramite i lavori svolti e le sfide aperte.

2 UX design per incentivare la fiducia nell'AI

In precedenti lavori abbiamo mostrato come l'AI possa giocare un ruolo importante nell'apprendimento, infatti, in [Battistoni *et al.*, 2020a] [Battistoni *et al.*, 2020b] essa è stata utilizzata per riconoscere gli errori e dare feedback agli utenti durante esercizi di comprensione della lingua dei segni. L'obiettivo era di permettere agli utenti di apprendere dai propri errori basandosi sulle valutazioni ottenute dall'Intelligenza Artificiale. Gli studi eseguiti sull'esperienza dell'apprendimento e le osservazioni dei ricercatori ci hanno permesso di affermare che, durante le attività di apprendimento, i partecipanti hanno instaurato con il sistema una sorta di rapporto positivo "discepolo-mentore". Tuttavia, l'apprendimento è stato bidirezionale: se da un lato l'utente apprendeva dall'AI, la stessa AI apprendeva le piccole possibili variazioni nei gesti, la forma distinta delle mani degli utenti tramite un processo di inferenza. In qualche modo, sia l'intelligenza umana che quella artificiale stavano collaborando allineandosi tra loro.

Tuttavia, questo tipo di relazione non è comune e generalmente gli utenti diffidano di sistemi basati su AI a qualsiasi livello [Shneiderman, 2020]. In [Battistoni *et al.*, 2021] abbiamo mostrato come la fiducia in questo tipo di sistemi può essere aumentata progettando una *user experience* adeguata, in particolare, agendo su tre fattori: l'approccio decisionale, il tipo di interazione e i sentimenti che il sistema è capace di provocare negli utenti. L'approccio decisionale può essere totalmente autonomo, quando la decisione viene presa solamente dal sistema, oppure, può essere collaborativa, quando la decisione viene presa coinvolgendo l'utente nel processo. Dipendendo dal contesto e dagli obiettivi del sistema questa scelta può influenzare il sentimento di fiducia verso la decisione adottata.

¹ <https://www.di.unisa.it/en/department/structures?id=143>

² <https://www.di.unisa.it/en/department/structures?id=135>

³ <https://www.di.unisa.it/en/department/structures?id=137>

Il tipo di interazione influenza il modo in cui l'AI e l'utente comunicano fra di loro, molto spesso il focus della sua progettazione è esclusivamente basato sulle necessità dell'utente, sottostimando che un'interazione non adeguata per l'AI possa creare problemi di mutua comprensione che influenzano negativamente l'esperienza e la percezione degli utenti. Infine, il tipo di sentimenti generati negli utenti possono riguardare gli elementi visti in precedenza, quali l'interazione e l'approccio decisionale, e il tipo di relazione, a volte quotidiana, che viene instaurata. Per esempio, nel caso degli assistenti vocali in commercio, anche se la AI resta invisibile, questa resta attiva esplorando il contesto e resta in ascolto pronta ad assistere l'utente. Ciò risulta in un'interazione passiva ma, allo stesso tempo, continua tra l'Intelligenza Artificiale e l'utente. Ciò può generare sentimenti di sospetto e di sfiducia o, viceversa, di efficienza e la consapevolezza di poter essere assistito in qualsiasi momento.

2.1 Un nuovo approccio alla progettazione

Il classico approccio *Human Centered Design* (HCD) sembra essere troppo limitato per guidare con successo i progettisti verso un design capace di gestire adeguatamente questi fattori, soprattutto per quei progettisti alle prime armi o con scarsa conoscenza dei temi legati alle relazioni tra AI e *user experience*. Per questo motivo, la nostra ricerca è stata indirizzata a definire nuovi approcci alla progettazione di sistemi interattivi basati su AI che enfatizzino il suo ruolo e importanza già durante le prime fasi del processo di progettazione. Da qui, in [Battistoni 2021], abbiamo proposto una prima estensione del tradizionale approccio HCD introducendo dei nuovi requisiti dell'AI. Essi consentono di definire, da un lato, come il sistema possa comprendere e interagire meglio con gli utenti, e dall'altro, il modo in cui questo appaia più naturale e familiare. Requisiti di esempio riguardano l'interazione, ovvero, il modo in cui, dato il contesto, l'interazione uomo-AI possa risultare facilmente comprensibile o facilmente apprendibile dalla stessa AI tramite, per esempio, un processo di *reinforcement learning*; oppure il tipo di approccio decisionale.

Rimane sfida della ricerca comprendere come integrare al meglio il ruolo dell'AI nel processo di progettazione e come permettere ai progettisti di poter ideare, costruire e valutare prototipi, anche solo *low-fidelity*, che coinvolgano le scelte progettuali per l'AI.

2.2 Supporto alla corretta analisi ed interpretazione dei risultati

Un'altra strada percorribile per aumentare la fiducia nell'AI è quella di supportare la corretta analisi e interpretazione dei suoi risultati. Infatti, lo studio di problematiche legate all'estrazione della conoscenza ed all'inferenza di metadati potrebbe portare ad una ricca proliferazione di risultati a valle

dell'esecuzione degli algoritmi. Per questo motivo, diventa necessaria la progettazione di tecniche per il ranking e la visualizzazione dei risultati, che supportino il monitoraggio di come gli stessi variano a seguito delle modifiche sui dati e che permettano l'interazione con l'utente al fine di analizzare e validare metadati [Caruccio *et al.*, 2019] [Breve *et al.*, 2020]. In tale contesto, sono stati progettati di diversi tool [Caruccio *et al.*, 2021] [Breve *et al.*, 2021] volti alla visualizzazione dell'evoluzione delle dei risultati di algoritmi per il *mining* di metadati che includono nuove metafore di visualizzazione e modelli di interazione.

Riferimenti bibliografici

[Battistoni *et al.*, 2020a] Battistoni, P., Di Gregorio, M., Romano, M., Sebillio, M., Vitiello, G., & Solimando, G. (2020, July). Sign Language Interactive Learning-Measuring the User Engagement. In International Conference on Human-Computer Interaction (pp. 3-12). Springer, Cham.

[Battistoni *et al.*, 2020b] Battistoni, P., Di Gregorio, M., Romano, M., Sebillio, M., & Vitiello, G. (2020). AI at the edge for sign language learning support. In the International Journal of Humanized Computing and Communication Vol. 1, No. 1 (2020) 23-42 (IJHCC), KS Press.

[Battistoni *et al.*, 2021] Battistoni, P., Romano, M., Sebillio, M., & Vitiello, G. (2021, July). A Change in Perspective About Artificial Intelligence Interactive Systems Design: Human Centric, Yes, But Not Limited to. In International Conference on Human-Computer Interaction (pp. 381-390). Springer, Cham.

[Shneiderman, 2020] Ben Shneiderman (2020) Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy, International Journal of Human-Computer Interaction, 36:6, 495-504, DOI: 10.1080/10447318.2020.1741118

[Caruccio *et al.*, 2019] Caruccio, L., Deufemia, V., & Polese, G. (2019) Visualization of (Multimedia) Dependencies from Big Data. Multimedia Tools and Applications, Springer: 78(23): 33151-33167.

[Breve *et al.*, 2020] Breve, B., Caruccio, L., Cirillo, S., Deufemia, V., & Polese, G. (2020) Visualizing Dependencies during Incremental Discovery Processes. In Proceedings of Workshops of the EDBT/ICDT 2020 Joint Conference, 3rd International Workshop on Big Data Visual Exploration and Analytics (BigVis).

[Breve *et al.*, 2021] Breve, B., Caruccio, L., Cirillo, S., Deufemia, V., & Polese, G. (2021) Dependency Visualization in Data Stream Profiling. Big Data Research, Elsevier: 25: 100240.

[Caruccio *et al.*, 2021] Caruccio, L., Cirillo, S., Deufemia, V., & Polese, G. (2021) Real-time visualization of profiling metadata upon data insertions. In Proceedings of Workshops of the EDBT/ICDT 2021 Joint Conference, 4th International Workshop on Big Data Visual Exploration and Analytics (BigVis).