

IL PROCESSO DI SVILUPPO PER L'IMPLEMENTAZIONE DI UN *CHATBOT RETRIEVAL-BASED*: UN CASO STUDIO

P. Zaccaglino*, Nicola Lopolito**

L&G Solution S.r.l., Via Fratelli Biondi, 3, 71122 Foggia FG
[*p.zaccaglino@legsolution.com](mailto:p.zaccaglino@legsolution.com); [**n.lopolito@legsolution.com](mailto:n.lopolito@legsolution.com)

SOMMARIO

Nell'attuale contesto digitale, l'adozione di *chatbot* nelle piattaforme online rappresenta una strategia cruciale per migliorare l'interazione con gli utenti e ottimizzare l'efficienza operativa. I *chatbot*, alimentati dall'intelligenza artificiale e dal *Natural Language Processing* (NLP), sono progettati per fornire risposte automatiche, immediate e personalizzate. Questo studio si focalizza sull'implementazione di un *chatbot retrieval-based*, evidenziando il processo di sviluppo, la valutazione e i risultati ottenuti. Il progetto NOVA PA di LG Solution Srl, che ha portato alla creazione della piattaforma AppaltiOK, integra un *chatbot retrieval-based* per supportare le interazioni con gli utenti. A differenza dei *chatbot* generativi, il *chatbot retrieval-based* seleziona risposte da un archivio predefinito, basandosi sulla similarità semantica delle domande degli utenti, garantendo risposte accurate e coerenti. Il presente articolo mostra, mediante il caso studio, gli step per lo sviluppo di tale metodologia. Il processo di sviluppo ha incluso la selezione di un *dataset* italiano di *Semantic Textual Similarity* (STS) e la sperimentazione di vari modelli di linguaggio multi-lingua, con il modello "*distiluse-base-multilingual-cased-v1*" scelto per le sue prestazioni superiori. L'architettura del modello si basa su un encoder simile a BERT, con un focus sul *Multi-Head Attention* per analizzare e comprendere il contesto delle domande. I test hanno dimostrato che il modello addestrato con tecniche di cosine *similarity loss* e *metric learning* ha migliorato significativamente le prestazioni, evidenziando un'alta efficienza operativa con risposte rapide e pertinenti. Infine, il *chatbot* ha dimostrato un'alta efficienza operativa, fornendo risposte rapide e confermando la sua idoneità per l'uso in contesti reali.

1. INTRODUZIONE

Nell'era digitale contemporanea, l'adozione dei *chatbot* nelle piattaforme online è divenuta una strategia fondamentale per migliorare l'interazione con gli utenti e ottimizzare l'efficienza operativa. Questi strumenti, basati sull'intelligenza artificiale e il *Natural Language Processing* (NLP), sono progettati per rispondere automaticamente alle domande degli utenti, fornendo assistenza immediata e personalizzata.

I primi passi nell'utilizzo della NLP ci riportano al 1950 [1] quando Alan Turing propose un programma tale che quando comunicano con gli utenti, questi ultimi non si rendono conto che stanno parlando con un software ma pensano di parlare con un essere umano. Alan Turing definì un test basato su questo presupposto, ora chiamato test di Turing. Sono stati adottati molti approcci diversi per risolvere il test di Turing. Oggi siamo arrivati a un punto in cui spesso è necessario che un *chatbot* si presenti agli umani come *bot* prima di iniziare una conversazione.

Oltre a supportare le funzionalità di un *chatbot* o di un assistente virtuale, la NLP viene utilizzata in molti altri campi come il riconoscimento vocale, la traduzione automatica, il *text mining*, la classificazione del testo e l'analisi del *sentiment*. La necessità di utilizzare *chatbot* e la corrispondente tecnologia NLP ha creato un nuovo settore del software specializzato nella creazione di *chatbot*. Sono

state sviluppate molte nuove società di software che offrono soluzioni diverse. Grandi aziende come Google, Amazon, Microsoft, IBM e Facebook offrono piattaforme di sviluppo *Chatbot* (CDP: *Chatbot Development Platforms*) complete come infrastruttura *cloud*. Le aziende più piccole stanno approfittando di queste piattaforme, agendo come intermediari aiutando altri clienti a utilizzare i *chatbot* nei loro servizi o utilizzando un CDP per le proprie esigenze.

Gli obiettivi principali dell'uso dei *chatbot* nelle piattaforme moderne comprendono miglioramenti significativi sia dal punto di vista dell'esperienza utente sia dal punto di vista operativo delle aziende.

Uno degli obiettivi primari dei *chatbot* è migliorare l'esperienza degli utenti sulle piattaforme digitali. I *chatbot* offrono risposte rapide e pertinenti, riducendo i tempi di attesa e aumentando la soddisfazione degli utenti. Uno studio condotto da Dasa et al. [2] ha dimostrato che l'uso dei *chatbot* nei servizi clienti ha incrementato significativamente la soddisfazione degli utenti grazie alla capacità di rispondere tempestivamente alle loro esigenze. Inoltre, un altro importante obiettivo è garantire la disponibilità continua del supporto, 24/24, 7/7. Questo è particolarmente utile per le piattaforme che operano a livello globale e devono rispondere a utenti in diverse zone orarie. La disponibilità costante dei *chatbot* contribuisce a migliorare la percezione del servizio da parte degli utenti, offrendo un supporto ininterrotto senza necessità di intervento

umano. Un approccio personale nel contatto digitale con il cliente è vitale per quanto riguarda il cliente e la sua preservazione [3]. Secondo Siswi e Wahyono [4], la soddisfazione si crea soddisfacendo le aspettative dei clienti. Ottimizzando l'esperienza del cliente, è possibile superare le aspettative e la soddisfazione del cliente. La soddisfazione del cliente porta alla fidelizzazione del cliente, che è importante da decenni [5]. I *chatbot* permettono anche alle aziende di automatizzare una vasta gamma di processi, riducendo così i costi operativi. Questo include la gestione delle richieste di supporto, la fornitura di informazioni e l'esecuzione di compiti ripetitivi. Uno studio di Innis e La Londe [6] ha dimostrato che l'implementazione dei *chatbot* può portare a una significativa riduzione dei costi operativi, migliorando al contempo l'efficienza dei processi interni. Un altro obiettivo cruciale è la personalizzazione del servizio. Grazie alle tecnologie di *machine learning* e analisi dei dati, i *chatbot* sono in grado di offrire un servizio su misura per ogni utente. Essi possono analizzare il comportamento degli utenti e adattare le risposte in base alle loro preferenze e necessità specifiche. [7] Kooli ha sottolineato come i *chatbot* possano migliorare l'apprendimento e l'interazione personalizzata, rendendo l'esperienza dell'utente più coinvolgente e rilevante.

I *chatbot*, infine, possono raccogliere una grande quantità di dati sulle interazioni con gli utenti. Questi dati possono essere analizzati per comprendere meglio le esigenze degli utenti, identificare trend e migliorare continuamente il servizio offerto. È da evidenziare anche come l'analisi dei dati derivati dall'interazione con i *chatbot* rappresenta un'opportunità cruciale per le aziende per affinare i propri servizi e sviluppare nuove funzionalità che rispondano meglio alle esigenze dei loro clienti. Un articolo di Hassani e Silva [8] evidenzia come i *chatbot*, come ChatGPT, possano assistere i *data scientist* nell'automazione di varie attività, inclusa la pulizia e il *pre-processing* dei dati, l'addestramento dei modelli e l'interpretazione dei risultati. Questi *chatbot* possono analizzare dati non strutturati, fornendo nuove intuizioni e migliorando i processi decisionali. Inoltre, l'articolo sottolinea l'importanza di sfruttare le capacità dei *chatbot* per personalizzare le interazioni con gli utenti, raccogliendo dati che possono essere utilizzati per migliorare continuamente il servizio offerto e sviluppare nuove funzionalità basate sulle esigenze degli utenti. Questo approccio non solo aumenta l'efficienza e la precisione dei flussi di lavoro, ma consente anche di identificare modelli e tendenze nei dati degli utenti che possono guidare lo sviluppo di funzionalità innovative e migliorare l'esperienza complessiva del cliente. La personalizzazione basata sui dati raccolti dai *chatbot* può portare a una maggiore soddisfazione e fedeltà dei clienti, offrendo un vantaggio competitivo significativo alle aziende.

Nel contesto della trasformazione digitale della Pubblica Amministrazione, il progetto NOVA PA¹, di LG Solution Srl, ha portato alla creazione di AppaltiOK, una piattaforma avanzata progettata per migliorare l'interazione tra le imprese fornitrici e le stazioni appaltanti che utilizzando la piattaforma Traspire. Questa piattaforma non solo facilita la gestione delle gare d'appalto, ma prevede anche l'integrazione di un *chatbot retrieval-based* per supportare gli utenti in modo veloce e preciso. Grazie al *chatbot*, gli utenti possono ricevere risposte immediate a domande frequenti e ottenere assistenza guidata nell'utilizzo delle varie funzionalità della piattaforma.

2. IL CASO STUDIO: IL *CHATBOT RETRIEVAL-BASED*

Il *chatbot* implementato in NOVA PA si basa su un approccio *retrieval-based*, il quale seleziona le risposte da un archivio predefinito in base alla similarità semantica delle domande poste dagli utenti. A differenza dei *chatbot* generativi che creano risposte ex novo, il *chatbot retrieval-based* di AppaltiOK utilizza tecniche avanzate di NLP per abbinare le domande degli utenti alle risposte più pertinenti disponibili nel database. L'obiettivo principale del *chatbot retrieval-based* è migliorare l'efficienza del supporto agli utenti, riducendo i tempi di attesa e fornendo risposte accurate e coerenti. Questo approccio garantisce che gli utenti ricevano informazioni affidabili e pertinenti, migliorando l'esperienza complessiva sulla piattaforma.

Dal punto di vista tecnico, a differenza dei *chatbot* generativi come ChatGPT, i *chatbot* basati su *retrieval* sono chiamati a gestire il compito del NLP di STS per identificare la domanda archiviata che presentasse la maggiore affinità semantica con la domanda posta dall'utente.

L'STS rappresenta una sfida significativa a causa delle complessità sintattiche e semantiche del testo. Infatti, variazioni anche minime nel modo in cui l'utente formula una domanda possono portare a similitudini sintattiche elevate ma a significati completamente diversi, e viceversa. Un esempio di questa dinamica è illustrato in Figura 1.

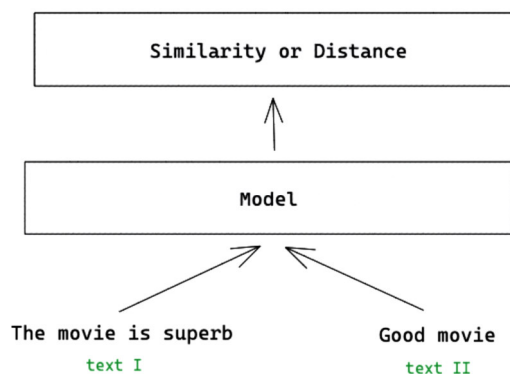


Figura 1: Esempio di *Semantic Textual Similarity*

¹ Progetto cofinanziato con il FESR Puglia POR Puglia 2014-20 – Progetto NOVA PA - CODICE PRATICA: SJA8705

3. IL PROCESSO DI SVILUPPO

L'implementazione del *chatbot retrieval-based* per la piattaforma Nova PA ha seguito un processo di sviluppo strutturato e metodico, volto a garantire prestazioni ottimali nel rispondere alle domande degli utenti. Di seguito, i principali passaggi seguiti nel processo di sviluppo.

La selezione del *dataset*

È stato selezionato un *dataset* italiano di STS dalla piattaforma *HuggingFace* per addestrare e valutare il modello. Questo *dataset* è stato scelto per la sua struttura che include coppie di frasi con punteggi di similarità, essenziali per il task di STS.

I modelli utilizzati

Sono stati testati vari modelli di linguaggio multi-lingua preaddestrati, tra cui *distiluse-base-multilingual-cased-v1* e *paraphrase-multilingual-mpnet-base-v2*. Il primo è stato selezionato per le sue prestazioni superiori. I risultati ottenuti, mostrati in tabella 1, comprendono i punteggi di MSE, MAE e EM per tutti i modelli testati.

Tabella 1: Risultati dei test sui modelli utilizzati

Modello	MAE	MSE	EM
<i>distiluse-base-multilingual-cased-v1</i>	0.85	1.16	35.68%
<i>distiluse-base-multilingual-cased-v2</i>	0.93	1.4	34.40%
<i>paraphrase-multilingual-mpnet-base-v2</i>	0.93	1.41	34.70%
<i>paraphrase-multilingual-MiniLM-L12-v2</i>	0.89	1.34	37.35%
<i>clips/mfaq</i>	2.3	7.49	11.24%
<i>aiknowyou/aiky-sentence-bertino</i>	1.87	5.15	16.24%
<i>efederici/sentence-bert-base</i>	1.05	1.78	31.70%

Sulla base di questi risultati, il modello "*distiluse-base-multilingual-cased-v1*" è stato selezionato per procedere con la fase successiva del progetto, che includeva l'addestramento. I risultati di Tabella 1 offrono un'analisi comparativa delle prestazioni di vari modelli di linguaggio multi-lingua preaddestrati, utilizzando tre metriche principali: *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE) e *Exact Match* (EM). Il modello "*distiluse-base-multilingual-cased-v1*" si distingue per le sue prestazioni eccellenti, con un MAE di 0.85 e un MSE di 1.16, i più bassi tra tutti i modelli testati. Il suo EM è del 35.68%, indicando una buona accuratezza complessiva.

Il modello "*distiluse-base-multilingual-cased-v2*" mostra un MAE di 0.93 e un MSE di 1.40, entrambi leggermente superiori rispetto alla versione v1. Tuttavia, mantiene buone prestazioni generali con un EM del 34.40%. Questi risultati indicano una leggera diminuzione nella precisione delle risposte esatte rispetto alla versione precedente.

Per il modello "*paraphrase-multilingual-mpnet-base-v2*", il MAE e il MSE sono rispettivamente di 0.93 e 1.41. Il suo EM, pari al 34.70%, è leggermente superiore a quello della versione *cased-v2*, suggerendo un miglioramento nella corrispondenza esatta delle risposte.

Il modello "*paraphrase-multilingual-MiniLM-L12-v2*" mostra le migliori prestazioni in termini di EM, con un valore del 37.35%. Il MAE e il MSE, rispettivamente 0.89 e 1.34, sono competitivi, indicando che questo modello è altamente performante, soprattutto nella precisione delle risposte esatte.

Al contrario, il modello "*clips/mfaq*" ha prestazioni significativamente peggiori, con un MAE di 2.30 e un MSE di 7.49, accompagnati da un EM molto basso dell'11.24%. Questo indica che il modello non è adeguato per il task specifico di similarità testuale semantica in contesto multi-lingua.

Il modello "*aiknowyou/aiky-sentence-bertino*" presenta un MAE di 1.87 e un MSE di 5.15, con un EM del 16.24%. Queste metriche inferiori rispetto ai modelli *distiluse* e *paraphrase* suggeriscono che le sue capacità di previsione sono meno accurate.

Infine, il modello "*efederici/sentence-bert-base*" mostra un MAE di 1.05 e un MSE di 1.78. Il suo EM del 31.70% indica una discreta precisione nel fornire risposte esatte, sebbene le sue prestazioni complessive siano inferiori rispetto ai migliori modelli *distiluse* e *paraphrase*.

Da quanto detto è, quindi, possibile dedurre che il modello "*distiluse-base-multilingual-cased-v1*" emerge come il più equilibrato e performante tra tutti, grazie ai suoi valori di MAE e MSE più bassi e a un buon livello di EM.

Tuttavia, il modello "*paraphrase-multilingual-MiniLM-L12-v2*" offre la migliore precisione per risposte esatte, nonostante i suoi MAE e MSE siano leggermente superiori.

Al contrario, modelli come "*clips/mfaq*" e "*aiknowyou/aiky-sentence-bertino*" dimostrano prestazioni significativamente inferiori, suggerendo che non sono adatti per l'implementazione in un sistema di *chatbot retrieval-based* focalizzato sulla similarità semantica testuale.

L'architettura del modello

Il modello si basa sull'architettura *Transformer* [9] con solo l'*encoder*, simile a BERT [10]. Utilizza il meccanismo di *Multi-Head Attention* per analizzare e comprendere il contesto delle domande poste dagli utenti. Originariamente, il *Transformer* includeva sia l'*encoder* che il *decoder*, come mostrato nella figura sotto. Tuttavia, la scelta dell'architettura specifica dipende dallo scopo del task. Per esempio, i *Large Language Model* come GPT-4 [11] adottano un'architettura solo *decoder*. In contrasto, per compiti che riguardano la rappresentazione del testo, come STS, si usa l'architettura con solo *encoder*, tipica di modelli come BERT. Una caratteristica distintiva dei modelli basati su architettura *Transformer* è l'impiego del *layer* di *Attention*, specificatamente del tipo *Multi-Head Attention*. Questo meccanismo è fondamentale per il funzionamento di tali modelli poiché permette di stabilire e analizzare le relazioni tra le parole di un testo. Attraverso l'*Attention*, il modello può valutare l'importanza relativa di ciascuna parola nel contesto delle altre, migliorando signi-

ficativamente la comprensione del testo. Più nel dettaglio, il *Multi-Head Attention* permette al modello di esaminare le informazioni da diverse prospettive contemporaneamente. Ogni “testa” nell’unità di *Multi-Head Attention* si concentra su diversi aspetti della relazione tra le parole, il che consente una rappresentazione più ricca e una migliore capacità di catturare varie dinamiche linguistiche. Questa funzionalità è risultata particolarmente utile per il task di comprensione del linguaggio naturale, come STS, dove la contestualizzazione delle parole è cruciale. Di seguito è illustrata l’architettura della pipeline utilizzata per predire la similarità tra due frasi. In particolare, dopo aver codificato il significato di entrambe le frasi nei vettori “u” e “v”, viene calcolata la loro similarità coseno per assegnare un punteggio (si veda la Figura 2).

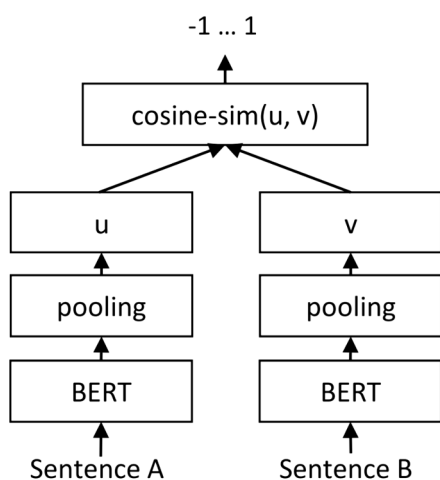


Figura 2: architettura della pipeline utilizzata

4. TEST E DISCUSSIONE DEI RISULTATI

Dopo aver sviluppato il modello, è stato essenziale eseguire una serie di test rigorosi per valutare le sue prestazioni e l’efficacia operativa.

Di seguito sono riportati i passaggi chiave e le considerazioni sui risultati ottenuti:

Addestramento del modello.

È stato utilizzato un approccio di cosine *similarity loss* per migliorare la capacità del modello di prevedere accuratamente i punteggi di similarità tra coppie di frasi. Un processo di *metric learning* con *triplet loss* è stato impiegato per aumentare l’accuratezza nelle metriche *Rank@N*, migliorando ulteriormente le prestazioni del modello. Nel contesto della valutazione delle prestazioni di modelli di apprendimento automatico, in particolare nei sistemi di recupero delle informazioni e nei modelli di classificazione, il termine *Rank@N* rappresenta una metrica utilizzata per misurare l’accuratezza del modello nel recuperare le risposte corrette entro una determinata posizione nell’elenco dei risultati ordinati. *Rank@N* indica, quindi, la percentuale di volte in cui la risposta corretta si trova tra le prime N posizioni dei risultati ordinati restituiti dal model-

lo. Ad esempio, *Rank@1* Indica la percentuale di volte in cui la risposta corretta si trova esattamente nella prima posizione dell’elenco dei risultati ordinati.

Applicazione su dataset proprietario.

Il *dataset* aziendale iniziale di circa 50 coppie domanda-risposta è stato ampliato tramite tecniche di *data augmentation*, generando versioni alternative delle domande originali per arricchire il *dataset*. Questo ha permesso di testare più efficacemente la qualità del modello.

Valutazione delle prestazioni.

Le prestazioni del modello sono state monitorate utilizzando varie metriche, tra cui MSE, MAE e EM. Queste metriche hanno permesso di valutare l’accuratezza e la precisione del modello.

Sperimentazione e testing.

Il modello è stato sottoposto a test preliminari e finali per verificarne la capacità di generalizzare e mantenere un livello elevato di prestazione su dati diversi non utilizzati durante la fase di addestramento. In uno scenario simulato con richieste parallele da parte di più utenti, il modello ha mostrato un’alta efficienza, fornendo risposte in circa 0.06 millisecondi per ogni utente, confermando la sua idoneità per contesti reali che richiedono interazioni rapide. Nelle tabelle 2 e 3 sono riportati i risultati dei test eseguiti.

Tabella 2: Risultati del Test set

Model	Rank@1	Rank@2	Rank@3	Rank@4	Rank@5
Untrained	82.05%	90.38%	94.23%	98.08%	99.36%
Trained (STS)	77.56%	88.46%	94.23%	97.44%	98.72%
Trained (ML)	80.13%	90.38%	95.51%	97.44%	98.72%

Tabella 3: Risultati del Validation set

Model	Rank@1	Rank@2	Rank@3	Rank@4	Rank@5
Untrained	87.39%	94.96%	96.64%	97.48%	97.48%
Trained (STS)	89.09%	94.96%	95.80%	97.48%	97.48%
Trained (ML)	89.09%	94.96%	95.80%	97.48%	97.48%

Le due tabelle presentano una valutazione dettagliata delle prestazioni del modello di *chatbot retrieval-based* sviluppato, esaminando i risultati ottenuti sui *Test set* e *Validation set*. Le considerazioni sui risultati offrono una comprensione approfondita dell’efficacia e dell’efficienza del modello nei diversi scenari di utilizzo. Nel *Test set*, il modello non addestrato ha mostrato risultati promettenti, con una precisione del 82.05% su *Rank@1* e del 90.38% su *Rank@2*. Il modello, cioè, possiede una capacità di base per riconoscere e recuperare risposte rilevanti, anche senza alcun addestramento specifico. Tuttavia, l’addestramento con il *dataset* di STS ha portato a un miglioramento nelle prestazioni di alcune metriche. Ad esempio, la precisione per *Rank@1* è scesa al 77.56%, ma le presta-

zioni per *Rank@3* e superiori sono rimaste elevate, mostrando una capacità migliorata di trovare risposte rilevanti entro le prime tre posizioni.

Nel *Validation set*, l'addestramento STS ha portato a un significativo miglioramento delle prestazioni, specialmente per *Rank@1* e *Rank@2*, che hanno raggiunto rispettivamente l'89.09% e il 94.96%. Questo suggerisce che il modello è diventato più preciso nel recuperare le risposte più rilevanti nelle prime posizioni, grazie all'addestramento specifico ricevuto. L'addestramento con *metric learning* (ML) ha ulteriormente migliorato le prestazioni del modello. Nel *Test set*, la precisione per *Rank@1* è salita all'80.13% e per *Rank@3* ha raggiunto il 95.51%, suggerendo un miglioramento nel recupero delle risposte più rilevanti. Nel *Validation set*, il modello addestrato con ML ha mantenuto prestazioni elevate, con *Rank@1* al 89.09% e *Rank@2* invariato rispetto all'addestramento STS, dimostrando robustezza e capacità di mantenere alte prestazioni anche su dati non visti durante l'addestramento.

In generale, quindi, i risultati mostrano che entrambi gli approcci di addestramento, STS e ML, hanno migliorato le prestazioni del modello rispetto alla versione non addestrata. Tuttavia, l'addestramento ML ha portato a un miglioramento maggiore in termini di precisione nel *Test set*, mentre nel *Validation set* le differenze tra i due metodi di addestramento sono risultate minime. Le prestazioni simili tra i due *set* indicano che il modello ha una buona capacità di generalizzare e mantenere alte prestazioni su dati non visti durante l'addestramento.

5. CONCLUSIONI

Il processo di test e valutazione ha dimostrato che il *chatbot retrieval-based* sviluppato per NOVA PA è altamente efficiente e accurato nel rispondere alle domande degli utenti. Grazie a un addestramento rigoroso, all'applicazione su un *dataset* proprietario

ampliato e a test approfonditi, il modello è stato in grado di fornire un supporto tempestivo e preciso, migliorando l'esperienza utente sulla piattaforma e ottimizzando l'efficienza operativa.ⁱ

BIBLIOGRAFIA

- [1] E. Adamopoulou and L. Moussiades, "Chatbots: History, technology, and applications," *Mach. Learn. Appl.*, vol. 2, p. 100006, Dec. 2020.
- [2] A. Daza, W. F. Peralta Robles, and J. A. Salazar Jiménez, "The Impact of Chatbots on Customer Satisfaction: A Systematic Literature Review," *TEM J.*, vol. 12, no. 3, 2023.
- [3] G. McLean and A. Wilson, "Evolving the online customer experience ... is there a role for online customer support?," *Comput. Hum. Behav.*, vol. 60, pp. 602–610, Jul. 2016.
- [4] A. A. Siswi and W. Wahyono, "The Role of Customer Satisfaction in Increasing Customer Loyalty," *Manag. Anal. J.*, vol. 9, no. 1, pp. 17–25, Mar. 2020.
- [5] P. B. Brandtzaeg and A. Følstad, "Why People Use Chatbots," in *Internet Science*, vol. 10673, I. Kompatsiaris, J. Cave, A. Satsiou, G. Carle, A. Passani, E. Kontopoulos, S. Diplaris, and D. McMillan, Eds. Cham: Springer International Publishing, 2017, pp. 377–392.
- [6] D. E. Innis and B. J. La Londe, "Modelling the Effects of Customer Service Performance on Purchase Intentions in the Channel," *J. Mark. Theory Pract.*, vol. 2, no. 2, pp. 45–60, Apr. 1994.
- [7] C. Kooli, "Chatbots in Education and Research: A Critical Examination of Ethical Implications and Solutions," *Sustainability*, vol. 15, no. 7, p. 5614, Mar. 2023.
- [8] H. Hassani and E. S. Silva, "The Role of ChatGPT in Data Science: How AI-Assisted Conversational Interfaces Are Revolutionizing the Field," *Big Data Cogn. Comput.*, vol. 7, no. 2, p. 62, Mar. 2023.
- [9] A. Vaswani et al., "Attention Is All You Need." *arXiv*, 2017.
- [10] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." *arXiv*, 2018.
- [11] OpenAI et al., "GPT-4 Technical Report." *arXiv*, 2023.

ⁱ Gli autori esprimono la loro gratitudine per la collaborazione e il supporto scientifico ricevuto dall'Università di Bologna e dalla KAD3 s.r.l., che hanno fornito risorse e conoscenze fondamentali per la realizzazione del progetto.