



Università
Ca'Foscari
Venezia

Corso di Laurea Magistrale
in Economia e Gestione delle Aziende

Tesi di Laurea

**L'impatto delle Intelligenze Artificiali Generative
e la ridefinizione del contesto lavorativo:
il caso del Customer Service**

Relatore

Prof. Massimiliano Nuccio

Correlatore

Prof. Carlo Romano Marcello Alessandro Santagiustina

Laureando

Riccardo Celotto
Matricola 889538

Anno Accademico

2022/2023

Sommario

INTRODUZIONE	1
1. EVOLUZIONE DELLE AI	3
1.1. Che Cos'è l'AI	3
1.1.1. <u>Nozioni Base</u>	3
1.1.1. <u>Contrasto con le Tecnologie Tradizionali</u>	4
1.2. Differenze dalle Tecnologie Tradizionali	6
1.2.1. <u>Programmazione vs Apprendimento</u>	6
1.2.2. <u>Machine Learning (ML) vs Deep Learning (DL)</u>	9
1.3. Che cosa sono le AI Generative	12
1.3.1. <u>Definizioni Base</u>	12
1.3.2. <u>Differenze dalle AI "Standard"</u>	14
1.4. Tecnologie Chiave	14
1.4.1. <u>Le Generative Adversarial Network (GANs)</u>	15
1.4.2. <u>I Large Language Models (LLMs)</u>	16
1.4.3. <u>I Variational Auto Encoders (VAEs)</u>	17
2. IL RUOLO DEI MODELLI DI LLM E DI GPT NELLO SVILUPPO DELLE INTELLIGENZE ARTIFICIALI GENERATIVE	19
2.1. Evoluzione dei Transformer e dei Large Language Models	19
2.1.1. <u>Architettura e Funzionalità dei Transformer</u>	19
2.1.2. <u>Importanza dei Transformer nello sviluppo dei LLM</u>	22
2.1.3. <u>I Large Language Models (LLM) e il Loro Ruolo nei Motori di NLP</u>	24
2.2. Generative Pre-Trained Transformer (GPT) e Chat GPT	25
2.2.1. <u>Che cos'è un GPT</u>	25
2.2.2. <u>Che cos'è ChatGPT</u>	27
2.2.3. <u>Foundation Models, Fine Tuning e Prompt Engineering</u>	29
2.3. Impatto di GPT nelle AI Generative	33
3. IL CUSTOMER SERVICE	37
3.1. Che cos'è il Customer Service	37
3.1.1. <u>Definizioni ed Origini</u>	37
3.1.2. <u>Multidimensionalità del Customer Service</u>	39
3.2. Evoluzione del Customer Service	40

3.2.1.	<u>Evoluzione Tecnologica</u>	42
3.2.2.	<u>Trasformazione Digitale</u>	43
3.2.3.	<u>La nascita del Customer Service automatico</u>	44
3.3.	Evoluzione dei Chatbot di Customer Service Pre-GPT-3	45
3.3.1.	<u>I primi Chatbot</u>	46
3.3.2.	<u>Gli Internet Chatbot</u>	47
3.3.3.	<u>I Chatbot “Tradizionali”</u>	48
3.3.4.	<u>I Chatbot AI Conversazionali</u>	50
3.4.	I Chatbot di Intelligenza Artificiale Generativa	52
3.4.1.	<u>I Tool di Intelligenza Artificiale Generativa</u>	53
3.4.2.	<u>I Chatbot di Intelligenza Artificiale Generativa</u>	54
4.	CASO STUDIO: EUROTHERM	59
4.1.	Scopo della Ricerca e Metodologia	59
4.2.	Evoluzione e Contesto Economico	61
4.2.1.	<u>Storia</u>	61
4.2.2.	<u>Il gruppo “E-Group”</u>	62
4.2.3.	<u>Il Settore HVAC</u>	63
4.3.	Fasi della trasformazione AI	65
4.3.1.	<u>Cultura aziendale e Milestone tecnologici</u>	65
4.3.2.	<u>Avvicinamento alla AI Generativa</u>	67
4.3.3.	<u>Embrace Gen AI</u>	68
4.4.	Aree della “Trasformazione Gen AI”	70
4.4.1.	<u>Progetti pilota</u>	71
4.4.2.	<u>Chatbot di Customer Service Pre-vendita “Zeromax”</u>	78
4.4.3.	<u>Impatto Atteso vs Effettivo</u>	82
4.5.	Risultati e Commenti finali	87
4.5.1.	<u>Risultati</u>	87
4.5.2.	<u>Sfide e Prospettive</u>	91
5.	CONCLUSIONI	93
	BIBLIOGRAFIA	97
	SITOGRAFIA	101
	APPENDICE	102

INTRODUZIONE

Nell'era digitale, l'evoluzione tecnologica ha assunto un ritmo senza precedenti, portando alla nascita e allo sviluppo delle Intelligenze Artificiali Generative, strumenti capaci di ridefinire in modo significativo il tessuto del contesto lavorativo. Questa tesi intende esplorare il funzionamento e l'impiego di tali tecnologie, con un focus nel settore del Customer Service.

Le AI Generative, con la loro capacità di creare contenuti nuovi e originali, analizzare grandi quantità di dati e apprendere autonomamente da queste informazioni, si pongono come catalizzatori di un cambiamento radicale. Non solo promettono un'ottimizzazione dei processi lavorativi attraverso l'automazione e la personalizzazione ma stimolano anche una riflessione più ampia sul futuro del lavoro, sulla relazione tra uomo e macchina e sull'etica nell'uso delle tecnologie avanzate.

Attraverso l'analisi di un caso studio e l'esame delle ricerche nel campo, questa Tesi si propone di delineare il contesto in cui le AI Generative stanno emergendo come strumenti rivoluzionari. Si discuterà il potenziale di queste tecnologie per trasformare il Customer Service, migliorando l'efficienza e l'esperienza utente, e si rifletterà sulle sfide che accompagnano la loro implementazione.

L'obiettivo è quello di offrire al lettore una comprensione profonda dell'impatto delle AI Generative, preparandolo ad approcciare con una visione critica e informata un potenziale approccio a questa Tecnologia.

1. EVOLUZIONE DELLE AI

1.1. Che Cos'è l'AI

1.1.1. Nozioni Base

L'Intelligenza Artificiale (AI), nel suo senso più ampio, si riferisce alla creazione di macchine che esibiscono capacità di pensiero e apprendimento paragonabili a quelle umane. Questa definizione, radicata nella ricerca fondamentale di scienziati come Alan Turing, che nel suo celebre articolo "Computing Machinery and Intelligence" (1950) pose le basi concettuali per la simulazione dell'intelligenza umana nelle macchine, rimane una pietra miliare nel campo. La definizione di AI si è evoluta per abbracciare non solo la capacità di eseguire compiti che richiederebbero intelligenza se svolti da esseri umani, ma anche la capacità di apprendere e adattarsi autonomamente (McCarthy et al. 1955) . Il concetto di "apprendimento" è centrale in questa definizione, distanziando l'AI dalle tecnologie tradizionali che operano su istruzioni fisse e predefinite. Le macchine AI, sono programmate per analizzare, interpretare e reagire a dati e ambienti in continua evoluzione, Russell e Norvig (2016). Questa capacità di apprendimento, implementata attraverso metodi di apprendimento automatico e deep learning, permette alle macchine di migliorare le proprie performance basandosi sull'esperienza, l'aspetto chiave che le differenzia dalle tecnologie convenzionali.

L'applicazione pratica di questi concetti è evidente in studi come quello di Krichevsky et al. (2012) dove l'uso di reti neurali profonde dimostra una straordinaria capacità di apprendimento e interpretazione visiva, un tempo considerata un'esclusiva del pensiero umano. Questi sviluppi segnano un passo significativo verso la realizzazione di sistemi AI che non solo imitano, ma in alcuni casi superano, le capacità cognitive umane in specifici compiti.

In sintesi, l'AI si configura come un campo in rapida evoluzione, che cerca di dotare le macchine di una forma di intelligenza paragonabile a quella umana, non limitata alla mera esecuzione di compiti programmati, ma estesa all'apprendimento, all'adattamento e al miglioramento autonomo. Questa progressione apre nuovi orizzonti nell'applicazione dell'AI e pone questioni fondamentali sull'interazione futuro tra umanità e macchine intelligenti.

1.1.1. Contrasto con le Tecnologie Tradizionali

Nell'analizzare le distinzioni tra l'Intelligenza Artificiale (AI) e le tecnologie tradizionali, emerge una comprensione più profonda delle loro differenze fondamentali. Queste differenze non si limitano solo alle modalità operative, ma si estendono alle capacità intrinseche e alla natura stessa delle tecnologie. Esaminando queste distinzioni, possiamo apprezzare pienamente il valore aggiunto e le potenzialità superiori offerte dall'AI.

Le tecnologie tradizionali, come i sistemi basati su database o i software di elaborazione dati, si distinguono per il loro approccio alla risoluzione dei problemi e all'esecuzione dei compiti. Le tecnologie tradizionali operano all'interno di un framework deterministico, seguendo una serie di istruzioni e regole predefinite, Ré e Suciú (2007). Il fatto di avere una programmazione predeterminata garantisce che ogni azione eseguita sia il risultato diretto e di un input specifico fornito dall'utente o dal programmatore, rendendolo quindi prevedibile. Questo approccio deterministico, torna molto efficace per compiti specifici e ben definiti, ma mostra limitazioni significative quando si tratta di adattarsi a nuovi contesti o situazioni inaspettate, Nguyen et al. (2016) .

Un esempio tipico di tecnologia tradizionale può essere trovato nei fogli di calcolo tradizionali (Es. Excel), che eseguono calcoli matematici specifici basati su formule preimpostate. Mentre questi strumenti sono estremamente utili per l'elaborazione di dati in maniera ordinata e ripetitiva, non possiedono la capacità di apprendere autonomamente o di eseguire analisi più complesse senza un input o un aggiornamento manuale. Questa mancanza di autonomia e adattabilità limita significativamente il loro utilizzo in contesti dinamici o in scenari che richiedono una maggiore flessibilità cognitiva.

Al contrario, le Intelligenze Artificiali, sfruttando tecniche avanzate di machine learning e reti neurali, offrono un approccio radicalmente diverso. I computer basati su Intelligenza Artificiale non si limitano ad eseguire compiti, ma apprendono e si adattano attraverso l'interazione con i dati, Bishop (2006). Questa capacità di adattamento, detta apprendimento autonomo, permette all' Intelligenza Artificiale di superare i limiti delle tecnologie tradizionali, adattandosi continuamente a nuovi dati e contesti operativi, in modo da migliorare le proprie prestazioni nel tempo e offrire soluzioni innovative a problemi complessi.

La distinzione tra AI e tecnologie tradizionali diventa particolarmente evidente quando consideriamo la capacità dell'AI di simulare aspetti dell'intelligenza umana come il riconoscimento di schemi, il processamento del linguaggio naturale e la percezione sensoriale. A differenza dei sistemi tradizionali, che sono confinati ai parametri della loro programmazione, l'AI si avvicina sempre più a un livello di funzionalità che imita, e in alcuni casi supera, le capacità cognitive umane in compiti specifici, come il processamento del linguaggio naturale o il riconoscimento di pattern complessi, Goodfellow et al. (2016). Questo salto qualitativo nell'elaborazione e nell'analisi dei dati apre nuovi orizzonti nell'automazione e nella risoluzione dei problemi, ridefinendo il futuro dell'interazione uomo-macchina e del progresso tecnologico.

Volendo proseguire la nostra analisi relativa alle differenze tra le tecnologie cosiddette tradizionali e le Intelligenze Artificiali, possono essere identificate tre caratteristiche chiave che permettono la distinzione tra le due e dimostrino la superiorità dell'Intelligenza Artificiale secondo parametri legati alla versatilità, adattabilità ed alle capacità cognitive.

- **Apprendimento e Adattamento Autonomo:** L'Intelligenza Artificiale, attraverso tecniche come il machine learning, si evolve in base all'esperienza e ai dati. Questo permette alle macchine di apprendere e migliorare autonomamente, Bishop et al. (2020). Questo a differenza delle tecnologie tradizionali, basate su algoritmi statici e predeterminati.
- **Capacità di Elaborazione e Analisi dei Dati:** Mentre i sistemi tradizionali operano entro i limiti delle loro programmazioni iniziali che abbiamo definito pre-deterministiche, l'Intelligenza Artificiale può elaborare ed estrarre significato da set di dati vasti e complessi e non uniformi. La ricerca di Krizhevsky et al. (2012) mostra come le reti neurali profonde possano gestire e interpretare enormi quantità di dati visivi in modi inaccessibili ai sistemi convenzionali.
- **Simulazione di Processi Cognitivi Umani:** Infine, L'Intelligenza Artificiale è nota soprattutto per superare le tecnologie tradizionali quando si tratta di simulare aspetti dell'intelligenza umana come il riconoscimento di schemi, il processamento del linguaggio naturale e la percezione sensoriale. Gli studi nel campo del Deep learning, come quelli di Goodfellow e al. (2014), dimostrano come le reti neurali profonde eseguano compiti che richiedono un elevato grado di intelligenza cognitiva.

Approfondendo queste differenze, emergono chiaramente le ragioni della superiorità dell'AI. La capacità dell'AI di apprendere e adattarsi autonomamente è una rivoluzione rispetto alla staticità delle tecnologie tradizionali. Questa adattabilità non si limita a migliorare l'efficienza operativa, ma apre nuove possibilità in termini di personalizzazione e innovazione.

L'elaborazione avanzata dei dati permette all'AI di riconoscere tendenze, fare previsioni e prendere decisioni informate in modi che sarebbero impossibili per i sistemi tradizionali. La capacità dell'AI di simulare processi cognitivi umani non solo migliora l'interazione uomo-macchina, ma permette anche l'esplorazione di compiti che richiedono creatività e intuizione, tradizionalmente considerati domini esclusivi dell'intelligenza umana.

La superiorità dell'AI si manifesta nella sua straordinaria capacità di superare i confini della programmazione tradizionale, aprendo orizzonti prima inimmaginabili e ridefinendo il futuro del progresso tecnologico.

1.2. Differenze dalle Tecnologie Tradizionali

1.2.1. Programmazione vs Apprendimento

Nel contesto dell'evoluzione tecnologica, una delle distinzioni più significative si manifesta nella comparazione tra le metodologie operative delle tecnologie tradizionali e quelle dell'Intelligenza Artificiale (AI). Questa distinzione non si limita solamente alle loro applicazioni pratiche, ma si estende a una differenza fondamentale nella filosofia di base di queste tecnologie.

La programmazione tradizionale, che costituisce la base dell'informatica e dello sviluppo del software, si caratterizza per l'adozione di un approccio metodico e strutturato alla creazione di sistemi informatici. In questo paradigma, i programmatori utilizzano linguaggi di programmazione (Es. C++, Java o Python) per scrivere codici dettagliati che definiscono specifiche funzioni e operazioni eseguibili da un computer o altro dispositivo elettronico.

Come riportato in precedenza, la modalità di programmazione tradizionale si basa su algoritmi definiti in modo esplicito e segue una logica condizionale per guidare il comportamento del sistema. Si traduce che ogni aspetto del funzionamento del sistema è delineato da istruzioni predeterminate e finite, rendendo il comportamento del software del tutto prevedibile e controllato. Questa natura deterministica della programmazione tradizionale è ciò che garantisce la stabilità e l'affidabilità dei sistemi software. Come sottolineato da Brooks (1995) la prevedibilità e la ripetibilità sono aspetti fondamentali in molte applicazioni software, in particolare in quelle dove la sicurezza e la precisione sono critiche. Tuttavia, questo approccio ha le sue limitazioni, specialmente in termini di flessibilità e adattabilità. I sistemi sviluppati con programmazione tradizionale non possono imparare o adattarsi autonomamente a nuove situazioni o dati, il che può limitare la loro efficacia in scenari dinamici o in rapida evoluzione.

I limiti della programmazione tradizionale iniziano a diventare evidenti quando vengono confrontate le prestazioni con le capacità di apprendimento e adattamento delle moderne tecnologie AI. Come discusso da Brooks et al. (1995), la programmazione tradizionale fornisce una base solida per la costruzione di software affidabili con uno scopo preciso e definito, per contro l'incapacità di adattarsi autonomamente a nuove situazioni rappresenta la più grande sfida nel contesto dell'evoluzione tecnologica attuale.

Se ci si pensa, l'AI adotta un approccio radicalmente diverso, caratterizzato dall'uso di algoritmi di apprendimento automatico. Sono proprio questi algoritmi a costituire un'evoluzione fondamentale rispetto alla programmazione tradizionale, permettendo l'introduzione di concetti avanzati come la flessibilità, l'adattabilità e l'apprendimento autonomo. Questi algoritmi, rappresentanti dell'essenza dell'apprendimento automatico (Machine Learning) e dell'apprendimento profondo (Deep Learning), dotano le macchine della capacità di analizzare, imparare e prendere decisioni basate su grandi volumi di dati, superando i limiti dei sistemi tradizionali.

Un aspetto chiave di questi algoritmi è l'apprendimento basato sull'esperienza, una metodologia che si distacca dalla necessità di programmazione esplicita per ogni specifico compito. Le AI imparano attraverso l'esposizione ai dati. Le reti neurali, per esempio, imparano a riconoscere modelli e schemi attraverso l'analisi di grandi dataset. Tale processo è ben illustrato nel lavoro di Krizhevsky et al. (2012), dove hanno sviluppato una rete neurale convoluzionale capace di classificare con precisione

immagini all'interno del vasto dataset di ImageNet. Questo tipo di apprendimento profondo (Deep Learning) dimostra come le AI possano acquisire competenze complesse senza una programmazione diretta e specifica.

La flessibilità degli algoritmi di apprendimento automatico si manifesta nella loro capacità di migliorare nel tempo. Questa caratteristica è evidenziata nell'evoluzione di sistemi come AlphaGo, sviluppato da DeepMind (AlphaGo è un noto software utilizzato per giocare al gioco del Go, noto gioco tradizionale Giapponese. AlphaGo è considerata una pietra miliare nel mondo dell'apprendimento automatico dato che fu la prima volta che una macchina riuscì a sconfiggere un Maestro umano nel gioco del Go, 2015). Silver et al. (2017) hanno mostrato che AlphaGo non solo ha appreso a giocare a Go a livelli competitivi, ma ha anche sviluppato strategie innovative attraverso l'apprendimento per rinforzo, un metodo che permette al sistema di ottimizzare le sue decisioni basandosi sui risultati delle partite precedenti.

In aggiunta, gli algoritmi di machine learning sono ottimizzati per il rilevamento di correlazioni e pattern nei dati che possono sfuggire all'analisi umana o tradizionale. La loro capacità di elaborazione e analisi di grandi quantità di dati in tempo reale li rende strumenti preziosi in campi come l'analisi predittiva e la rilevazione di tendenze, Jordan e Mitchell (2015)

Nell'ambito operativo della tecnologia informatica quindi, la contrapposizione tra la programmazione tradizionale e gli algoritmi di apprendimento automatico delle Intelligenze Artificiali (AI), mette in evidenza le differenze fondamentali nel modo in cui questi due approcci, molto differenti tra loro, affrontano il trattamento dei dati e la risoluzione di problemi. Tale contrapposizione rivela non solo le loro diverse capacità operative, ma anche le varie potenzialità e limitazioni inerenti a ciascun approccio.

Uno degli aspetti chiave di questa distinzione riguarda la flessibilità operativa. Mentre la programmazione tradizionale è basata su set di istruzioni immutabili e definite, che offrono stabilità e prevedibilità, gli algoritmi di apprendimento automatico delle Intelligenze Artificiali sono intrinsecamente dinamici, riuscendo ad adattarsi ed ad apprendere continuamente da nuovi dati o ambienti. Questa differenza si traduce in una maggiore capacità delle AI di affrontare problemi non strutturati e di adattarsi a scenari in evoluzione, una flessibilità che è assente nella programmazione tradizionale.

D'altra parte, la programmazione tradizionale offre vantaggi in termini di controllo e comprensione del comportamento del sistema. La natura deterministica della programmazione tradizionale rende i sistemi più prevedibili e affidabili in applicazioni dove errori o deviazioni possono avere conseguenze gravi. In contrasto, sebbene gli algoritmi di apprendimento automatico siano potenti nell'elaborazione di grandi set di dati e nella generazione di modelli predittivi, possono talvolta mancare di trasparenza (un fenomeno noto come "black box" nel machine learning), rendendo difficile per gli utenti comprendere completamente il processo decisionale della macchina.

Inoltre, la capacità degli algoritmi di apprendimento automatico di generare soluzioni innovative e di scoprire pattern nascosti nei dati rappresenta un chiaro vantaggio rispetto alla programmazione tradizionale. Tuttavia, questo richiede una grande quantità di dati di alta qualità per l'addestramento, una risorsa che non è sempre disponibile o facilmente accessibile. La scelta tra i due approcci dipende completamente quindi dalle specifiche esigenze dell'applicazione, dalla natura dei dati disponibili secondo quantità e qualità e dagli obiettivi dell'utente finale. Questa progressione apre nuovi orizzonti nell'applicazione dell'AI e pone questioni fondamentali sull'interazione futura tra umanità e macchine intelligenti.

1.2.2. Machine Learning (ML) vs Deep Learning (DL)

Emergendo dal contesto di 'Programmazione Vs Apprendimento', discusso nel sottocapitolo precedente, Machine Learning (ML) e Deep Learning (DL) rappresentano evoluzioni significative nell'approccio all'apprendimento nell'ambito dell'intelligenza artificiale (AI).

La storia e l'evoluzione di queste tecnologie rivelano progressi significativi. Il concetto di ML risale agli anni '50 e '60 con i primi lavori su algoritmi di apprendimento, Samuel (1959), mentre il DL ha guadagnato popolarità negli ultimi due decenni, specialmente dopo i progressi nel riconoscimento dell'immagine e del linguaggio naturale LeCun et al. (2015). La crescente disponibilità di grandi set di dati e la potenza computazionale avanzata hanno contribuito a questa evoluzione, permettendo al DL di emergere come una tecnologia fondamentale nell'era moderna dell'AI.

Il Machine Learning, è il campo di studio che fornisce ai computer l'abilità di apprendere senza essere esplicitamente programmati in maniera tradizionale. Diventa quindi un

concetto che segna una svolta dall'approccio alla tradizionale programmazione, Mitchell (1997). Il Deep Learning, come sottoinsieme del Machine Learning, approfondisce questa evoluzione utilizzando livelli di reti neurali per riuscire ad apprendere dai dati, LeCun et al. (2015). Le reti neurali tendono ad imitare il funzionamento del cervello umano, permettendo al Deep Learning di elaborare dati non strutturati e complessi mantenendo livelli di profondità elevati ed una precisione senza precedenti.

Il Machine Learning utilizza algoritmi di analisi per elaborare ed interpretare pattern nei set di dati, adattandosi e efficientando gradualmente per migliorare le prestazioni, Bostrom e Yudkowsky (2014). Tali algoritmi possono variare passando da tecniche di apprendimento supervisionato a tecniche di apprendimento non supervisionato, ciascuna con applicazioni e limitazioni specifiche. Il Deep Learning, per contro, sfrutta reti neurali profonde ed ampie per gestire volumi di dati molto grandi, come dimostrato da Goodfellow et al. (2016). Tali reti sono sorprendentemente efficaci per l'identificazione di correlazioni complesse nei dati, rendendo quindi il Deep Learning indispensabile in vari campi come ad esempio il riconoscimento di immagini e lo sviluppo di linguaggio naturale.

La progressione dal Machine Learning tradizionale al Deep Learning è evidenziata non solo dalla loro capacità di elaborare dati diversi ma anche dalla crescente complessità dei problemi che possono risolvere. Questa transizione segna un punto di svolta nell'evoluzione dell'AI, spostando il focus dalla semplice automazione alla creazione di sistemi capaci di apprendimento e ragionamento autonomi.

Per andare a definire Machine Learning e Deep Learning, è necessario comprendere che il Deep Learning è a tutti gli effetti un sottoinsieme specializzato del Machine Learning. Spiegando la differenza tra i due approcci in maniera semplice, un sistema di Machine Learning può essere addestrato per riconoscere Email Spam usando tecniche di apprendimento supervisionato, che gli permette di analizzare delle caratteristiche specifiche delle email e imparare da esempi etichettati in precedenza, riconoscendo quindi le Email Spam, Friedman et al. (2001). Il Deep Learning, invece, può essere utilizzato per sviluppare un sistema di riconoscimento facciale, utilizzando reti neurali convoluzionali per elaborare e interpretare milioni di immagini facciali Krizhevsky et al. (2012).

Confrontando in maniera approfondita Machine Learning e Deep Learning si rivelano differenze significative soprattutto in termini di algoritmi e modelli utilizzati, quantità di dati richiesti ed elaborati, potenza computazionale necessaria e tipologia di use case che meglio risolvono. Gli algoritmi di Machine Learning includendo tecniche come regressioni lineari e alberi decisionali, sono tipicamente meno complessi e richiedono una minore quantità di dati per essere addestrati. Gli algoritmi di Machine Learning quindi, sono spesso utilizzati per analizzare dati strutturati o per risolvere problemi dove le relazioni tra i dati sono più evidenti Hastie et al. (2009). Il Deep Learning, invece, sfrutta le reti neurali profonde, le quali richiedono enormi set di dati per identificare sottili pattern e correlazioni (almeno decine di migliaia). Queste reti ritrovano un utilizzo particolarmente efficace per analizzare dati non strutturati, come immagini e testi, Goodfellow et al. (2016).

In termini di potenza computazionale, il Deep Learning richiede generalmente risorse significativamente maggiori rispetto al Machine Learning tradizionale, a causa della complessità e della dimensione delle reti neurali impiegate. Mentre un modello di Machine Learning può essere addestrato efficacemente su un computer standard, i modelli di Deep Learning spesso necessitano di hardware specializzato come GPU o TPU per ridurre i tempi di addestramento Dean et al. (2012).

Per quanto riguarda le applicazioni pratiche, il ML, con la sua capacità di apprendere da set di dati strutturati e di fare previsioni o classificazioni, ha aperto la strada alla creazione di sistemi AI che possono generare contenuti basandosi su pattern appresi, Jordan & Mitchell (2015). Ad esempio, i sistemi di raccomandazione utilizzati da piattaforme di streaming e e-commerce si basano su algoritmi di ML per analizzare le preferenze degli utenti e generare raccomandazioni personalizzate. Il Deep Learning, dal canto suo, ha spinto ancora più in là i confini, permettendo lo sviluppo di AI generative che possono creare immagini, musica, testi e perfino video che sono quasi indistinguibili da quelli creati da esseri umani. Questo è possibile grazie alle reti neurali profonde, potendo elaborare e imitare le complessità delle varie espressioni umane e artistiche. Alcuni esempi significativi di tali sistemi comprendono i Generative Adversarial Networks (GANs) e i Transformer, che hanno mostrato risultati sorprendenti nel campo della generazione di immagini e testi, Vaswani et al. (2017).

Questa distinzione tra Machine Learning e Deep Learning è fondamentale per comprendere quale tecnologia sia più appropriata a seconda delle esigenze specifiche, sia in termini di risorse disponibili sia di obiettivi dell'analisi. Il Machine Learning e il Deep Learning non solo hanno rivoluzionato l'approccio all'analisi dei dati e alla risoluzione dei problemi, ma hanno soprattutto posto le basi per lo sviluppo di Intelligenze Artificiali Generative (GAI). Queste ultime rappresentano l'ennesimo passo nell'evoluzione dell'Intelligenza Artificiale, come verrà esplorato nel prossimo sottocapitolo, e sono fortemente influenzate dalle capacità e dagli approcci del ML e del DL.

In conclusione, ML e DL non sono solo strumenti potenti per l'analisi e la comprensione dei dati, ma fungono anche da catalizzatori per l'avanzamento delle AI generative, ampliando le possibilità di ciò che le AI possono creare e come possono interagire con il mondo reale.

1.3. Che cosa sono le AI Generative

1.3.1. *Definizioni Base*

Le AI Generative, come descritto da Goodfellow et al. (2014), sono avanzati sistemi di intelligenza artificiale distinti per la loro capacità di creare nuovi dati e contenuti, imitando quelli reali. Questi sistemi si basano su algoritmi di apprendimento automatico per analizzare e comprendere grandi set di dati esistenti, e successivamente utilizzano queste informazioni per creare contenuti nuovi e unici.

Le caratteristiche chiave delle Generative AI includono:

- **Apprendimento Automatico (Machine Learning):** Queste AI utilizzano algoritmi di apprendimento automatico per analizzare e imparare da grandi quantità di dati. Questo apprendimento è spesso non supervisionato o semi-supervisionato. (Anantrasirichai & Bull, 2020); (Goodfellow et al., 2020).
- **Capacità Generativa:** A differenza delle AI tradizionali che si limitano ad analizzare e categorizzare i dati, le Generative AI possono creare nuovi contenuti. Questo include testi, immagini, musica, e persino codice software. (Caporusso et al., 2019).

- **Adattabilità e Personalizzazione:** Sono in grado di adattarsi a vari contesti e possono essere personalizzate per esigenze specifiche, rendendole utili in diversi settori come l'arte, la scienza, la medicina e l'intrattenimento. (Struik & Kondratenko, 2021).
- **Interattività e Risposta in Tempo Reale:** Queste AI possono interagire con gli utenti in tempo reale, offrendo risposte e contenuti generati in base alle richieste specifiche degli utenti. (Hong et al., 2020).

Per capire meglio questo concetto della generazione dal nulla, sviluppando una metafora, immaginate le Intelligenze Artificiali Generative (Generative AI) come dei "maestri alchimisti" del mondo digitale. Queste AI non si limitano a comprendere o analizzare i dati, ma li trasformano in qualcosa di completamente nuovo e originale partendo da materiali esistenti, trasformandoli in risultati completamente nuovi e inediti. Il processo di generazione inizia sempre con l'apprendimento: queste AI, come studenti attenti, osservano e studiano enormi quantità di dati esistenti, imparando da essi schemi, stili e strutture.

Ad esempio, consideriamo una Generative AI focalizzata sulla generazione di immagini. Inizialmente, la rete neurale della AI viene "addestrata" esponendola a migliaia, o anche milioni, di immagini. Durante questa fase di addestramento, la AI analizza questi dati, apprendendo dettagli come forme, colori e composizioni.

Una volta completata la fase di addestramento, la Generative AI può iniziare il suo processo di "creazione". Utilizzando le conoscenze acquisite, inizia a generare nuove immagini. Queste immagini non sono copie di quelle che ha visto durante l'addestramento, ma creazioni originali, basate sui pattern appresi. Il risultato è un'opera d'arte digitale unica, un prodotto dell'apprendimento della AI e della sua "creatività" algoritmica.

Un esempio particolarmente noto di questo processo è quello delle Generative Adversarial Networks (GANs). Le GANs utilizzano due reti neurali in competizione: una generativa, che crea immagini, e una discriminativa, che giudica le immagini. La rete discriminativa valuta se l'immagine generata è "reale" o no, e questa valutazione aiuta la rete generativa a perfezionare ulteriormente le sue creazioni, Costa et al. (2020)

In sintesi, le Generative AI, attraverso processi di apprendimento profondo e meccanismi come le GANs, hanno aperto nuovi orizzonti nella generazione digitale, permettendo la creazione di contenuti che prima potevamo solo immaginare. Queste tecnologie rappresentano non solo un incredibile progresso tecnico, ma anche un nuovo campo di espressione artistica e creativa nel mondo digitale. Questa progressione apre nuovi orizzonti nell'applicazione dell'AI e pone questioni fondamentali sull'interazione futuro tra umanità e macchine intelligenti.

1.3.2. Differenze dalle AI “Standard”

Il confronto tra le Generative AI e le AI tradizionali rivela differenze fondamentali nella loro natura e nelle loro capacità operative. Il campo delle Generative AI si distingue nettamente dalle AI tradizionali, entrambe radicate nella scienza dell'intelligenza artificiale, ma con approcci e obiettivi diversi.

Le AI tradizionali, come descritto da Russell e Norvig (2016), sono principalmente incentrate sull'analisi, interpretazione e reazione ai dati, con applicazioni che spaziano dal riconoscimento di pattern alla classificazione e all'elaborazione del linguaggio naturale. Questi sistemi si avvalgono di algoritmi di machine learning come l'apprendimento supervisionato e non supervisionato, per estrarre informazioni utili dai dati, ma hanno limiti nella creazione di nuovi contenuti.

Al contrario, le Generative AI, esemplificate dai lavori su GANs di Goodfellow et al. (2014), trascendono la mera analisi dei dati esistenti. Questi sistemi innovativi sono progettati non solo per comprendere i dati, ma per generare nuovi dati e contenuti, spesso indistinguibili da quelli reali. La loro capacità di generazione, che va oltre la replica dei dati esistenti, è una forma di creazione basata sull'apprendimento dai dati. Pertanto, le Generative AI contribuiscono attivamente alla creazione di nuove realtà digitali, trovando applicazioni in ambiti come l'arte digitale e la simulazione.

1.4. Tecnologie Chiave

Nel mondo delle Generative AI, diverse tecnologie chiave emergono come pilastri fondamentali che supportano e guidano il loro sviluppo. Le tecnologie chiave delle

Generative AI, ossia le Generative Adversarial Networks (GANs), i Large Language Models (LLM) come GPT-3 e i Variational Auto Encoders (VAEs), sono ritenute dagli addetti ai lavori come fondamentali per il loro impatto rivoluzionario e la loro capacità di spingere i confini dell'intelligenza artificiale. Le GANs hanno trasformato la generazione di immagini, introducendo un nuovo livello di realismo. I LLM hanno ridefinito il processamento del linguaggio, rendendo possibile la generazione di testo di alta qualità indistinguibile da quello umano. I VAEs hanno aperto nuove possibilità nella modellazione di dati complessi.

Insieme, queste tecnologie hanno esteso significativamente il campo di applicazione delle Generative AI, dimostrando un'ampia gamma di applicazioni pratiche e di ricerca, non solo permettendo la creazione di contenuti nuovi e innovativi, ma anche aprendo e continuando ad aprire nuove strade nel campo dell'AI..

1.4.1. Le Generative Adversarial Network (GANs)

Le Generative Adversarial Networks (GANs), una rivoluzionaria tecnologia nel campo delle Generative AI, sono state introdotte da Ian Goodfellow e colleghi nel loro paper fondamentale del 2014, "Generative Adversarial Nets". Questa tecnologia funziona tramite l'interazione di due reti neurali: una rete generativa, che produce immagini, e una discriminativa, che le valuta. Questo meccanismo di competizione porta alla generazione di immagini di alta qualità e realistiche.

Il funzionamento delle GANs si basa su un processo iterativo di addestramento, dove la rete generativa cerca di creare immagini sempre più convincenti, mentre quella discriminativa impara a distinguere meglio tra immagini reali e generate.

Volendo esprimere il processo iterativo delle GANs tramite una metafora, è quindi simile a una gara tra due artisti, dove ogni artista impara dall'altro. La rete generativa, come un pittore, si sforza di creare immagini sempre più realistiche. Contemporaneamente, la rete discriminativa agisce come un critico d'arte, affinando la sua abilità nel distinguere tra le creazioni del generatore e immagini autentiche. Con ogni iterazione, il generatore diventa più abile nel creare immagini realistiche, mentre il discriminatore diventa più esperto nel rilevare le sottili differenze tra le immagini reali e quelle generate. Questo ciclo di feedback porta a miglioramenti continui in entrambe le reti, migliorando la qualità delle immagini generate e raffinando la complessità dei modelli generativi.

Le GANs hanno trovato applicazioni rivoluzionarie in numerosi settori. Nell'arte digitale, ad esempio, sono state utilizzate per creare nuove forme di espressione artistica, come dimostrato da progetti che combinano stili pittorici classici e moderni per generare opere d'arte uniche. In medicina, le GANs hanno un ruolo cruciale nella generazione di immagini mediche realistiche per la formazione e la ricerca, come illustrato da Nie et al. (2017) nel loro lavoro sulla generazione di immagini mediche sintetiche.

In altri campi, come la moda e il design, le GANs contribuiscono a creare prototipi di design innovativi. Nel settore dei videogiochi e dell'intrattenimento, vengono utilizzate per generare ambienti e personaggi realistici, migliorando l'esperienza di gioco. Inoltre, nel campo dell'apprendimento automatico, le GANs sono impiegate per migliorare l'accuratezza dei modelli attraverso la generazione di dati di addestramento supplementari.

Nonostante l'efficacia delle Generative Adversarial Networks, queste reti affrontano alcune sfide tecniche significative. Una delle principali è la difficoltà nel mantenere la stabilità durante l'addestramento, un problema noto come "mode collapse", dove il generatore produce un numero limitato di output, indipendentemente dall'input. Questo aspetto è stato ampiamente discusso in vari studi, tra cui quello di Salimans et al. (2016), che esplorano metodi per mitigare questi problemi.

Inoltre, c'è il rischio che le GANs producano risultati non realistici o distorti, un problema evidenziato in numerosi contesti, tra cui la generazione di immagini facciali. Nonostante queste sfide, le GANs restano una tecnologia fondamentale nelle Generative AI, fornendo una piattaforma per innovazioni future e nuove applicazioni in diversi settori.

1.4.2. I Large Language Models (LLMs)

Large Language Models (LLM) come GPT: Sviluppato da OpenAI, GPT ha segnato una svolta significativa nella generazione di linguaggio naturale. Presentato per la prima volta nello studio di Tom Brown et al (2020), GPT-3 ha dimostrato la capacità di generare testo coerente e contestualmente rilevante su una vasta gamma di argomenti, rivoluzionando il campo del NLP.

I Large Language Models (LLM) operano attraverso una sofisticata rete neurale basata su un'architettura chiamata "transformer". Questa architettura permette ai modelli di processare e generare linguaggio naturale analizzando e comprendendo il contesto di

enormi quantità di dati testuali. Il loro funzionamento si basa sull'apprendimento da un vasto corpus di testo, durante il quale il modello impara pattern, strutture linguistiche, e varie sfumature del linguaggio. Utilizzando tecniche come l'apprendimento per rinforzo e il transfer learning, i LLM sono in grado di generare testi coerenti e contestualmente pertinenti, spaziando da risposte conversazionali a contenuti creativi complessi.

Volendo spiegare in maniera più semplice, il funzionamento dei LLM come GPT può essere paragonato a quello di un bibliotecario estremamente erudito, che ha letto e memorizzato il contenuto di milioni di libri. Quando gli viene posta una domanda, questo "bibliotecario digitale" non cerca una risposta esatta come farebbe un archivio, ma costruisce una risposta basandosi su tutto ciò che ha "letto". È come se creasse una nuova pagina di un libro, combinando tutte le sue letture per produrre qualcosa di informativo e rilevante per la domanda posta.

Per capire come vengono scelte le informazioni da trasformare in output di testo, è necessario comprendere che i LLM operano secondo principi di determinazione probabilistica. Quando generano testo, stanno essenzialmente scegliendo ogni parola successiva basandosi sulla probabilità: "quale parola o frase ha più senso dopo quelle già scritte?" Il risultato è un testo che rappresenta la successione più probabile di parole, basata su l'immensa quantità di dati linguistici che il modello ha analizzato. Questo approccio probabilistico significa che mentre i risultati sono spesso sorprendentemente coerenti e pertinenti, sono basati su probabilità anziché su una comprensione umana del contesto o dell'intento.

1.4.3. Variational Auto Encoders (VAEs)

Variational Autoencoders (VAEs): I VAEs sono una tecnologia fondamentale nel campo delle Generative AI, introdotta da Kingma e Welling (2013). I VAEs sono reti neurali utilizzate per la generazione e la modellazione di dati, in particolare per immagini e altri tipi di dati complessi.

Tecnicamente, i VAEs funzionano codificando i dati in una rappresentazione compressa, una sorta di "essenza" dei dati. Questa fase impiega un approccio probabilistico, che introduce una variazione casuale per garantire che la rappresentazione compressa possa essere utilizzata per generare nuove istanze di dati simili, ma non identiche, a quelle

originali. Durante la decodifica, i VAEs utilizzano questa rappresentazione per ricostruire i dati, creando nuove istanze che mantengono le proprietà fondamentali del set originale.

Riprendendo la metafora sul pittore, il funzionamento dei VAEs può essere paragonato al lavoro di un pittore che crea un ritratto. Inizialmente, l'artista studia il soggetto (codifica) e crea uno schizzo compresso che cattura le caratteristiche essenziali. Successivamente, utilizza questo schizzo per ricreare il ritratto, aggiungendo dettagli per renderlo il più fedele possibile all'originale (decodifica).

I Variational Auto Encoders (VAEs) hanno trovato applicazioni in una serie di ambiti grazie alla loro capacità di generare dati che rispecchiano fedelmente la distribuzione del set di addestramento. Nel campo della generazione di immagini, i VAEs sono impiegati per creare variazioni di volti umani o paesaggi, fornendo una vasta gamma di immagini realistiche per la formazione di modelli di intelligenza artificiale. Un esempio notevole è la loro applicazione nella generazione di immagini mediche sintetiche, che possono essere utilizzate per la formazione dei medici senza ricorrere a immagini di pazienti reali. Inoltre, i VAEs sono stati utilizzati anche nella modellazione di dati complessi in altri settori, come la finanza e la meteorologia, per simulare scenari e prevedere tendenze.

Nonostante i VAEs siano strumenti potenti, affrontano alcune sfide tecniche, tra cui la tendenza a produrre immagini con una qualità inferiore, spesso percepite come "sfocate". Questo problema è stato ampiamente studiato e documentato in vari lavori accademici, come quelli di Kingma e Welling, i quali hanno esplorato approfonditamente le capacità e le limitazioni dei VAEs. Tuttavia, rimangono una tecnologia fondamentale per le Generative AI, grazie alla loro versatilità nell'applicazione, dalla generazione di immagini alla modellazione di dati complessi. Queste tecnologie sono la spina dorsale delle applicazioni più innovative delle Generative AI, definendo il campo come un'area di ricerca e sviluppo dinamica e in rapida evoluzione.

2. IL RUOLO DEI MODELLI DI LLM E DI GPT NELLO SVILUPPO DELLE INTELLIGENZE ARTIFICIALI GENERATIVE

2.1. Evoluzione dei Transformer e dei Large Language Models

Nei capitoli precedenti, abbiamo esplorato le basi delle AI e della sua evoluzione e il mondo delle AI Generative come GPT, nello specifico un Large Language Model (LLM), che hanno ridefinito le possibilità dell'Intelligenza Artificiale nel campo del linguaggio naturale. Tuttavia, per comprendere appieno il salto evolutivo rappresentato da questi modelli, è essenziale rivolgere lo sguardo all'architettura sottostante che ne ha permesso la realizzazione, ovvero i Transformer. Questi modelli di deep learning hanno non solo semplificato ma anche potenziato la capacità dei computer di comprendere e generare il linguaggio naturale, fungendo da pilastri per i moderni LLM.

2.1.1. Architettura e Funzionalità dei Transformer

Un transformer è un modello di deep learning innovativo che ha apportato cambiamenti significativi non solo nel campo del Natural Language Processing (NLP) ma anche in altre aree dell'intelligenza artificiale. Questi modelli, distinti per la loro capacità di elaborare sequenze di dati in modo estremamente efficiente e sofisticato, rappresentano un salto qualitativo rispetto ai modelli precedenti. Introdotti da Vaswani et al. nel 2017, i transformer hanno rappresentato una svolta nell'Intelligenza Artificiale, in particolare nel Natural Language Processing (NLP). Mentre i precedenti modelli come le reti neurali ricorrenti (RNN) e le Long Short-Term Memory (LSTM) hanno gettato le basi, i transformer hanno rivoluzionato l'approccio al trattamento del linguaggio naturale grazie alla loro architettura innovativa. Il loro meccanismo di 'attenzione', che consente di focalizzarsi in modo dinamico sulle parti rilevanti di un input, ha permesso una comprensione e una generazione del testo più contestualmente pertinente e sofisticata rispetto ai modelli precedenti.

Il meccanismo di attenzione, situato al cuore dell'architettura dei transformer, fu introdotto da Vaswani et al. (2017). Questa caratteristica consente al modello di

focalizzarsi su diverse parti di un input, come una frase o un passaggio di testo, per calcolare una rappresentazione del contesto più ricca e sfaccettata.

Questa funzionalità di attenzione si differenzia dalle precedenti architetture, come le reti RNN o le LSTM, poiché non tratta l'input sequenzialmente. Invece, il modello analizza l'intera sequenza in una volta, permettendo di ponderare dinamicamente l'importanza relativa di ciascuna parte dell'input. Tale approccio consente ai transformer di catturare relazioni complesse e a lungo raggio all'interno del testo, un aspetto fondamentale per la comprensione accurata del linguaggio naturale.

Nel contesto del NLP, questa capacità si traduce in una comprensione del contesto notevolmente migliorata. Per esempio, in una frase, il modello può assegnare più peso a determinate parole o frasi, a seconda di quanto siano rilevanti per il significato complessivo del testo. Questo permette ai transformer di generare risposte o traduzioni che non solo sono linguisticamente corrette, ma anche contestualmente appropriate. Tale caratteristica è stata fondamentale per il successo di modelli avanzati basati su transformer, come BERT di Devlin et al. (2019), che utilizza un approccio bidirezionale per comprendere il contesto di ogni parola all'interno di una frase.

Un altro aspetto cruciale del meccanismo di attenzione è la sua efficienza computazionale, ovvero la capacità di processare elementi di una sequenza in parallelo, piuttosto che in serie. Questa capacità ha portato a una riduzione significativa dei tempi di addestramento e ha reso possibile l'elaborazione di set di dati molto più grandi. Questa capacità di processare simultaneamente tutti gli elementi di una sequenza di dati rappresenta un avanzamento tecnologico fondamentale, Vaswani et al. (2017).

A differenza delle RNN e delle LSTM, che elaborano i dati in modo sequenziale e sono quindi limitate dalla dipendenza temporale tra i dati, i transformer eliminano questa restrizione attraverso l'elaborazione parallela. Questa caratteristica consente ai transformer di valutare l'intera sequenza di input in un unico passaggio, piuttosto che analizzare ciascun elemento uno alla volta. Tale approccio non solo accelera il processo di apprendimento, ma consente anche ai modelli di catturare relazioni complesse e a lungo raggio all'interno dei dati. L'efficienza computazionale fornita dall'elaborazione parallela è particolarmente vantaggiosa quando si gestiscono grandi set di dati, come nei LLM, Vaswani et al. (2017). Questo aspetto è stato un fattore determinante nel migliorare la scalabilità dei modelli di NLP e nel ridurre significativamente i tempi di addestramento.

La capacità di processare simultaneamente grandi quantità di dati rende i transformer ideali per l'uso con unità di elaborazione grafica (GPU), che sono ottimizzate per il calcolo parallelo. Questo ha contribuito in modo notevole all'accelerazione dello sviluppo e all'ampia adozione dei transformer nel campo dell'intelligenza artificiale.

Infine l'ultimo elemento di differenza è l'architettura dei Transformer: la struttura a blocchi. La struttura architettonica dei transformer è caratterizzata da una serie di blocchi modulari, ognuno dei quali svolge un ruolo cruciale nell'elaborazione dell'input. Come delineato da Vaswani et al. (2017), ciascun blocco di questa architettura è composto da un livello di attenzione, seguito da una rete neurale feed-forward. Questa disposizione strutturale non solo migliora l'efficienza complessiva del modello, ma offre anche un livello di flessibilità senza precedenti nella progettazione e nell'applicazione dei modelli di deep learning.

La composizione modulare dei transformer si basa su una serie di blocchi ripetuti, ognuno dei quali svolge un ruolo specifico e cruciale nell'elaborazione dell'input. All'interno di ciascun blocco, troviamo innanzitutto un livello di attenzione, il quale consente al modello di analizzare l'intera sequenza di input, attribuendo un peso relativo alle diverse parti di tale sequenza. Questo meccanismo di attenzione è fondamentale per una comprensione più profonda e dettagliata del contesto e delle relazioni intrinseche nel testo, superando i limiti delle architetture precedenti che lavoravano i dati in maniera sequenziale.

Segue poi, all'interno di ogni blocco, una rete neurale feed-forward. Questa rete, che agisce sull'output del livello di attenzione, è progettata per elaborare ulteriormente e raffinare la rappresentazione ottenuta. Il lavoro di Vaswani et al. (2017) ha messo in luce come queste reti contribuiscano a migliorare la comprensione e la generazione del linguaggio, offrendo una trasformazione più sofisticata dell'input ricevuto.

Un aspetto rilevante di questa architettura a blocchi è la sua natura modulare e ripetitiva, che offre una flessibilità senza precedenti nella progettazione dei modelli. La possibilità di impilare o configurare diversamente questi blocchi consente agli sviluppatori e ai ricercatori di adattare l'architettura ai requisiti specifici di un'ampia gamma di compiti di NLP. Questa versatilità si è dimostrata particolarmente utile in diverse applicazioni, dalla traduzione automatica alla generazione di testo, confermando la potenza e la flessibilità dei transformer.

2.1.2. Importanza dei Transformer nello sviluppo dei LLM

La rivoluzione apportata dai transformer nel trattamento del linguaggio naturale ha segnato un'avanzata significativa per i Large Language Models (LLM). Questo progresso è innanzitutto attribuibile al meccanismo di attenzione, una caratteristica centrale dell'architettura dei transformer che ha permesso una comprensione e un'analisi del testo senza precedenti in termini di profondità e contesto. Prima dell'avvento dei transformer, i modelli basati su reti neurali ricorrenti (RNN) e Long Short-Term Memory (LSTM) dominavano il campo del NLP. Tuttavia, questi modelli presentavano limitazioni significative, specialmente nella gestione di sequenze lunghe e nella cattura di dipendenze contestuali complesse.

L'introduzione dei transformer, come illustrato da Vaswani et al. (2017), ha apportato una svolta decisiva. Il loro meccanismo di attenzione permette ai modelli di focalizzarsi su parti specifiche del testo, ponderando l'importanza relativa di ciascuna parola o frase in funzione del contesto generale. Questo approccio ha portato a una comprensione molto più ricca e sfumata del linguaggio, superando i limiti dei modelli precedenti che lavoravano i dati in maniera sequenziale e spesso perdevano informazioni contestuali cruciali.

Nel contesto degli LLM come GPT, l'impatto di questa innovazione è stato particolarmente evidente. Brown et al. (2020) hanno osservato come GPT-3, grazie all'architettura basata su transformer, sia stato in grado di eseguire una varietà di compiti di NLP con una sorprendente autonomia e precisione. A differenza dei modelli tradizionali, GPT-3 non richiedette un'ampia supervisione o specifiche pre-programmazioni per eseguire compiti complessi. Il modello fu in grado di generare testo, rispondere a domande, e persino creare contenuti creativi con un livello di competenza che prima era irraggiungibile per i modelli di AI.

Questa capacità avanzata di analisi e generazione del linguaggio ha reso i LLM basati su transformer strumenti estremamente potenti e versatili. La loro applicabilità si estende oltre i tradizionali compiti di NLP, influenzando molteplici ambiti aziendali. La versatilità e l'efficacia dei LLM basati su transformer hanno aperto nuove prospettive nel campo dell'AI, spingendo i confini di ciò che è possibile realizzare attraverso l'elaborazione del linguaggio naturale.

Uno degli aspetti più rilevanti che i transformer hanno apportato allo sviluppo dei Large Language Models (LLM) è la scalabilità. La natura dell'architettura dei transformer ha consentito un significativo aumento della dimensione e della complessità dei set di dati utilizzabili per l'addestramento di questi modelli. Questo aumento di scalabilità si è rivelato cruciale per migliorare la capacità di generazione e comprensione del linguaggio dei LLM, Vaswani et al. (2017).

Prima dell'introduzione dei transformer, i limiti computazionali e architetturali delle neurali ricorrenti (RNN) e Long Short-Term Memory (LSTM) impedivano l'addestramento di modelli su set di dati di dimensioni estremamente ampie. Con l'avvento dei transformer, questa barriera è stata superata. La loro capacità di elaborare simultaneamente grandi blocchi di dati ha permesso di utilizzare corpus di testo di dimensioni senza precedenti. Brown et al. (2020) nel loro studio su GPT-3 hanno dimostrato come l'addestramento su un vasto set di dati abbia migliorato notevolmente la capacità del modello di comprendere e generare linguaggio in modo coerente e contestualmente pertinente.

Questa scalabilità ha influenzato non solo la quantità di dati che possono essere trattati, ma anche la qualità delle prestazioni dei LLM. I modelli possono ora essere addestrati su una gamma di testi molto più ampia, che include una varietà di stili, argomenti e strutture linguistiche. Questa diversità di addestramento ha permesso ai LLM di acquisire una comprensione del linguaggio più ricca e sfumata, migliorando la loro capacità di rispondere in modo appropriato a una gamma più ampia di richieste.

Inoltre, la scalabilità offerta dai transformer ha reso possibile l'esplorazione di nuove frontiere nel campo dell'AI. I ricercatori possono ora sperimentare con LLM di dimensioni e complessità maggiori, spingendo i confini di ciò che è possibile realizzare nel trattamento del linguaggio naturale. Questo progresso ha aperto la strada a nuove applicazioni e ha migliorato la capacità dei modelli di affrontare compiti di NLP sempre più sfidanti.

In conclusione, la scalabilità introdotta dall'architettura dei transformer ha avuto un impatto decisivo sullo sviluppo dei LLM. Questa caratteristica ha non solo permesso di superare le precedenti limitazioni computazionali, ma ha anche migliorato in modo significativo la qualità e la versatilità dei modelli di linguaggio, portando a un'era di progressi senza precedenti nel campo dell'intelligenza artificiale.

2.1.3. I Large Language Models (LLM) e il Loro Ruolo nei Motori di NLP

I Large Language Models (LLM) hanno assunto un ruolo centrale nel funzionamento di motori avanzati di elaborazione del linguaggio, come la serie Generative Pre-trained Transformer (GPT) di OpenAI. Questi modelli rappresentano un salto qualitativo nell'ambito dell'intelligenza artificiale, offrendo capacità di comprensione e generazione del linguaggio naturale a livelli senza precedenti.

L'essenza dei LLM nei motori come GPT risiede nella loro straordinaria capacità di generare testo coerente e contestualmente pertinente. A differenza dei precedenti modelli di NLP, i LLM basati su transformer, come GPT, sono addestrati su set di dati di vasta portata, acquisendo così una comprensione del linguaggio umano incredibilmente ampia. Brown et al. (2020) hanno esplorato in profondità il modello GPT-3, mettendo in luce come la sua architettura avanzata permetta al modello di eseguire una varietà di compiti di NLP, dal completamento di testo alla risposta a domande, con una ridotta necessità di supervisione specifica per ogni compito.

Il Natural Language Processing (NLP) è un campo di ricerca multidisciplinare che si trova all'incrocio tra informatica, intelligenza artificiale e linguistica. Si occupa dello sviluppo di algoritmi e sistemi che permettono ai computer di comprendere, interpretare e manipolare il linguaggio umano in modo significativo. Il NLP mira a colmare il divario tra la comunicazione umana e la comprensione computazionale.

Jurafsky e Martin (2009) forniscono una panoramica completa del campo, definendo il NLP come l'applicazione di algoritmi computazionali al linguaggio sia scritto che parlato. La disciplina si estende dalla comprensione del significato semantico e del contesto delle parole alla capacità di estrarre e utilizzare la conoscenza dai documenti di testo.

Negli ultimi 5 anni, il campo del NLP ha compiuto progressi significativi grazie allo sviluppo di modelli di deep learning. Il NLP è utilizzato in una vasta gamma di applicazioni, dalla traduzione automatica alla generazione di testo, dall'analisi dei sentimenti al riconoscimento vocale. L'obiettivo è quello di rendere le interazioni tra uomo e macchina il più naturali e intuitive possibile, migliorando la capacità dei computer di comprendere e rispondere in maniera appropriata al linguaggio umano.

Un altro aspetto significativo dei LLM nei motori come GPT è la loro capacità di apprendimento few-shot o zero-shot. Questo significa che, una volta addestrati, i LLM possono comprendere e rispondere a richieste o compiti per i quali non sono stati specificamente addestrati. Tale flessibilità è rivoluzionaria, poiché consente ai modelli di adattarsi a nuove sfide e contesti con minimo o nessun intervento aggiuntivo.

Inoltre, i LLM come GPT hanno cambiato il paradigma dell'interazione uomo-macchina. Con la loro capacità di generare testo che imita strettamente lo stile e la coerenza umana, questi modelli hanno aperto la strada a nuove applicazioni in ambiti aziendali. Questo ha notevolmente ampliato il campo di applicazione dell'intelligenza artificiale, rendendo possibile un'interazione più naturale e intuitiva tra gli utenti e i sistemi basati su AI, Wen et al. (2023)..

In sintesi, i LLM hanno svolto un ruolo fondamentale nel definire le capacità dei motori avanzati come GPT. La loro evoluzione ha portato a un netto miglioramento nelle capacità di elaborazione del linguaggio naturale, permettendo una generazione di testo più sofisticata e un'applicazione più ampia nel campo dell'AI. Con i loro progressi continui, i LLM stanno plasmando il futuro dell'intelligenza artificiale, dimostrando un potenziale ancora inesplorato nel campo dell'elaborazione del linguaggio naturale, Web et al. (2023).

2.2. Generative Pre-Trained Transformer (GPT) e Chat GPT

2.2.1. Che cos'è un GPT

La serie Generative Pre-trained Transformer (GPT) rappresenta una pietra miliare nel campo dell'Intelligenza Artificiale e del Natural Language Processing (NLP). Sviluppata da OpenAI, la serie GPT è una famiglia di modelli di deep learning che hanno rivoluzionato il modo in cui i computer comprendono il linguaggio naturale grazie alla sua architettura avanzata e alle sue capacità di generazione di testo.

Il nucleo dell'architettura GPT è basato sull'architettura dei transformer, introdotta da Vaswani et al. (2017). Il design innovativo dei transformer permette a GPT di processare e generare testo con un livello di coerenza e contestualizzazione notevolmente superiore

rispetto ai precedenti modelli di NLP basati su reti neurali ricorrenti (RNN) o Long Short-Term Memory (LSTM).

I modelli GPT sono pre-addestrati su vasti dataset di testo, permettendo loro di apprendere una ricca comprensione del linguaggio naturale prima di essere impiegati in compiti specifici. Questo pre-addestramento, come esemplificato nel lavoro di Radford et al. (2018), consente a GPT di generare testo che è non solo grammaticalmente corretto, ma anche ricco dal punto di vista semantico e contestuale. La capacità di generazione di testo di GPT si estende a una varietà di applicazioni, dalla creazione di contenuti alla risposta automatica a domande, evidenziando la sua versatilità come strumento di NLP.

Una delle caratteristiche distintive di GPT è la sua capacità di apprendimento non supervisionato, che lo rende eccezionalmente potente in termini di flessibilità e adattabilità. A differenza dei modelli tradizionali che richiedono un addestramento specifico per ciascun compito, GPT può essere utilizzato in una varietà di contesti con minimo aggiustamento. Questa versatilità è stata ulteriormente sviluppata in GPT-3, come descritto da Brown et al. (2020), il quale ha dimostrato una notevole capacità di apprendimento few-shot e zero-shot, eseguendo compiti per i quali non è stato specificamente addestrato.

Ogni iterazione della serie GPT ha costruito sulle fondamenta del suo predecessore, portando a miglioramenti sostanziali in termini di dimensioni, complessità e capacità.

- **GPT-1:** L'Inizio dell'Era dei Transformer nel NLP. Introdotta da Radford et al. (2018), GPT-1 ha segnato l'inizio dell'era dei transformer nel NLP. Con 117 milioni di parametri, GPT-1 ha superato i modelli precedenti basati su RNN e LSTM, dimostrando una notevole abilità nella generazione di testo e nell'apprendimento non supervisionato. Questo modello ha gettato le basi per l'utilizzo dell'architettura dei transformer in ambito NLP, ponendo le premesse per innovazioni future.
- **GPT-2:** GPT-2, presentato da Radford et al. (2019), ha rappresentato un salto qualitativo significativo. Con 1,5 miliardi di parametri, più di 10 volte maggiore rispetto a GPT-1, GPT-2 ha migliorato notevolmente la generazione di testo rispetto al suo predecessore, offrendo una comprensione del linguaggio più profonda e sfumata. Il modello ha dimostrato un'abilità eccezionale nel

completamento di testi e nella creazione di contenuti convincenti, segnando un passo importante nello sviluppo dei modelli di linguaggio.

- **GPT-3:** L'iterazione GPT-3, descritta da Brown et al. (2020), divenne un punto di riferimento nel campo dell'NLP. Con i suoi 175 miliardi di parametri, GPT-3 ha stabilito un nuovo standard per la generazione di testo e la comprensione del linguaggio. Il modello ha mostrato capacità avanzate di apprendimento few-shot, consentendo una adattabilità notevole a vari compiti di NLP senza necessità di addestramento specifico.
- **GPT-3.5 (Turbo):** Con GPT-3.5, OpenAI ha introdotto miglioramenti incrementali rispetto a GPT-3. Non è stato condiviso il modello GPT-3.5 grezzo ma solamente applicato a funzionalità specifiche come “Turbo”. Questo modello intermedio ha ulteriormente affinato le capacità di generazione puntando su un notevole incremento della velocità di sviluppo di testo. Attualmente (2024) è il modello più veloce di OpenAI nella generazione di testo.
- **GPT-4:** GPT-4 rappresenta l'ultima evoluzione della serie al momento della scrittura (2024), portando le capacità di elaborazione del linguaggio a un livello ancora superiore. Grazie ai suoi 1,76 Triloni di parametri, aumenta di 3 ordini di grandezza la base di Training rispetto ai predecessori GPT-3 e 3.5. Attualmente è il modello più performante conosciuto.

In conclusione, i modelli GPT rappresentano un'innovazione significativa nel campo dell'intelligenza artificiale, segnando un punto di svolta nella capacità dei computer di interagire e comprendere il linguaggio umano. Con il loro continuo sviluppo, i modelli GPT stanno plasmando il futuro del NLP e dell'AI, offrendo nuove prospettive e capacità in questo campo in rapida evoluzione.

2.2.2. Che cos'è ChatGPT

ChatGPT, sviluppato da OpenAI, rappresenta un'evoluzione significativa nel panorama dei modelli basati sulla tecnologia Generative Pre-trained Transformer (GPT). Questo modello si distingue per il suo focus specializzato sull'interazione conversazionale, mirando a simulare una conversazione naturale e umana. La sua implementazione segna un punto di svolta nell'ambito dei chatbot e dei sistemi di assistenza virtuale.

Al cuore di ChatGPT vi è l'architettura transformer, simile a quella utilizzata nei vari GPT-3/3.5/4, ma con una serie di ottimizzazioni mirate a migliorare le interazioni conversazionali. Come sottolineato da Radford et al. (2019), mentre i modelli GPT standard sono stati addestrati principalmente per generare testo basato su prompt iniziali, ChatGPT è stato ulteriormente affinato per rispondere in modo coerente e contestualmente appropriato all'interno di una conversazione. Questa ottimizzazione permette a ChatGPT di mantenere il contesto attraverso scambi di dialogo multi-turn, una caratteristica che lo distingue dai precedenti modelli di chatbot.

Lo scambio di dialogo multi-turn è una caratteristica cruciale nei sistemi di intelligenza artificiale conversazionale. Questo approccio permette ai modelli di mantenere e gestire una conversazione su più turni di dialogo, creando un'interazione più naturale e umana. Un dialogo multi-turn si riferisce a una conversazione in cui l'interazione non si limita a una singola domanda e risposta, ma si sviluppa su più turni di scambio, con ciascuna risposta che tiene conto del contesto precedente. Questo concetto è fondamentale per simulare conversazioni reali, dove gli interlocutori ricordano e fanno riferimento a ciò che è stato detto precedentemente. Il concetto di dialogo multi-turn è stato esplorato in diversi studi accademici nel campo dell'NLP e dell'intelligenza artificiale. Sutskever et al. (2014), in uno studio sulla Sequence to Sequence Learning, hanno gettato le basi per modelli capaci di gestire conversazioni complesse, sebbene i primi approcci fossero ancora limitati nella gestione di contesti conversazionali prolungati.

Nel caso di ChatGPT, il modello utilizza l'architettura dei transformer per gestire dialoghi multi-turn. Questo viene realizzato tramite l'addestramento del modello su set di dati di conversazioni, permettendogli di imparare come le risposte sono correlate l'una all'altra nel corso di uno scambio. ChatGPT, grazie a questa capacità, può generare risposte che non solo sono pertinenti all'ultimo messaggio ricevuto, ma che tengono conto dell'intera storia della conversazione. Una delle sfide principali nell'AI conversazionale è la capacità di mantenere coerenza e contesto lungo una serie di scambi comunicativi. ChatGPT affronta questa sfida utilizzando tecniche avanzate di apprendimento profondo e strategie di trattamento del contesto, come descritto da Brown et al. (2020). Queste tecniche consentono a ChatGPT di comprendere e rispondere a sequenze di domande e commenti, mantenendo un filo logico coerente e pertinente. Questa capacità di contestualizzazione rende ChatGPT un modello particolarmente avanzato per applicazioni che richiedono un

elevato grado di interazione naturale, come il supporto clienti automatizzato e la formazione virtuale.

In sintesi, ChatGPT di OpenAI rappresenta un notevole progresso nel campo dell'intelligenza artificiale conversazionale. La sua architettura ottimizzata per il dialogo e la sua capacità di mantenere contesto e coerenza in una conversazione lo rendono un modello unico e innovativo, con ampie applicazioni pratiche e teoriche nel settore dell'AI.

2.2.3. *Foundation Models, Fine Tuning e Prompt Engineering*

I Foundation Models rappresentano una classe di modelli di apprendimento automatico che hanno rivoluzionato il campo dell'intelligenza artificiale. Sono modelli addestrabili su larga scala e scalabili che forniscono una conoscenza generale estesa, utilizzabile come base per svariate applicazioni. Caratterizzati da un numero elevato di parametri, questi modelli sono addestrati su vasti set di dati, acquisendo così una comprensione profonda e versatile del linguaggio naturale e di altri tipi di dati. I Foundation Models sono addestrati su dataset di dimensioni senza precedenti, che coprono una vasta gamma di stili linguistici, argomenti e contesti. Questo addestramento intensivo permette ai modelli di sviluppare una comprensione profonda e sfumata del linguaggio naturale, una capacità fondamentale per applicazioni come ChatGPT. Come descritto da Brown et al. (2020), la vasta conoscenza linguistica acquisita da questi modelli consente loro di interpretare e rispondere a una varietà di input conversazionali in modo contestualmente appropriato.

L'addestramento su larga scala dei Foundation Models offre una flessibilità senza pari, consentendo a ChatGPT di adattarsi e rispondere efficacemente a una vasta gamma di richieste conversazionali. Questa capacità di adattamento è resa possibile dalla natura dei modelli pre-addestrati, che possono essere affinati per compiti specifici. La ricerca di Radford et al. (2019) ha illustrato come i modelli basati su GPT-2 possano essere personalizzati per vari compiti di NLP. ChatGPT estende questa flessibilità, sfruttando la base di GPT-3 per creare interazioni conversazionali più naturali e coinvolgenti.

GPT-4 di OpenAI è uno degli esempi più emblematici di un Foundation Model. Con i suoi 1.76 trilioni di parametri, GPT-4 ha dimostrato capacità senza precedenti in termini di generazione di testo, comprensione del linguaggio e adattabilità a vari compiti di NLP con pochi o nessun esempio di addestramento OpenAI et al. (2023). Questo modello utilizza l'architettura e le capacità di GPT-4 per ottimizzare le risposte in contesti

dialogici. In conclusione, i Foundation Models sono cruciali per lo sviluppo di ChatGPT, fornendo la base di conoscenza e la comprensione contestuale necessarie per interazioni conversazionali avanzate.

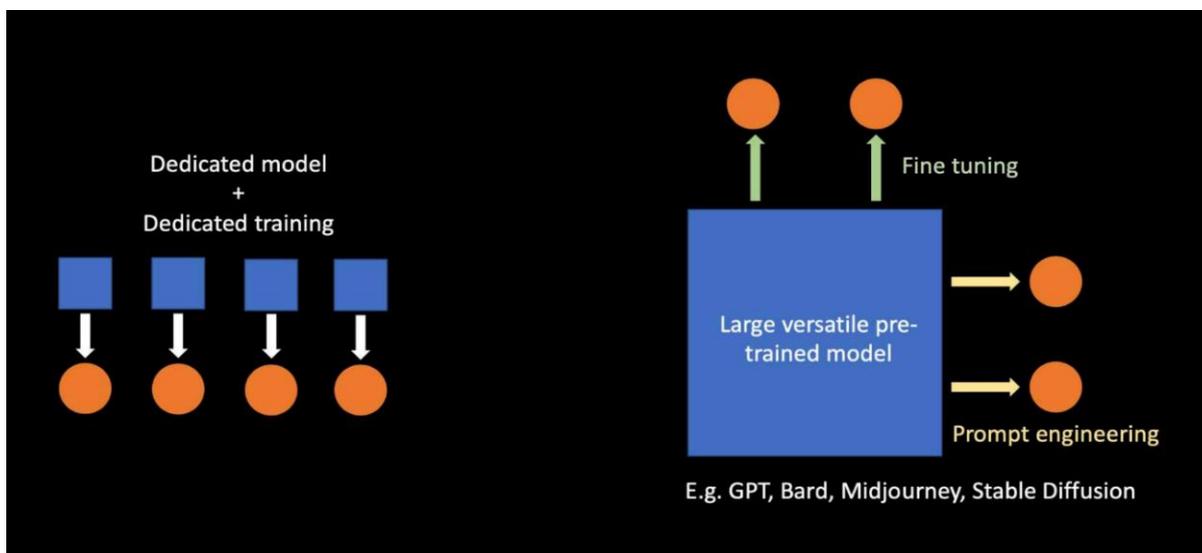


Figura 1: Confronto concettuale tra due approcci distinti nella creazione di modelli di intelligenza artificiale: i "Foundation Models" e i "Dedicated Models", AI Department, UNOX.

Per capire la potenzialità dei Foundation Models e la differenza dai modelli "Standard", detti Dedicated, facciamo riferimento alla figura 1. A sinistra, i "Dedicated Models" sono rappresentati da una serie di blocchi con frecce dirette verso particolari punti. Questo simboleggia modelli di AI specifici con un addestramento dedicato per ciascun compito. Ogni modello è progettato, addestrato e ottimizzato per eseguire un unico compito o un set ristretto di compiti, implicando una specializzazione stretta e una mancanza di flessibilità. A destra, il "Foundation Model" è rappresentato da un grande blocco blu da cui emergono due frecce, una che indica il "Fine tuning" e l'altra "Prompt engineering". Questo illustra un singolo modello di AI di grandi dimensioni e versatile che, una volta pre-addestrato su una vasta gamma di dati, può essere adattato a una varietà di compiti. Il "Fine tuning" si riferisce al processo di addestramento successivo su un set di dati più specifico per migliorare la performance del modello in compiti particolari. "Prompt engineering" indica l'uso di input attentamente progettati per guidare il modello verso l'esecuzione di compiti specifici senza la necessità di addestramento supplementare.

Nell'approccio tradizionale dei modelli dedicati, ogni sistema di intelligenza artificiale è costruito e perfezionato per un'attività specifica. Questi modelli richiedono set di addestramento individuali e sono ottimizzati per massimizzare le prestazioni in quel

singolo compito. Il loro addestramento è intensamente focalizzato, e di conseguenza, il loro ambito di applicazione è stretto e delimitato.

Al contrario, i "Foundation Models" incarnano un paradigma di ampio respiro. Essi sono modelli pre-addestrati di vasta scala, formati su un corpus eclettico di dati che abbraccia una molteplicità di domini e compiti. Il loro addestramento generale li rende straordinariamente versatili, capaci di adattarsi a una gamma di applicazioni successive attraverso tecniche come il "fine tuning" e "Prompt Engineering". La differenza sostanziale tra i due approcci si riflette nella loro capacità di generalizzazione e specializzazione. I modelli dedicati, sebbene altamente competenti nel loro ambito specifico, mancano della flessibilità per adattarsi a compiti al di fuori di quello per cui sono stati espressamente addestrati. I "Foundation Models", invece, grazie alla loro vasta conoscenza pre-acquisita e alle capacità di generalizzazione, possono essere rapidamente adattati a nuovi compiti, dimostrando una malleabilità e un'efficienza notevolmente superiori. In definitiva, la scelta tra un modello dedicato e un "Foundation Model" si basa su una valutazione tra la necessità di specializzazione intensa e la preferenza per l'adattabilità e la versatilità. Nel contesto del 2024, si osserva una tendenza crescente verso l'adozione di "Foundation Models", come indicato dagli esempi di GPT e altri modelli analoghi, che rappresentano una direzione strategica nell'evoluzione dell'intelligenza artificiale.

Il "fine tuning" è una fase cruciale nello sviluppo di modelli di intelligenza artificiale ad alte prestazioni, in particolare quando si tratta di "Foundation Models". Questo processo implica l'adattamento di un modello pre-addestrato su un vasto set di dati a compiti specifici attraverso un addestramento supplementare su un set di dati più ristretto e focalizzato. Il fine tuning affina le conoscenze generali acquisite durante il pre-addestramento per migliorare la performance del modello su compiti specializzati. In pratica, il modello è "ritoccato" con dati specifici del dominio, che possono includere terminologia tecnica, stili di dialogo particolari o tipologie di testo uniche. Questo processo permette al modello di adattarsi e di eseguire con maggiore precisione i compiti per i quali il pre-addestramento generale non sarebbe sufficientemente specifico. L'importanza del fine tuning è stata ampiamente documentata in letteratura. Howard e Ruder (2018) hanno introdotto il concetto di "Universal Language Model Fine-tuning" (ULMFiT), che ha dimostrato l'efficacia del fine tuning nel trasferire le conoscenze di un modello di lingua pre-addestrato a compiti di classificazione del testo. Questo metodo ha

ispirato una serie di innovazioni nel fine tuning di modelli più avanzati, come GPT-3 e GPT-4, e ha portato a miglioramenti significativi nella precisione dei modelli nelle loro applicazioni dedicate.

Nel contesto dei Foundation Models, il fine tuning è particolarmente importante per raggiungere un alto grado di specializzazione. Brown et al. (2020) hanno dimostrato che GPT-3, nonostante la sua versatilità, beneficia notevolmente del fine tuning per compiti come la comprensione del testo e la generazione di risposte in contesti specifici. Il fine tuning consente ai modelli di superare le sfide poste da compiti che richiedono una comprensione più profonda di nicchie specifiche o contesti particolarmente complessi. Il fine tuning è diventato una componente standard nel workflow di sviluppo dei modelli di intelligenza artificiale moderni. Consente ai ricercatori e ai praticanti di sfruttare la potenza computazionale e la conoscenza generale dei Foundation Models, adattandoli per soddisfare esigenze precise e per migliorare le prestazioni in una varietà di applicazioni, dalla medicina personalizzata al servizio clienti automatizzato. In definitiva, il fine tuning è essenziale per trasformare un modello generico in uno strumento specializzato, rendendolo uno dei pilastri dell'AI applicata.

Il "Prompt Engineering" è una pratica emergente nel campo dell'intelligenza artificiale, particolarmente rilevante per l'operatività dei modelli generativi come ChatGPT e altri sistemi di Generative AI. Consiste nell'elaborare e progettare input specifici (prompt) che guidino il modello a generare output desiderati senza necessità di un addestramento ulteriore. Il prompt engineering sfrutta la comprensione pre-addestrata del linguaggio dei modelli AI per produrre risposte specifiche e accurate. Questa pratica si basa sull'ipotesi che, sebbene i modelli siano stati addestrati su vasti dataset, il modo in cui una richiesta è formulata può influenzare significativamente la qualità e la pertinenza della risposta generata. Gli studi di GPT-3 hanno dimostrato l'efficacia del prompt engineering nel guidare il modello verso compiti specifici utilizzando esempi mirati nel prompt iniziale, Brown et al. (2020). Nel contesto dei chatbot e della General AI, il prompt engineering è diventato uno strumento fondamentale. Permette agli sviluppatori di modellare le interazioni senza dover ricorrere a complesse pipeline di addestramento o a interventi manuali. Attraverso l'ingegnerizzazione di prompt ben congegnati, è possibile indirizzare modelli come ChatGPT per svolgere compiti di assistenza clienti, tutoraggio, terapia conversazionale e altre applicazioni interattive e di NLP.

La pratica del prompt engineering è cruciale per sfruttare appieno le capacità dei modelli di intelligenza artificiale. La precisione del prompt può significativamente elevare la qualità dell'output, riducendo ambiguità e migliorando la rilevanza delle risposte. Questo è particolarmente importante in applicazioni dove le risposte accurate sono critiche, come nell'assistenza medica o legale, per esempio. La capacità di manipolare e dirigere l'AI attraverso prompt sofisticati è divenuta una competenza chiave nel campo dell'intelligenza artificiale. L'impiego del prompt engineering nel campo dei Foundation Models evidenzia la flessibilità e l'efficienza di questi sistemi AI avanzati. Inoltre, evidenzia il cambiamento paradigmatico nel modo di interagire con l'intelligenza artificiale: da un modello di addestramento e fine-tuning specifico per compito, a un approccio basato su interazioni ingegnerizzate che sfruttano la capacità del modello di inferire e generare risposte da prompt ingegnosi.

2.3. Impatto di GPT nelle AI Generative

Le Intelligenze Artificiali Generative, guidate da avanzamenti come i modelli GPT di OpenAI, hanno aperto un ventaglio di possibilità che permeano la vita quotidiana. Queste tecnologie non solo migliorano le operazioni esistenti ma creano anche nuovi paradigmi di interazione, innovazione e creatività. Al giorno d'oggi, AI come GPT si sono integrate in maniera sempre più invisibile e indispensabile in numerosi aspetti della vita di tutti i giorni. L'AI ha rivoluzionato il campo dell'educazione e del lavoro, offrendo strumenti che possono personalizzare l'apprendimento e aumentare la produttività. Dalla generazione di materiale didattico personalizzato alla semplificazione dei flussi di lavoro complessi, l'AI Generativa sta potenziando gli individui ad essere più efficienti ed efficaci nelle loro attività.

Nel novembre 2022, il rilascio di ChatGPT da parte di OpenAI ha rappresentato un punto di svolta nell'adozione di massa dei chatbot conversazionali. Questo modello avanzato di intelligenza artificiale si è distinto come un esemplare preminente nell'ambito dei sistemi conversazionali basati sull'AI, non solo per la sua sofisticata capacità di generare testo, ma anche per la sua straordinaria abilità di condurre interazioni naturali e significative. Prima di ChatGPT, i chatbot erano spesso limitati da interazioni rigide e scriptate che compromettevano l'esperienza dell'utente. Con l'introduzione di ChatGPT, basato

sull'architettura di GPT-3, si è assistito a un notevole avanzamento nella fluidità e nella naturalità dei dialoghi generati dall'AI. Questo ha permesso una comunicazione intuitiva, caratterizzata da una comprensione contestuale precedentemente irraggiungibile per i chatbot.

ChatGPT ha raggiunto un milione di utenti in soli cinque giorni dal lancio, un'impresa che ha segnato un record per la base di utenti in più rapida crescita per un'applicazione consumer. Questa adozione esponenziale è proseguita nel tempo, con un incremento degli utenti del 9900% nei primi due mesi, portando il numero totale di utenti attivi a 100 milioni entro gennaio 2023. Entro aprile 2023, il traffico stimato degli utenti aveva raggiunto i 173 milioni, confermando la tendenza di una rapida espansione dell'utenza (McKinsey, 2023). Il sito web di ChatGPT ha ricevuto un numero stimato di 1,66 miliardi di visite nell'ultimo mese registrato (Ottobre 2023), che, nonostante una leggera diminuzione rispetto al mese precedente, rappresentava ancora un incremento significativo dall'inizio dell'anno. Inoltre, attualmente ChatGPT attira circa 1 miliardo di visitatori mensili, consolidando la sua posizione come uno strumento di interazione AI di massima rilevanza (Notta.ai). Le previsioni economiche di OpenAI per ChatGPT sono altrettanto notevoli, con un'entrata stimata di \$200 milioni entro la fine del 2023 e una previsione di raggiungere 1 miliardo di dollari entro la fine del 2024. Questi numeri mettono in luce non solo la velocità di adozione del chatbot, ma anche il suo potenziale impatto economico a lungo termine.

La capacità di ChatGPT di mantenere il contesto su più turni di conversazione e di adattarsi agli input degli utenti in tempo reale ha catalizzato un rapido incremento nella sua adozione. Questo è stato possibile grazie a innovazioni tecniche significative che hanno migliorato l'usabilità e l'accessibilità di queste AI. La ricerca di Bender et al. (2021) ha evidenziato come il design e l'ingegneria dei prompt di ChatGPT abbiano permesso una personalizzazione senza precedenti nelle risposte dell'AI, rendendolo estremamente versatile in una varietà di contesti applicativi. L'adozione di massa di ChatGPT è stata ulteriormente accelerata dalla sua implementazione in settori chiave come il supporto clienti e l'istruzione. La facilità con cui ChatGPT poteva essere integrato in piattaforme esistenti ha permesso alle organizzazioni di offrire servizi migliorati e più personalizzati. Le organizzazioni hanno riconosciuto il valore aggiunto di un'interazione cliente più naturale e coinvolgente, portando a una rapida espansione dell'utilizzo di questi sistemi.

In sintesi, l'adozione di ChatGPT ha dimostrato l'immensa potenzialità delle AI conversazionali e ha evidenziato un interesse crescente per le tecnologie che possono fornire interazioni naturali e intuitive. L'ampio utilizzo di ChatGPT evidenzia un cambiamento nel modo in cui gli utenti interagiscono con la tecnologia, preferendo soluzioni che offrono risposte immediate e contestualizzate. L'adozione di massa di ChatGPT è un chiaro indicatore della fiducia degli utenti nella capacità di questa tecnologia di migliorare vari aspetti della vita quotidiana e professionale.

L'arrivo di ChatGPT sul mercato ha segnato un'epoca di intensificato interesse e attenzione globale verso le AI generative. Questo fenomeno ha stimolato un'ondata di innovazioni e ha influenzato notevolmente il comportamento dei competitor nel campo dell'intelligenza artificiale. Dopo il lancio di ChatGPT, vi è stata una marcata crescita nell'interesse verso le AI, non solo all'interno della comunità tecnologica, ma anche tra il grande pubblico. La facilità di utilizzo e l'ampia gamma di applicazioni di ChatGPT hanno attirato l'attenzione dei media, degli investitori, e degli accademici, ponendo l'AI generativa sotto un riflettore di rilevanza globale. Ciò ha aperto un dialogo più ampio sul potenziale e le implicazioni future di queste tecnologie. Il lancio di ChatGPT ha innescato un interesse senza precedenti verso le AI generative, non solo all'interno della comunità tecnologica, ma anche tra il grande pubblico. Questo incremento dell'attenzione è evidenziato da un significativo aumento delle discussioni online sull'AI, in particolare dopo il rilascio di GPT-4 in aprile 2023, e da ChatGPT che ha raggiunto circa 100 milioni di utenti attivi mensili già a gennaio 2023, a soli due mesi dal suo lancio (University of Rochester, 2023). Le ricerche hanno esplorato la percezione pubblica e accademica di queste tecnologie, mostrando un interesse che travalica le professioni IT, coinvolgendo individui di vari settori professionali. L'analisi di oltre 300.000 tweet e più di 150 articoli scientifici ha rivelato che ChatGPT è generalmente percepito come un sistema di alta qualità, evocando sentimenti prevalentemente positivi e emozioni di gioia nelle discussioni sui social media. Tuttavia, sono state sollevate preoccupazioni etiche riguardo alle capacità di ChatGPT (ExplodingTopics, 2023).

Inoltre, i temi di maggiore interesse includono lo sviluppo e la formazione dei modelli AI, l'AI nel business, la creatività e l'influenza potenziale dell'AI sulla società (ExplodingTopics, 2023). Questa crescente attenzione porta un cambio nel panorama dell'interazione uomo-macchina, ponendo l'AI generativa sotto un riflettore di rilevanza globale e aprendo un dialogo più ampio sulle sue potenzialità e implicazioni future.

3. IL CUSTOMER SERVICE

3.1. Che cos'è il Customer Service

3.1.1. Definizioni ed Origini

Il concetto di customer service è fondamentale nel contesto aziendale moderno, essendo un punto di contatto cruciale tra un'organizzazione e i suoi clienti. Il customer service viene definito come l'insieme delle interazioni tra un'azienda e i suoi clienti, con l'obiettivo di soddisfare le esigenze e le aspettative dei clienti per garantire la loro soddisfazione e fidelizzazione, Zeithaml e Berry (1985). La definizione stilata nel 1985 pone l'obiettivo del customer service nella soddisfazione e fidelizzazione, lasciando un'impronta generica di intesa che tale servizio sia svolto a seguito di un pre-esistente relazione o alla chiusura di una ipotetica vendita, con l'obiettivo di fidelizzare. In questo contesto, il customer service si evolve da una semplice funzione di assistenza post-vendita a una componente strategica dell'esperienza complessiva del cliente, come evidenziato da Gronroos (1990), i temi principali che aggiunge alla definizione includono l'importanza di gestire i punti di contatto critici con i clienti, l'impiego di strategie di marketing orientate al servizio, la creazione di valore attraverso interazioni di servizio. Presentando per la prima volta il concetto di servitizzazione di un processo verso un cliente e non di un prodotto. Questa evoluzione riflette il cambiamento nelle aspettative dei consumatori e la crescente complessità delle interazioni cliente-azienda. Sempre negli anni '90, la letteratura in materia sottolinea l'importanza di un customer service di qualità come fattore chiave per il successo aziendale. Questo si basa su elementi quali l'affidabilità, l'empatia, la reattività, la sicurezza e l'aspetto tangibile del servizio offerto, Zeithaml e Parasuraman (1990). Solo nei primi anni 2000, Associato al termine Customer Service vengono posizionate le parole "Risoluzioni dei problemi", infatti nel 2005 l'obiettivo principale del customer service era quello di risolvere problemi, fornire informazioni e migliorare l'esperienza complessiva del cliente con l'azienda, Johnston e Clark (2005) La storia del customer service si intreccia involontariamente con l'evoluzione delle esigenze dei consumatori e delle capacità delle aziende di rispondere efficacemente a queste nuove esigenze. Da un inizio focalizzato principalmente sull'assistenza post-vendita, il customer service si è evoluto per coprire un'ampia gamma di interazioni, che vanno dall'assistenza pre-acquisto all'assistenza continua post-acquisto. Questo sviluppo è stato dettagliato da

autori come Solomon (2014) nel suo libro "Managing Customer Experience and Relationships", che sottolinea l'importanza di un servizio clienti integrato in tutte le fasi del percorso del cliente.

La portata e la funzione del customer service hanno subito un'espansione significativa nel corso degli anni. Le aziende ora vedono il customer service sia come un centro di costo che come un'opportunità per distinguersi dalla concorrenza, aumentare la soddisfazione del cliente e costruire relazioni a lungo termine. Il customer service moderno richiede una stretta integrazione, verticale ed orizzontale, con diverse aree aziendali, inclusi marketing, vendite e supporto tecnico, per garantire un'esperienza cliente coerente e positiva, Bitner e Gremler (2018). Nel contesto del percorso del cliente, il customer service assume un ruolo chiave, soprattutto nell'era del digitale, in diverse fasi: pre-acquisto, acquisto e post-acquisto. In ogni fase, il servizio clienti contribuisce a plasmare l'esperienza complessiva del cliente influenzando la percezione del marchio e la probabilità di ripetere.

L'evoluzione dei bisogni dei clienti e la conseguente risposta aziendale hanno svolto un ruolo chiave nel plasmare il paesaggio del customer service come lo conosciamo oggi. Inizialmente, i bisogni dei clienti erano relativamente semplici e si concentravano principalmente sulla qualità del prodotto e sulla risoluzione dei problemi post-vendita. Tuttavia, con l'evoluzione dell'economia e il progresso tecnologico, questi bisogni sono diventati più complessi e diversificati. I clienti hanno iniziato a cercare esperienze più personalizzate e risposte rapide alle loro richieste, sia prima, sia durante che dopo l'acquisto.

Le aziende, rispondendo a queste mutevoli esigenze, hanno adottato approcci più innovativi e orientati al cliente nel fornire il servizio. L'introduzione di tecnologie come i call center, i numeri verdi, la risposta vocale interattiva (IVR) e, successivamente, Internet, e-mail e chat dal vivo ha permesso alle aziende di offrire un servizio più efficiente e accessibile. Questo cambiamento è stato guidato dalla necessità di mantenere la vicinanza al cliente in un mercato sempre più competitivo, Solomon (2014). Con l'avvento dei social media e il crescente utilizzo di dispositivi mobili, le aziende hanno dovuto adattarsi ulteriormente, offrendo supporto e interazioni in tempo reale attraverso queste nuove piattaforme. La personalizzazione del servizio è diventata una priorità, con le aziende che utilizzano dati e analisi per comprendere meglio le esigenze dei clienti e

offrire soluzioni su misura. L'integrazione dell'intelligenza artificiale e dei chatbot nel customer service, rappresenta l'ultima frontiera di questo processo evolutivo, offrendo risposte immediate e personalizzate 24/7 ai clienti. L'evoluzione dei bisogni dei clienti e la risposta delle aziende a questi cambiamenti hanno guidato l'evoluzione del customer service da un approccio tradizionale, centrato sul prodotto, a uno moderno, orientato all'esperienza e alla personalizzazione del cliente. Questo cambiamento ha richiesto un costante aggiornamento delle strategie aziendali e l'adozione di nuove tecnologie per rimanere competitivi e soddisfare efficacemente le esigenze dei clienti.

3.1.2. Multidimensionalità del Customer Service

L'analisi del Customer Service attraverso una lente multidimensionale si rivela essenziale per disvelare la sua complessa natura e il suo significativo contributo alle dinamiche aziendali odierne. Questo approccio, esplorato dettagliatamente nel capitolo, si basa su tre assi fondamentali che ne costituiscono il fulcro concettuale.

Il primo asse esamina le funzioni intrinseche del Customer Service all'interno dell'azienda. Questa analisi include un esame approfondito delle varie attività e responsabilità che definiscono il Customer Service, quali la risoluzione di problemi, il supporto durante la fase pre-vendita, la cura delle relazioni con la clientela, e il trattamento dei feedback e delle recensioni. Studi e ricerche accademiche offrono un'ampia evidenza dell'importanza di queste funzioni nel migliorare l'esperienza del cliente e nel rafforzare la sua lealtà verso l'azienda, Homburg et al. (2017), di Kumar e Reinartz (2016).

Il secondo pilastro pone l'accento sull'importanza dell'integrazione del Customer Service nelle varie funzioni aziendali, evidenziando come una collaborazione sinergica e una comunicazione efficace tra il Customer Service e altre aree, quali marketing, vendite, sviluppo del prodotto e analisi di mercato, siano essenziali per il successo e l'efficacia dell'assistenza al cliente. Tale integrazione consente al Customer Service di contribuire significativamente agli obiettivi aziendali e al continuo miglioramento dell'offerta di prodotti e servizi, Payne e Frow (2017).

Infine, si considera il posizionamento del Customer Service nell'arco dell'esperienza cliente, dall'acquisto alla post-vendita. È imperativo che il Customer Service moduli la propria strategia e azioni in base alle esigenze e alle aspettative dei clienti in ciascuna

fase, con l'obiettivo di ottimizzare il coinvolgimento e il valore generato dall'interazione con l'azienda, Lemon e Verhoef (2016).

Questi tre assi forniscono una struttura solida per comprendere il ruolo vitale del Customer Service nel contesto aziendale moderno, evidenziando la necessità di strategie di assistenza al cliente che siano efficaci, integrate e orientate al futuro. Nell'ambito del customer service contemporaneo, le funzioni si sono evolute per rispondere a esigenze sempre più complesse e differenziate dei consumatori e delle aziende stesse. Il customer service non è più solo un punto di contatto per la risoluzione dei problemi, ma si è trasformato in un elemento strategico e multidimensionale che incide significativamente sull'esperienza e sulla soddisfazione del cliente e sulle dinamiche strategiche aziendali di raccolta dati e customer success.

Attualmente il Customer Service adotta un ruolo strategico e si posiziona in modo multifunzionale all'interno delle organizzazioni, agendo come un fulcro tra l'azienda stessa e i suoi clienti. Le principali modalità di posizionamento si presentano in 4 categorie principali:

- **Integrazione Aziendale:** Fa parte sia della catena del valore che del team di gestione delle relazioni con i clienti (CRM), contribuendo a molteplici aspetti del business.
- **Interfaccia Cliente-Azienda:** Funge da punto di contatto principale per i clienti, sia nel pre che nel post-acquisto, influenzando direttamente la loro esperienza e soddisfazione.
- **Supporto Strategico:** Informa le strategie aziendali fornendo insight sui clienti, le loro esigenze e comportamenti.
- **Operazioni di Business:** È coinvolto nelle operazioni quotidiane, dalla gestione delle richieste dei clienti alla risoluzione dei problemi.

3.2. Evoluzione del Customer Service

La storia del customer service si può tracciare fin dalle origini dell'economia di mercato, quando il concetto di assistenza al cliente iniziò a prendere forma. Originariamente, il customer service era una funzione diretta e personale, spesso gestita direttamente dai

proprietari delle aziende o dai venditori, con un forte focus sul rapporto faccia a faccia con il cliente. Con la rivoluzione industriale e l'emergere di mercati più vasti e diversificati, il customer service iniziò a evolversi per adattarsi alle nuove esigenze di un pubblico più ampio.

Nei primi esempi di customer service, si osserva un'enfasi sull'interazione diretta e personale, con una forte attenzione alla qualità del prodotto e alla soddisfazione immediata del cliente. Questo approccio era limitato dalla scala e dalla portata dell'operatività aziendale, concentrato principalmente su piccoli mercati locali, Grönroos (1994).

Con l'espansione dell'economia industriale e il sorgere di mercati più ampi e diversificati, si rese necessario per le aziende adottare metodi più strutturati di customer service. Nacquero così i call center. I call center hanno permesso alle aziende di gestire un volume maggiore di richieste di assistenza clienti, mantenendo un contatto diretto, sebbene a distanza. Aksin et al. (2007). La nascita del marketing e il focus sull'esperienza del cliente hanno segnato un'importante svolta nella storia del customer service. Le aziende iniziarono a investire in dipartimenti dedicati, che si occupavano non solo dell'assistenza post-vendita, ma anche di gestire le relazioni con i clienti, con l'obiettivo di aumentare la soddisfazione e la fedeltà, Lovelock e Wirtz (2011).

L'avvento di Internet ha segnato un'altra svolta significativa. Le aziende hanno iniziato a utilizzare i siti web come strumento di customer service, offrendo FAQ, sezioni di aiuto, moduli di contatto e più avanti, i Chatbot. I siti web hanno fornito un nuovo livello di accessibilità e convenienza per i clienti, consentendo loro di ottenere assistenza in qualsiasi momento, aumentando la portata non solo ad aree geografiche, come nel caso dei call center, ma anche nelle fasce orarie, Holloway e Beatty (2003). Con la diffusione dei social media, un'altra trasformazione ha preso piede. Le piattaforme social sono diventate strumenti vitali per il customer service, permettendo un'interazione diretta e tempestiva con i clienti, manifestando l'importanza dei social media come canali diretti di comunicazione per le aziende, Jansen et al. (2009). Infine, l'introduzione dei chatbot basati su intelligenza artificiale ha rappresentato l'ultima evoluzione in questo settore. I chatbot hanno offerto un modo per automatizzare e personalizzare l'assistenza clienti, Questi sistemi AI possono gestire richieste non solo di base ma anche avanzate e molto specifiche, liberando risorse umane per compiti più complessi.

3.2.1. Evoluzione Tecnologica

L'evoluzione tecnologica del customer service ha segnato profondamente la storia delle interazioni azienda-cliente, partendo dalle prime forme di comunicazione e registrazione fino all'introduzione di tecnologie avanzate come l'intelligenza artificiale e la realtà aumentata. Questo percorso, che abbraccia millenni di innovazioni, riflette non solo il progresso tecnologico, ma anche il cambiamento nelle aspettative e nelle esigenze dei clienti. Dalle prime forme di registrazione su gettoni e papiro nell'antichità fino al primo reclamo di servizio clienti registrato su una tavoletta di argilla nel 1750 a.C., si evidenzia l'importanza della documentazione e della comunicazione nelle transazioni commerciali. Questi metodi primitivi di registrazione e reclamo hanno posto le basi per lo sviluppo di sistemi di customer service strutturati.

L'invenzione del telefono nel 1876 ha rappresentato una svolta significativa, permettendo per la prima volta la comunicazione diretta a distanza tra aziende e clienti. Questa innovazione ha aperto la strada alla creazione dei primi call center negli anni '60, destinati a gestire un volume crescente di chiamate e a migliorare l'efficienza del servizio clienti. Con l'introduzione dell'email nel 1972, le aziende hanno avuto accesso a un nuovo canale di comunicazione digitale, che ha ulteriormente rivoluzionato le interazioni con i clienti. L'avvento di Internet nel 1991 ha trasformato radicalmente le modalità di accesso alle informazioni e ai servizi, consentendo una comunicazione più rapida e versatile.

Nel corso degli anni '90 e 2000, l'email è diventata uno strumento standard di comunicazione aziendale, mentre i social media hanno aperto nuove vie per l'assistenza clienti, promuovendo l'interazione diretta e pubblica. L'emergere di smartphone e app negli anni 2010 ha reso possibile l'accesso ai servizi clienti in movimento, migliorando ulteriormente l'accessibilità e la comodità per i consumatori. L'introduzione di chatbot multi risposta negli anni 2010 ha segnato un ulteriore passo avanti, automatizzando le risposte ai clienti e migliorando l'efficienza del servizio. Contemporaneamente, l'analisi avanzata dei dati ha permesso di personalizzare l'esperienza del cliente, offrendo un servizio più mirato e efficace.

Recentemente, lo sviluppo di chatbot basati sull'intelligenza artificiale e l'uso della realtà aumentata e virtuale hanno offerto esperienze immersive per il supporto e la dimostrazione dei prodotti, sottolineando il continuo impegno verso l'innovazione nel campo del customer service. Queste milestone tecnologiche non solo dimostrano

l'evoluzione del customer service nel corso dei secoli, ma sottolineano anche l'importanza di adattarsi continuamente alle nuove tecnologie per soddisfare le esigenze in evoluzione dei clienti. Per approfondimenti, si rimanda alla consultazione di pubblicazioni accademiche e libri che trattano la storia della tecnologia del customer service e le sue implicazioni per le pratiche aziendali moderne.

3.2.2. Trasformazione Digitale

La trasformazione digitale ha rivoluzionato il settore del Customer Service, introducendo cambiamenti sostanziali nella modalità di interazione tra aziende e clienti. Questo fenomeno ha portato a una serie di sviluppi significativi:

Connettività e Velocità delle Informazioni: L'era digitale ha garantito una connettività senza precedenti, accelerando lo scambio di informazioni tra aziende e clienti e ottimizzando i processi aziendali interni. Questo ha migliorato notevolmente l'efficienza del servizio offerto, un principio che si estende a vari settori, Gomber et al. (2017).

Nuovi Modelli di Business: L'adozione delle tecnologie digitali ha aperto la strada a modelli di business innovativi. In particolare, nel settore dei servizi finanziari, le FinTech e altre imprese hanno sfruttato queste tecnologie per introdurre prodotti e servizi migliorati, cambiando radicalmente il modo in cui comunicano e interagiscono con i clienti, Warner & Wäger (2019).

Digitalizzazione dei Processi di Servizio: Le aziende, specialmente nel settore delle telecomunicazioni, stanno trasformando i loro servizi per adattarli all'era digitale. Ciò comporta la digitalizzazione dei processi di servizio clienti attraverso l'intero ciclo di vita del cliente, influenzando non solo le relazioni esterne ma anche l'organizzazione interna e l'offerta di valore al cliente, Valdez-de-Leon (2016).

Agilità e Rinnovamento Strategico: La trasformazione digitale richiede un impegno costante verso l'innovazione e l'adozione di nuove tecnologie digitali. L'agilità organizzativa diventa cruciale per il rinnovo strategico, implicando un cambiamento non solo nei modelli di business ma anche nella collaborazione, nella cultura aziendale e nell'approccio al mercato, Warner & Wäger (2019).

La digitalizzazione del customer service ha segnato una svolta nel rapporto tra aziende e clienti, introducendo modalità di interazione innovative che rispecchiano le mutevoli

aspettative dei consumatori nell'era digitale. La letteratura recente sottolinea come l'adozione di chatbot e assistenti virtuali basati sull'intelligenza artificiale abbia trasformato il servizio clienti, offrendo risposte immediate e personalizzate 24/7 e alleggerendo il carico di lavoro dei team di supporto, Jain et al. (2021). Questi strumenti non solo migliorano l'efficienza operativa ma arricchiscono anche l'esperienza del cliente, un aspetto cruciale per la fidelizzazione e il successo aziendale a lungo termine.

L'espansione dei canali digitali ha ulteriormente democratizzato l'accesso al customer service, consentendo ai clienti di interagire con le aziende attraverso social media, app mobili e messaggistica istantanea, oltre ai tradizionali canali telefonici ed email. Questa diversificazione dei punti di contatto favorisce una comunicazione più fluida e diretta, permettendo alle aziende di cogliere feedback in tempo reale e di adattarsi rapidamente alle esigenze dei clienti, Heinonen et al. (2010).

Inoltre, la capacità di raccogliere e analizzare grandi volumi di dati sui clienti rappresenta un'altra pietra miliare della digitalizzazione del customer service. Utilizzando tecniche avanzate di data analytics, le aziende possono ora personalizzare l'esperienza di servizio in modo più accurato, anticipando le esigenze dei clienti e offrendo soluzioni proattive. Questo approccio orientato ai dati non solo aumenta la soddisfazione del cliente ma promuove anche una relazione più profonda e duratura con il marchio.

La trasformazione digitale del customer service implica anche un cambiamento nella cultura organizzativa, richiedendo un impegno da parte di tutti i livelli aziendali per adottare un mindset centrato sul cliente e sfruttare la tecnologia come leva per migliorare l'esperienza complessiva del cliente. Plekhanov et al. (2023).

3.2.3. La nascita del Customer Service automatico

L'evoluzione dei sistemi di Risposta Vocale Interattiva (IVR) rappresenta un'epoca significativa nel panorama del Customer Service, delineando le premesse per l'emergere dei chatbot moderni. Concepiti per rispondere alla necessità di ottimizzare la gestione delle chiamate nei centri di assistenza, gli IVR hanno implementato tecnologie software avanzate per dirigere gli utenti attraverso percorsi vocali predefiniti, consentendo loro di selezionare le informazioni richieste attraverso il tastierino telefonico o mediante comandi vocali. Questa rivoluzione, che trae origine già dagli anni '30, ha ricevuto un impulso decisivo con l'avvento della tecnologia Dual Tone Multi-Frequency (DTMF) nel

1962, introdotta da Bell. Tale innovazione ha sostituito i tradizionali sistemi a composizione rotativa con quelli a pulsante, modificando radicalmente l'interazione degli utenti con i dispositivi telefonici, Dresler et al. (2009).

La tecnologia DTMF ha aperto la strada alla creazione di menu telefonici interattivi, facilitando operazioni come "premere 1 per le vendite, premere 2 per il servizio clienti", inaugurando così una forma preliminare di automazione nel settore dell'assistenza clienti. Con l'aggiunta del riconoscimento vocale agli inizi degli anni '80, le modalità di interazione si sono ulteriormente espanse, consentendo alle persone di rispondere vocalmente oltre che tramite selezione manuale. Questo avanzamento ha segnato un passo cruciale verso l'automazione dei processi nei centri di assistenza, anticipando l'introduzione dei primi prototipi di chatbot vocali, Ben-Haim et al. (2016)

L'impulso verso lo sviluppo degli IVR è scaturito dalla necessità per le aziende di gestire efficacemente un numero crescente di chiamate, minimizzando i tempi di attesa per i consumatori e incrementando la produttività operativa. Rivoluzionando il customer service, gli IVR hanno permesso alle compagnie di garantire supporto continuativo 24 ore su 24, diminuendo la necessità di intervento umano per le richieste più semplici e consentendo una personalizzazione dell'esperienza utente. Questi sistemi hanno, inoltre, gettato le fondamenta per l'evoluzione verso strumenti di assistenza automatizzati più avanzati, culminati nell'integrazione di chatbot dotati di intelligenza artificiale e capacità di apprendimento continuo. In conclusione, l'avvento e lo sviluppo degli IVR non solo hanno inaugurato l'era dell'automazione nel Customer Service ma hanno anche facilitato la transizione verso l'adozione di chatbot, dispositivi sempre più evoluti capaci di emulare interazioni umane e fornire un servizio clienti su misura e efficiente.

3.3. Evoluzione dei Chatbot di Customer Service Pre-GPT-3

L'evoluzione dei chatbot rappresenta un'importante trasformazione nel panorama tecnologico aziendale, in particolare nel Customer Service, passando da semplici meccanismi di risposta automatizzata per l'assistenza clienti a complessi sistemi integrati nelle strategie di vendita, marketing e coinvolgimento dei consumatori. Esplorando l'evoluzione dei chatbot e del loro ruolo nell'automazione dei servizi clienti e delle

vendite, tratteremo un arco evolutivo che parte dalle prime applicazioni di Interactive Voice Response (IVR) nei Call Center fino all'avvento e all'espansione dell'intelligenza artificiale generativa (G-AI) nel 2023. Questo percorso è segnato da tappe fondamentali come l'introduzione di web chat, email, e l'utilizzo di piattaforme come Messenger, WhatsApp e Slack, che hanno gradualmente esteso l'impiego dei chatbot oltre i tradizionali ambiti di applicazione. Il presente lavoro si basa su un'analisi dettagliata delle diverse generazioni di chatbot, partendo dai sistemi basati su regole, caratteristici delle prime fasi, ai più recenti sistemi alimentati da G-AI, capaci di apprendere autonomamente da interazioni precedenti e di elaborare risposte sempre più accurate e simil-umane, Louis Moynihan (2024).

3.3.1. I primi Chatbot

L'evoluzione dei chatbot dagli anni '60 fino all'era pre-NLP rappresenta un periodo fondamentale nella storia dell'intelligenza artificiale conversazionale, segnando l'inizio dell'esplorazione dell'interazione uomo-macchina. La creazione di ELIZA nel 1964 da parte di Joseph Weizenbaum al MIT segnò una pietra miliare significativa. ELIZA, progettato per emulare una conversazione con un terapeuta Rogeriano, utilizzava regole di elaborazione del testo per riconoscere le parole chiave e rispondere con domande predefinite, offrendo agli utenti l'illusione di un dialogo umano, Weizenbaum (1966). Questo esperimento non solo dimostrava il potenziale dei computer nel simulare il dialogo umano ma apriva anche la strada a ulteriori innovazioni nel campo dei sistemi di dialogo basati su regole, Wikipedia (2024)

Nei decenni successivi, l'avanzamento tecnologico e l'espansione di Internet hanno catalizzato lo sviluppo dei chatbot, consentendo loro di diventare più sofisticati ed efficienti. Benché i primi chatbot fossero limitati nella comprensione del linguaggio naturale e nelle capacità di risposta, la loro adozione come strumenti di supporto al cliente dimostrava già un chiaro vantaggio: la possibilità di fornire assistenza ininterrotta, superando i limiti geografici e temporali tipici dell'assistenza umana. Questa fase di sviluppo dei chatbot, pur mantenendosi ancorata a meccanismi decisionali semplici e a interazioni basate su scenari predefiniti, ha gettato le basi per l'integrazione futura di tecnologie più avanzate, come l'elaborazione del linguaggio naturale (NLP), che avrebbero trasformato i chatbot in sistemi conversazionali ancora più interattivi e personalizzati.

L'interesse accademico e industriale verso i chatbot e l'intelligenza artificiale negli anni successivi alla creazione di ELIZA ha evidenziato la crescente rilevanza di queste tecnologie nell'ambito dell'interazione digitale, ponendo le premesse per la loro evoluzione verso applicazioni più complesse e capaci di offrire un'esperienza utente significativamente migliorata. Il percorso di crescita dei chatbot, dall'essere semplici risponditori automatizzati a strumenti interattivi capaci di svolgere ruoli più complessi nelle operazioni aziendali, sottolinea l'importanza dell'innovazione tecnologica continua nel campo dell'intelligenza artificiale, Tanbits (2023).

3.3.2. Gli Internet Chatbot

L'espansione dell'accessibilità a Internet attraverso la connessione broadband agli inizi del nuovo millennio ha rappresentato un punto di svolta cruciale per il settore del customer service, inaugurando un'era di automazione e interazione digitale senza precedenti. La diffusione capillare dei siti web aziendali, facilitata da questa nuova ondata di connettività, ha ridefinito le modalità con cui le imprese entrano in contatto con la propria clientela, introducendo canali di comunicazione innovativi come l'email e il webchat. In questo contesto, società all'avanguardia quali Zendesk.com, LivePerson.com e Intercom.com hanno assunto un ruolo pionieristico, specializzandosi rispettivamente nella gestione del servizio clienti tramite email e nella fornitura di piattaforme di webchat, delineando così i contorni di un nuovo capitolo nell'evoluzione dell'automazione nel servizio di assistenza.

La predominanza di input basati su testo in questi canali ha aperto la strada a possibilità di automazione finora inesplorate, superando le limitazioni tecnologiche dell'epoca, come il riconoscimento vocale nelle sue fasi embrionali e la rigida composizione tramite tastiera telefonica. In particolare, il webchat, grazie all'utilizzo di linguaggi di programmazione quali HTML e JavaScript, ha offerto agli utenti un'interfaccia interattiva altamente accessibile, attraverso cui è possibile navigare con facilità grazie a risposte automatizzate organizzate in menu simili a pulsanti McTear et al. (2016).

Questo nuovo panorama digitale ha esercitato un profondo impatto sui call center, determinando una migrazione degli utenti dai canali tradizionali a piattaforme online per l'acquisto di beni e servizi e promuovendo l'utilizzo di dispositivi mobili. Di conseguenza, è emersa l'industria degli SMS outbound, caratterizzata da una propria automazione

basata su regole. L'adozione di software SAAS (Software as a Service) per gestire questi canali emergenti ha dato vita all'omnicanalità, estendendo l'automazione a un'ampia gamma di canali di comunicazione. Nonostante l'automazione negli IVR sia rimasta in gran parte legata a logiche basate su regole e implementata prevalentemente in loco, l'automazione via email e webchat ha offerto nuove opportunità per un'elaborazione più profonda e complessa dei compiti, Kumar & Telang (2012).

L'evoluzione dei software SAAS per call center ha costituito un momento decisivo, offrendo agli operatori, oltre che agli utenti finali, una nuova interfaccia. Questo ha portato a distinguere tra automazioni rivolte agli utenti e quelle destinate agli agenti, queste ultime frequentemente associate al concetto di "Human Assist AI". La trasformazione dell'input testuale in automazione e la flessibilità offerta dai sistemi SAAS e basati su cloud hanno ampliato le possibilità di estensione del servizio. Ciononostante, nonostante la prevalenza dell'email come canale principale, questo soffre di bassi tassi di apertura e conversione, mentre il webchat, sebbene faciliti una comunicazione più immediata, trova limitazioni nella sua adozione, particolarmente nei paesi con un forte utilizzo del desktop, e non supporta iniziative outbound come quelle di marketing, Hoy (2018).

3.3.3. I Chatbot “Tradizionali”

Nel panorama del servizio clienti digitale, Dagli anni 2010 in poi, l'adozione di chatbot operanti secondo principi logici definiti, ovvero “Rule-based”, ha segnato un passo significativo verso l'ottimizzazione delle interazioni azienda-cliente tramite macchina. Questi strumenti, progettati per navigare le richieste dei clienti seguendo percorsi di risposta "If-Then", emergono come soluzioni efficaci per gestire con precisione un'ampia varietà di quesiti. Sviluppati per far fronte all'esigenza di elaborare un volume crescente di interazioni in modo efficiente, questi sistemi di risposta automatizzati si avvalgono dell'identificazione di parole chiave o frasi specifiche negli input degli utenti per fornire feedback immediato e pertinente.

L'introduzione di chatbot basati su regole nel contesto del servizio clienti si inserisce in un contesto di sviluppo tecnologico, in particolare nell'ambito dell'elaborazione del linguaggio naturale (NLP) e della domanda per un servizio continuativo, disponibile 24 ore su 24. La caratteristica distintiva, e al contempo limitante, di questi chatbot risiede

nella loro capacità di operare all'interno di parametri predeterminati, offrendo una soluzione affidabile per applicazioni dove le risposte possono essere uniformate, come nei settori bancario, dell'e-commerce e dei servizi pubblici.

Differenti tipologie di chatbot basati su regole si sono evolute per soddisfare vari aspetti del servizio clienti. Ad esempio, i chatbot con opzioni a pulsante facilitano la navigazione dell'utente attraverso menù intuitivi, mentre i chatbot basati su parole chiave elaborano le richieste testuali identificando termini specifici per fornire risposte mirate. Altri modelli avanzati, quali i chatbot ad albero decisionale, impiegano sequenze di domande per affinare le esigenze dell'utente e indirizzarlo verso la soluzione o il servizio adeguato.

L'impiego di chatbot basati su regole nel servizio clienti portava vantaggi significativi, migliorando l'efficienza operativa e la capacità di gestire numerose richieste simultaneamente, senza necessità di interazione umana. Questa forma di automazione ha elevato la qualità del servizio offerto, incrementando la soddisfazione dei clienti grazie a risposte rapide e mirate, e permettendo al personale di dedicarsi a questioni più delicate che richiedono un approccio personalizzato.

La struttura degli alberi decisionali nei chatbot “Rule-based” incarna un avanzamento significativo nella programmazione informatica, mirato a facilitare l'interazione tra utenti e macchine attraverso una sequenza logica di quesiti e risoluzioni. Questi sistemi si inaugurano con un nodo principale, da cui si dipanano molteplici sentieri guidati da criteri definiti, indirizzando l'utente verso soluzioni o azioni specifiche basate sulle proprie selezioni. Ogni scelta effettuata dall'interlocutore dirige verso un ulteriore nodo, configurando il processo come un albero decisionale, in cui ogni biforcazione rappresenta un'alternativa dialogica possibile, Russell e Norvig (2016).

Sul fronte tecnico, l'elaborazione di un chatbot che sfrutta alberi decisionali esige un'attenta pianificazione delle interazioni potenziali. Gli sviluppatori sono chiamati a prevedere le interrogazioni degli utenti e a organizzare le repliche in modo coerente, per garantire che ogni percorso intrapreso conduca a un esito appropriato. Tale processo di anticipazione si basa sull'analisi della frequenza di specifiche domande o sull'utilizzo di dati accumulati per identificare le traiettorie di dialogo più ricorrenti.

I benefici di questo metodo includono rapidità ed efficienza nel trattamento di richieste standard, poiché il sistema è capace di erogare risposte prestabilite senza necessità di

complesse interpretazioni linguistiche. Inoltre, la configurazione ad albero assicura una guida uniforme dell'utente verso le informazioni necessarie, ottimizzando l'esperienza per quesiti comuni o procedure routinarie, Klopfenstein et al. (2017).

Nonostante ciò, una limitazione cruciale dei chatbot fondati su alberi decisionali è la loro intrinseca rigidità. La gestione delle interazioni è vincolata da regolamenti e itinerari predefiniti, rendendo arduo il dialogo dinamico o l'interpretazione di input inattesi. Tale inflessibilità diminuisce la capacità di adattamento rispetto a sistemi basati su intelligenza artificiale di nuova generazione, in grado di apprendere dalle interazioni e di modificare conseguentemente le proprie risposte. Ulteriormente, lo sviluppo e il mantenimento di un chatbot articolato su alberi decisionali possono rivelarsi complessi e gravosi con l'aumentare delle interazioni possibili, necessitando aggiornamenti continui per incorporare nuove varianti o cambiamenti nei servizi proposti, Jurafsky & Martin (2018).

3.3.4. I Chatbot AI Conversazionali

Infine, l'ultimo passo evolutivo dei chatbot nel Customer Service prima di arrivare all'Intelligenza Artificiale Generativa, sono i Chatbot AI cosiddetti Conversazionali. Tali Chatbot sono gli stessi che comunemente vediamo utilizzati nella maggior parte dei siti web o nelle app, e sono la massima rappresentazione evolutiva di un chatbot con Intelligenza Artificiale "Tradizionale".

L'utilizzo dell'AI conversazionale nei chatbot di Customer Service, quindi, consente di interpretare e processare le richieste degli utenti in linguaggio naturale, grazie all'impiego di Natural Language Processing (NLP) e Natural Language Understanding (NLU). Queste tecnologie permettono ai chatbot di comprendere l'intento dietro alle domande degli utenti, di estrarre informazioni rilevanti e di fornire risposte accurate e pertinenti, simulando una conversazione umana, totalmente diverso quindi rispetto ai chatbot a struttura ad albero che hanno solamente risposte predefinite, Henderson et al. (2019). Un aspetto fondamentale dell'AI conversazionale nel Customer Service è senza ombra di dubbio la sua capacità di apprendimento continuo. Attraverso il Machine Learning, i chatbot si adattano e migliorano nel tempo, ampliando la loro conoscenza e affinando le risposte fornite in base alle interazioni precedenti. Questo processo di apprendimento consente ai chatbot di diventare sempre più efficienti nella gestione delle richieste,

offrendo un servizio sempre più personalizzato e proattivo, avvicinandosi sempre di più al simulare una conversazione con una persona umana, Young et al. (2018).

L'Intelligenza Artificiale Conversazionale (Conversational AI) rappresenta quindi un avanzamento significativo nella capacità delle macchine di interpretare e rispondere al linguaggio umano in modo simile all'interazione umana. Il processo alla base di questa tecnologia è articolato in diverse fasi cruciali che consentono la trasformazione del linguaggio naturale in un formato comprensibile per il sistema e la generazione di risposte coerenti e personalizzate.

La prima fase di questo processo è rappresentata dall'elaborazione del linguaggio naturale (Natural Language Processing), che ha il compito di analizzare l'input testuale fornito dall'utente. Questa fase include la correzione ortografica, l'identificazione di sinonimi, l'interpretazione grammaticale, il riconoscimento del sentimento e la decomposizione della richiesta in elementi base (parole e frasi) per facilitare la comprensione da parte del sistema, Hirschberg e Manning (2015). Successivamente, intervengono modelli avanzati di apprendimento automatico noti come Natural Language Understanding (NLU). Questi modelli permettono all'intelligenza artificiale di determinare con precisione l'intento dietro la richiesta dell'utente e di estrarre gli elementi informativi rilevanti, facilitando così l'identificazione della risposta più adeguata alla domanda post, Young et al. (2018).

Una volta compresa la richiesta, il sistema deve formulare una risposta che sia non solo corretta ma anche personalizzata. A questo proposito, la Conversational AI utilizza le informazioni elaborate tramite l'NLU per generare risposte che simulino un'interazione umana, rendendo l'esperienza dell'utente il più naturale e soddisfacente possibile. Questo processo di generazione della risposta è spesso affidato a meccanismi di Natural Language Generation (NLG), che traducono i dati strutturati in un output linguistico coerente, Gatt e Krahmer (2018). Gli esempi più famosi di AI conversazionale sono le famose Siri o Alexa, rispettivamente di Apple e Amazon.

Nonostante i progressi tecnologici e l'ampio utilizzo nell'assistenza clienti, l'Intelligenza Artificiale Conversazionale (AI Conversazionale) presenta comunque delle limitazioni e delle sfide. Tra i limiti funzionali, si evidenzia la difficoltà nel comprendere pienamente il linguaggio naturale umano, specialmente in presenza di slang, dialetti regionali o espressioni idiomatiche. Questo può portare a incomprensioni o risposte non pertinenti, influenzando negativamente l'esperienza dell'utente. Allargandosi dagli idiomi

particolari, la gestione di conversazioni complesse che richiedono ragionamenti astratti o empatia risulta ancora una sfida per i sistemi AI Conversazionali attuali, limitando l'efficacia nei contesti più sofisticati di interazione, questo punto è la chiave di volta che poi ha permesso ai Chatbot di AI Generativa di prendere piede dopo l'uscita di GPT-3. Le principali motivazioni del passaggio all'AI generativa sono identificabili nel limite dell'AI conversazionale e di conseguenza nell'insoddisfazione dei clienti. Nello specifico l'AI conversazionale essendo limitata a dare risposte predefinite su contenuti preesistenti, tende ad assumere comportamenti devianti o ridondanti quando viene stressata o sottoposta a richieste che non trovano una risposta diretta all'interno del materiale su cui è stata direttamente addestrata. Tale comportamento porta a risposte inaccurate o fuori tema, portando di conseguenza ad un elevato grado di insoddisfazione verso il cliente che sta interagendo con il Chatbot, il tutto si aggiunge alla mancanza di Empatia. Questa limitazione diventa particolarmente evidente in contesti di interazione avanzati, dove l'AI deve interpretare e rispondere a sfumature emotive e complessità concettuali che richiedono una profonda comprensione del contesto umano e delle dinamiche interpersonali. Questa difficoltà nell'emulare la capacità umana di gestire conversazioni ricche e multidimensionali limita l'efficacia dell'AI, particolarmente nelle situazioni che richiedono una vera comprensione delle emozioni umane e delle intenzioni sottostanti, riducendo la sua capacità di fornire risposte veramente personalizzate e pertinenti. Questa limitazione nell'AI si collega direttamente all'insoddisfazione dei clienti, poiché l'empatia autentica è fondamentale per una vera comprensione e connessione, Perry, A. (2023).

3.4. I Chatbot di Intelligenza Artificiale Generativa

Lo sviluppo dei chatbot di customer service ha raggiunto un punto di svolta con l'avvento di GPT-3 e successivamente GPT-4, due versioni avanzate della tecnologia Generative Pre-trained Transformer. Questi modelli di AI generativa hanno portato a una rivoluzione nel campo, consentendo ai chatbot di comprendere e generare risposte in modo più naturale e contestualizzato. La capacità di elaborare e produrre linguaggio simile a quello umano ha notevolmente migliorato l'interazione tra aziende e clienti, rendendo i servizi di assistenza più efficienti e personalizzati. Con GPT-3 e GPT-4, i chatbot sono diventati

strumenti potenti per il customer service, capaci di gestire un'ampia gamma di richieste con un livello di comprensione precedentemente irraggiungibile

3.4.1. I Tool di Intelligenza Artificiale Generativa

Per capire l'applicazione dei Foundation models GPT-3 e GPT-4 come Chatbot di Customer Support, vediamo la necessità di fare un salto indietro e capire come funzionano le applicazioni di Intelligenza Artificiale Generativa.

Il miglior modo per comprendere come questa nuova tecnologia sia strutturata e sia presente nel nostro mondo di tutti i giorni, e cosa centri con il Customer Service è di paragonarla ad una porzione dell'universo, ovvero il Sistema solare. Se il sistema solare è la rappresentazione dell'ecosistema Mondiale delle tecnologie di Intelligenza Artificiale, possiamo identificare nel pianeta Terra il Foundation Model GPT-4, ovvero il motore di Intelligenza Artificiale Generativa sviluppato da OpenAI e rilasciato a Marzo 2023. E' conoscenza comune che i motori GPT siano diventati noti al pubblico globale grazie all'applicazione ChatGPT, che in maniera indiretta ha permesso al modello GPT-4 (Pianeta Terra) di dimostrare al mondo le sue capacità. Ed è anche questo il motivo per cui la maggior parte delle persone confondono GPT e ChatGPT come se fossero la medesima cosa. Riprendendo la metafora spaziale, se dovessimo posizionare ChatGPT nella mappa del sistema solare, sarebbe sicuramente la Luna, ovvero un satellite del pianeta terra, ovvero il modello GPT-4. Si capisce quindi che ChatGPT non è nient'altro che un'applicazione che utilizza la potenza e le capacità di GPT4 veicolate verso un'interfaccia chatbot e vincolate alla botta e risposta sotto forma di testo o immagini.

Come esiste ChatGPT, però, attorno al pianeta GPT-4 orbitano moltissimi altri satelliti, decina di migliaia di satelliti, questi corpi celesti prendono il nome di "Tool di AI Generativa". I satelliti, ovvero i "Tool" non sono altro che diverse interfacce poste sopra allo stesso motore, GPT-4, veicolandolo e vincolandolo a lavorare per ottenere output diversi da input diversi. Se dovessimo fare degli esempi, ci sono tool che permettono, dato un testo o solamente un parola di creare, con tanto di immagini e struttura, una presentazione power point pronta all'uso in decine di secondi, oppure tool che dato un riassunto di un attività d'impresa, è in grado di generare un Business Analysis con piano finanziario nel giro di qualche minuto, e così via. Gli use cases di utilizzo sono molteplici

e in continua crescita, il denominatore comune però, sono sempre i foundation model che stanno alla base, come in questo caso GPT-4.

Ritornando al Customer Service, quindi, i Chatbot che utilizzano Intelligenza Artificiale Generativa non sono altro che dei satelliti in orbita attorno a GPT-3 o GPT-4, ovvero dei Tool di Generative AI.

3.4.2. I Chatbot di Intelligenza Artificiale Generativa

Lo sviluppo dei chatbot di customer service ha raggiunto un punto di svolta con l'avvento di GPT-3 e successivamente GPT-4, due versioni avanzate della tecnologia Generative Pre-trained Transformer. Questi modelli di AI generativa hanno portato a una rivoluzione nel campo, consentendo ai chatbot di comprendere e generare risposte in modo più naturale e contestualizzato. La capacità di elaborare e produrre linguaggio simile a quello umano ha notevolmente migliorato l'interazione tra aziende e clienti, rendendo i servizi di assistenza più efficienti e personalizzati. Con GPT-3 e GPT-4, i chatbot sono diventati strumenti potenti per il customer service, capaci di gestire un'ampia gamma di richieste con un livello di comprensione precedentemente irraggiungibile. I Chatbot di AI Generativa offrono una serie di vantaggi distinti, che corrispondono quasi direttamente ai vantaggi che può offrire un motore di AI Generativa, questo perché ritornando alla metafora del sistema solare, i chatbot in questione, di ultima generazione, non sono altro che satelliti che orbitano attorno a Foundation models completi, sono però veicolati e vincolati, tramite operazioni di Fine Tuning, a generare output contestuali ai contenuti sui quali sono stati sintonizzati.

Diventa naturale conseguenza che in base al grado di libertà e personalizzazione con cui impostiamo il Chatbot ne dipenderanno conseguentemente le possibilità che un cliente può esplorare interfacciandosi con esso. Le principali caratteristiche che apportano vantaggi ai Chatbot di AI Generativa sono le stesse che, per definizione, rendono la AI Generativa di gran lunga superiore alla AI Tradizionale, e da cui, di conseguenza si sviluppano degli aspetti migliorativi nello specifico caso circoscritto dei Chatbot basati su AI Generativa. Tali aspetti sono:

- **Comprensione Contestuale Migliorata:** I chatbot basati su GPT-3 e GPT-4 possiedono capacità avanzate di comprensione del linguaggio naturale (Natural Language Understanding - NLU), consentendo loro di interpretare le richieste

degli utenti con un grado di precisione precedentemente inaccessibile. La loro architettura, addestrata su vasti dataset di testo, permette una gestione più efficace delle ambiguità linguistiche, rendendo le interazioni più fluide e naturali, Brown et al. (2020).

- Personalizzazione e Adattabilità: A differenza dei chatbot convenzionali, i sistemi basati su GPT-3 e GPT-4 possono generare risposte personalizzate in tempo reale, adattandosi al contesto e al tono della conversazione. Ciò migliora significativamente l'esperienza utente, facendo sentire il cliente compreso e valorizzato, Brown et al. (2020).

A cascata, l'integrazione di chatbot AI generativi nel servizio clienti non solo migliora l'efficienza e la qualità delle interazioni ma apre anche nuove possibilità per un servizio altamente personalizzato e reattivo. Questo si trasforma non solo in risposte più accurate o campi di applicazione più vasti, ma anche alla multimedialità delle interazioni. Con l'utilizzo di AI Generativa, i Chatbot quindi acquisiscono questi vantaggi multimodali:

- Lettura e Creazione di Contenuti Multimediali: Possibilità di interpretare e comprendere non solo richieste testuali, ma anche input video o audio. Lato Output è in grado di generare automaticamente contenuti multimediali come video tutorial o guide passo-passo in risposta a richieste specifiche. Ad esempio, di fronte a una domanda su come configurare un dispositivo, il chatbot può creare e fornire un video tutorial personalizzato che guida l'utente attraverso il processo. Oppure la richiesta di riparazione di un prodotto può essere fatta tramite il caricamento di una foto del guasto, senza necessità di spiegazioni ulteriori.
- Risposte Dinamiche e Contestuali: Grazie alle capacità di svolgere task autonomamente ad alle capacità di apprendimento dinamico, è possibile integrare informazioni provenienti da diverse fonti in tempo reale per fornire risposte che riflettono l'attuale contesto di mercato, disponibilità del prodotto, o tendenze, adattandosi dinamicamente alle circostanze del momento.
- Gestione di Richieste Complesse: La capacità del Chatbot di avere in plancia la conoscenza generale di un modello di LLM con più di 175 Miliardi di Parametri (GPT-3) o addirittura più di 100 Trilioni (GPT-4) permette l'integrazione di informazioni provenienti da diverse fonti in tempo reale, oltre a quelle a cui fare direttamente riferimento fornite dall'organizzazione nel processo di Fine-Tuning, permette di gestire richieste complesse, utilizzando il ragionamento per andare

oltre le query predefinite. Ad esempio, può gestire domande multi faccettate che richiedono una comprensione approfondita del contesto e dell'intento dell'utente, fornendo soluzioni complete.

- **Supporto Multilingua:** I chatbot di AI Generativa non hanno praticamente limitazioni dal punto di vista del linguaggio, essendo allenate su più di 180 Lingue, permettendo un'interazione multilingua e multicanale adottando lo stesso modello in qualsiasi contesto geografico.
- **Offerta di Raccomandazioni Personalizzate:** I chatbot possono integrarsi con i sistemi CRM esistenti e altri servizi per tracciare informazioni comportamentali sui clienti, offrendo suggerimenti personalizzati su quali prodotti acquistare o come risolvere determinati problemi. Inoltre, sfruttando le capacità di apprendimento automatico, i chatbot possono migliorare gradualmente queste raccomandazioni, adeguando il loro servizio per soddisfare diversi segmenti di consumatori.
- **Completamento Automatico dei Compiti:** I chatbot AI con tecnologia RPA (Robotic Process Automation) possono utilizzare flussi di lavoro automatizzati per completare vari compiti per conto di dipendenti e clienti. Questo può includere la compilazione di moduli, il follow-up con i clienti dopo un acquisto o anche la modifica di una sottoscrizione o profilo.
- **Sentiment Analysis:** Attraverso una combinazione di NLP (Natural Language Processing), linguistica computazionale e strumenti di ML (Machine Learning), i chatbot hanno sviluppato una comprensione più profonda dell'emozione umana. Possono rilevare come si sente un cliente in punti specifici del percorso cliente e utilizzare queste informazioni per determinare i prossimi passi migliori per il servizio.
- **Interazioni Omni-canale:** I chatbot non sono più solo uno strumento per la messaggistica sul sito web e le conversazioni sui social media. Le aziende di oggi possono costruire "assistenti virtuali" progettati con la stessa tecnologia dei chatbot, in grado di rispondere al linguaggio umano parlato.
- **Task Enanching:** Il tempo di confezionamento delle chiamate può ora essere ridotto significativamente grazie ai chatbot alimentati da AI che hanno la capacità di catturare e riassumere note da chiamate e altre interazioni, andando a sviluppare in catena task eseguibili e smistate alla business unit responsabile. Questo

risparmierà tempo all'Advisor nel concludere una chiamata o interazione, consentendogli di concentrarsi invece su altri compiti a valore aggiunto.

Le possibilità sono illimitatamente più vaste con l'adozione di tecnologia Generativa nei chatbot di Customer Service delle organizzazioni, ma un aspetto che non deve passare sottocchio è la facilità di implementazione. L'adozione delle Application Programming Interfaces (API) basate su GPT-3 e GPT-4 nel campo dei chatbot rappresenta un'innovazione dirompente nel settore del supporto al cliente e dell'interazione digitale. Questa metodologia si distingue radicalmente dalle pratiche tradizionali, che prevedono la creazione manuale di chatbot, introducendo un approccio incentrato sull'integrazione piuttosto che sulla costruzione ex-novo. Diversamente dai sistemi conversazionali tradizionali, che richiedono estese fasi di sviluppo, inclusa l'elaborazione di flussi di dialogo, l'addestramento su specifici dataset e lo sviluppo di algoritmi per il trattamento del linguaggio naturale (NLP), l'implementazione dei chatbot basati su GPT-3 e GPT-4 si avvale dell'utilizzo delle API fornite da OpenAI. L'integrazione tramite API elimina la necessità di affrontare la gestione di infrastrutture IT complesse o di disporre di un team dedicato all'intelligenza artificiale, consentendo alle organizzazioni di incorporare le funzionalità di GPT-3 e GPT-4 nei loro sistemi esistenti, come i siti web, con un impegno minimo in termini di tempo e risorse.

Un ulteriore vantaggio offerto dalle API è la possibilità di personalizzare il modello (fine-tuning) per adeguarlo alle esigenze specifiche di un'azienda. Attraverso la fornitura di esempi di interazioni target, è possibile affinare ulteriormente il modello, garantendo che le risposte del chatbot riflettano fedelmente il tono, lo stile e le peculiarità del brand. Questo processo di personalizzazione continua assicura che il chatbot mantenga la sua efficacia e rilevanza nel tempo, contribuendo a ottimizzare l'esperienza utente e la soddisfazione del cliente. La riduzione dei costi e degli oneri manutentivi costituisce un altro elemento chiave. La responsabilità dell'infrastruttura sottostante e degli aggiornamenti del modello è a carico di OpenAI, permettendo alle aziende di concentrarsi sull'impiego del chatbot senza doversi occupare della gestione tecnica. Questo aspetto non solo comporta una diminuzione dei costi operativi ma accelera anche i tempi di implementazione, favorendo l'agilità aziendale e la capacità di rispondere prontamente alle mutevoli esigenze dei clienti.

4. CASO STUDIO: EUROTHERM

Il terzo capitolo delinea l'evoluzione e l'innovazione portata dai chatbot di Intelligenza Artificiale Generativa nel campo del Customer Service, evidenziando come la transizione da sistemi conversazionali tradizionali a modelli generativi avanzati come GPT-3 e GPT-4 abbia segnato una svolta significativa nell'interazione tra organizzazioni e clienti. La capacità di questi chatbot di offrire risposte contestualizzate, personalizzate e di gestire complessità e multimedialità senza precedenti non solo ha migliorato l'efficienza e l'efficacia del supporto al cliente, ma ha anche aperto nuovi orizzonti per la personalizzazione del servizio e l'engagement del cliente. La facilità di implementazione tramite API, inoltre, ha reso l'adozione di queste tecnologie accessibile a un'ampia gamma di aziende, promettendo una trasformazione profonda nel panorama del customer service.

Nel quarto capitolo, ci addenteremo in un caso studio concreto che illustra l'applicazione pratica e i potenziali cambiamenti derivanti dall'implementazione di un chatbot di AI generativa all'interno di Eurotherm, un'azienda operante nel settore HVAC. Il capitolo ha l'obiettivo di Capire come l'Intelligenza Artificiale Generativa, e non solo i chatbot di AI Generativa, possano influire nei processi, nell'operatività e nelle strategie di un'organizzazione.

4.1. Scopo della Ricerca e Metodologia

Questo capitolo presenta un caso studio di successo dell'implementazione della Generative AI in azienda, nello specifico nel customer service di Eurotherm, azienda leader nel settore del comfort climatico indoor. Attraverso l'intervista al responsabile dell'innovazione in azienda, analizziamo le fasi della trasformazione AI, le aree di impatto e i risultati ottenuti.

La ricerca si propone di esplorare l'impiego dell'Intelligenza Artificiale Generativa (AI Generativa) nel contesto del customer service di Eurotherm, azienda di punta nel settore del comfort climatico indoor. L'obiettivo principale è comprendere in che modo l'adozione di questa tecnologia possa influenzare i processi strategici e operativi dell'azienda, con un focus particolare sull'assistenza ai clienti. Questo studio mira a rispondere alla domanda di ricerca centrale: "Come l'AI Generativa può impattare i

processi strategici ed operativi in un'azienda, con un focus sul Customer Service?". Lo scopo di questa ricerca è duplice. Da un lato, si intende illustrare come l'AI Generativa possa essere sfruttata per migliorare i processi di customer service in un'azienda leader nel suo settore, fornendo un modello replicabile per altre realtà aziendali interessate a innovare i propri servizi. Dall'altro, lo studio mira a contribuire al dibattito accademico sull'impiego dell'AI Generativa nei processi aziendali, offrendo una testimonianza concreta del suo potenziale impatto sul miglioramento dell'efficienza operativa e sulla qualità dell'interazione con i clienti.

La metodologia adottata per questa ricerca è di tipo qualitativo, incentrata sull'analisi di un caso studio specifico. L'approccio è basato su un'intervista approfondita con il responsabile dell'innovazione e il Chief Digital Officer (CDO) di Eurotherm. Questo metodo è stato scelto per la sua efficacia nel fornire insight dettagliati e contestualizzati sulle esperienze dirette degli individui coinvolti nel processo di digitalizzazione e implementazione dell'AI all'interno dell'organizzazione.

L'intervista si è concentrata su diverse aree tematiche, tra cui:

- Fasi della Trasformazione AI: Descrizione sequenziale dell'introduzione e dell'evoluzione dell'AI Generativa nel customer service, evidenziando le tappe principali, le sfide incontrate e le soluzioni adottate.
- Aree di Impatto: Analisi degli effetti dell'AI Generativa sulle dinamiche interne del customer service, comprese le modifiche ai flussi di lavoro, l'efficienza operativa e la qualità dell'assistenza clienti.
- Risultati Ottenuti: Valutazione dell'impatto dell'AI Generativa sulle prestazioni complessive del customer service, con particolare attenzione ai benefici quantificabili come la riduzione dei tempi di risposta, l'aumento della soddisfazione dei clienti e l'ottimizzazione delle risorse.
- Percezioni e Feedback: Raccolta delle percezioni del personale e dei clienti riguardo l'efficacia e l'accettazione dell'AI Generativa nel contesto del servizio offerto, inclusi eventuali feedback negativi e le strategie adottate per affrontarli.

L'intervistato, Fabio Aprea (F.A.), attualmente Head of Digital Transformation e Chief Digital Officer di Eurotherm. F.A. rappresenta un esempio della mobilità interna e dell'evoluzione professionale all'interno del contesto aziendale, dove la capacità di adattamento e la continua formazione giocano un ruolo chiave. Iniziando la sua carriera

più di vent'anni fa all'interno dell'azienda Eurotherm, ha descritto in modo molto chiaro l'evoluzione tecnologica dell'azienda. Iniziando la sua carriera nell'ambito degli acquisti e della logistica, la sua transizione verso il ruolo di responsabile del controllo di gestione ha segnato un punto di svolta, permettendogli di approfondire le competenze analitiche e di interpretazione dei dati aziendali. Dal post Covid, con la promozione alla posizione di responsabile della digitalizzazione ha consolidato ulteriormente la sua centralità nel processo di trasformazione tecnologica e innovativa dell'azienda.

4.2. Evoluzione e Contesto Economico

4.2.1. Storia

Eurotherm è un Azienda che si distingue nel settore del riscaldamento e raffrescamento (HVAC), specializzata nell'offerta di soluzioni idroniche di riscaldamento e raffreddamento a pavimento. Eurotherm si è affermata nel panorama del riscaldamento e raffrescamento a pavimento a partire dal 1984, anno in cui Cleto Pezzeri, fondatore, intraprese il suo percorso imprenditoriale in Trentino-Alto Adige. La sua visione di valorizzare le potenzialità del riscaldamento radiante ha incontrato un mercato ricettivo e ha permesso all'azienda di crescere e specializzarsi nella fornitura di impianti idronici ad alta efficienza energetica. Con l'intento di posizionarsi come leader in Italia per il comfort climatico, Eurotherm ha sviluppato soluzioni volte a garantire un comfort abitativo ottimale attraverso la tecnologia radiante. All'inizio del nuovo millennio, la direzione aziendale passa nelle mani di Christian Pezzeri, sotto la cui guida Eurotherm si consolida e si espande a livello internazionale. L'azienda, che oggi conta oltre 100 collaboratori e si avvale di un sistema di qualità certificato, ha radicato la propria presenza in Italia e proseguito con determinazione il proprio percorso di internazionalizzazione. Con una presenza significativa nei principali paesi europei, Eurotherm vanta una vasta esperienza nella realizzazione di impianti sia in ambito nazionale che internazionale. Impegnata nella promozione del benessere abitativo e lavorativo, Eurotherm partecipa attivamente alla vita associativa del settore e collabora con istituzioni accademiche per fomentare la cultura del comfort radiante e contribuire allo sviluppo tecnologico del settore.

Nel 2010, Eurotherm e le sue business unit, Eureka ed Eurass, si sono integrate in E-Group, affiancandosi a Enetec, azienda del gruppo specializzata in ricerca e sviluppo, progettazione e produzione di componenti per sistemi radianti. La sinergia tra le aziende consente a E-Group di gestire l'intera filiera del riscaldamento e raffrescamento a pavimento, assicurando un'eccellenza operativa che si traduce, oltre alla creazione di un vantaggio competitivo non indifferente, nella qualità e affidabilità dei sistemi Eurotherm, Sito web Eurotherm, Info Eurotherm (2023).

4.2.2. Il gruppo “E-Group”

Eurotherm, azienda investigata per l'implementazione di Intelligenza Artificiale Generativa all'interno dell'organizzazione, è parte del gruppo “E-Group. E-Group è un gruppo industriale italiano leader nel settore del comfort climatico indoor, con un fatturato di oltre 80 milioni di euro distribuito su circa 2500 clienti distribuiti in 30 mercati differenti e una forza lavoro di più di 250 persone. Il gruppo opera attraverso due società controllate, Eurotherm, specializzata nella progettazione e commercializzazione di sistemi radianti e di sistemi di gestione del clima, ed Enetec, focalizzata sulla produzione di componenti per impianti radianti, E-group (2023).

Le attività del gruppo si possono riassumere in tre macro-aree:

- Ricerca e sviluppo: E-Group investe costantemente in ricerca e sviluppo per migliorare le proprie tecnologie e proporre soluzioni innovative e all'avanguardia nel settore del comfort climatico indoor.
- Produzione: Enetec, la società del gruppo dedicata alla produzione, produce una vasta gamma di componenti per impianti radianti, tra cui tubazioni, pannelli radianti e sistemi di regolazione.
- Commercializzazione e assistenza: Eurotherm si occupa della commercializzazione e assistenza dei sistemi radianti e dei sistemi di gestione del clima del gruppo E-Group. L'azienda opera attraverso una rete di vendita capillare in Italia e all'estero.

L'investimento costante in ricerca e sviluppo testimonia la volontà di E-Group di anticipare le esigenze del mercato e di rispondere con prodotti che non solo garantiscono il massimo benessere degli utenti finali ma sono anche rispettosi dell'ambiente. Questo approccio proattivo si riflette nella collaborazione con istituti di ricerca e università, allo

scopo di esplorare nuove tecnologie e materiali che possano contribuire a ridurre l'impatto ambientale degli impianti di climatizzazione, rendendoli al tempo stesso più efficienti ed efficaci.

La produzione di Enetec che si articola in due stabilimenti, uno in Italia ed uno in Germania, è esemplare di come l'innovazione tecnologica possa coniugarsi con l'attenzione alla qualità e alla sostenibilità. L'alta specializzazione raggiunta nella fabbricazione di componenti per sistemi radianti, unita a un rigido controllo delle fasi produttive, assicura soluzioni in grado di soddisfare le più stringenti normative in materia di efficienza energetica e impatto ambientale.

Eurotherm, dall'altra parte, si pone come punto di riferimento nella distribuzione di sistemi per il comfort climatico, offrendo una gamma completa di soluzioni per il riscaldamento e il raffrescamento degli ambienti. L'assistenza tecnica pre e post vendita, la formazione continua e il supporto all'installazione sono solo alcuni degli aspetti che evidenziano il desiderio dell'azienda di costruire una reputazione orientata alla massima soddisfazione e al raggiungimento di obiettivi comuni di benessere e sostenibilità.

L'orientamento strategico di E-Group si proietta verso il futuro con l'obiettivo di affermarsi come leader internazionale nel settore del comfort climatico indoor, attraverso soluzioni innovative che sappiano integrare efficienza energetica, sostenibilità e tecnologia avanzata. La visione dell'azienda, incentrata sulla creazione di un ambiente abitativo salubre e rispettoso dell'ecosistema, trova conferma nel suo impegno per lo sviluppo di progetti e iniziative che promuovano la consapevolezza e l'adozione di pratiche volte alla riduzione dell'impronta carbonica e al miglioramento della qualità della vita, E-group (2023).

4.2.3. Il Settore HVAC

Il mercato dei sistemi HVAC (Heating, Ventilation, and Air Conditioning) rappresenta un importante segmento industriale, caratterizzato da una crescente enfasi sull'efficienza energetica, l'innovazione tecnologica e la sostenibilità ambientale. In particolare, il contesto italiano riflette una tendenza globale verso soluzioni di climatizzazione più efficienti e rispettose dell'ambiente, spinta da normative rigorose e da una crescente consapevolezza ecologica dei consumatori.

L'industria HVAC in Italia, valutata in 4,5 miliardi di euro nel 2023, è proiettata a crescere con un tasso di crescita annuale del 5,8% nei prossimi 5 anni. Questo sviluppo è sostenuto da diversi fattori chiave, tra cui l'innovazione tecnologica, come l'integrazione dell' IoT (Internet of Things) e l'Intelligenza Artificiale nei sistemi di climatizzazione, che permettono una gestione più efficiente e personalizzata degli ambienti interni, e la spinta alla sostenibilità ambientale, HVAC Market Analysis Report, GVR (2023).

Inoltre, la legislazione italiana, come il Decreto Legislativo 192/2005 e il Decreto Legislativo 102/2014, mirando a promuovere l'efficienza energetica e l'uso di tecnologie sostenibili nei sistemi HVAC, hanno portato il settore a subire crescite economiche e tecnologiche importanti negli ultimi 10 anni. Queste normative, insieme agli incentivi governativi come il Super Bonus 110%, stimolano la domanda di soluzioni di raffreddamento e riscaldamento innovative e a basso impatto ambientale. Nonostante le prospettive positive, il settore HVAC italiano deve affrontare alcune sfide significative, tra cui la volatilità dei prezzi delle materie prime e una concorrenza sempre più intensa e capillarizzata, composta da piccole aziende che si allacciano a grandi produttori. La presenza di numerosi attori nel mercato, da aziende locali a multinazionali, crea un ambiente altamente competitivo che, considerando il contesto economico settoriale di forte spinta con l'entrata in vigore del Superbonus, sta spingendo le imprese a investire continuamente in ricerca e sviluppo per differenziarsi, HVAC Report, Fortune Business Inside (2022).

Le opportunità per il mercato HVAC in Italia includono il potenziale di crescita nel segmento residenziale, grazie al rinnovato interesse per la riqualificazione energetica degli edifici, e nel settore commerciale, dove la necessità di sistemi efficienti e tecnologicamente avanzati è in aumento. Inoltre, la tendenza verso soluzioni HVAC intelligenti e interconnesse apre nuove possibilità per lo sviluppo di servizi di manutenzione predittiva e di gestione ottimizzata degli impianti, contribuendo a una maggiore efficienza operativa e a riduzioni significative dei costi energetici. Il mercato HVAC in Italia si trova di fronte a un momento di trasformazione, guidato dall'innovazione tecnologica e dalle esigenze di sostenibilità.

4.3. Fasi della trasformazione AI

4.3.1. Cultura aziendale e Milestone tecnologici

Eurotherm, si da subito si è distinta per una propensione all'innovazione tecnologica proattiva, lo dimostra come già a metà degli anni 2000, procedette l'inserimento in azienda di un responsabile informatico, le cui mansioni e responsabilità gravavano completamente ad ambiti di infrastruttura informatica e tecnologica. Tale inserimento aziendale è un chiaro sintomo di come l'ascesa alla proprietà da parte di Christian Pezzeri abbia dato inizio alla macchina di innovazione che poi sarebbe diventata Eurotherm. Il punto di partenza, quindi, delle evoluzioni tecnologiche dell'organizzazione, partono dalla metà degli anni 2000. La figura di un responsabile dell'infrastruttura informatica ha permesso un pieno coinvolgimento nel panorama tecnologico dell'epoca, permettendo l'implementazione di nuovi sistemi. Dal 2011 in poi, le principali Milestone tecnologiche raggiunte dall'azienda sono state l'adozione del CRM, del MES per la produzione, e l'adozione di un nuovo ERP e la totale delegazione della gestione dei Server aziendali verso una Server Farm Esterna. La lungimiranza e il livello di cultura innovativa aziendale, però, si evincono dalla scelta strategica, poi trasformata in vantaggio competitivo, dell'adozione della Suite Microsoft 365, nel 2019, prima del Covid, trasformando il sistema aziendale in un sistema completamente su Cloud, come riferito dallo stesso F.A. *“L'azienda, con il titolare, è sempre stata pro-tecnologia. Siamo passati alla Suite 365 di Microsoft, in Cloud, prime del Covid, che è stata la scelta vincente, visto quello che è successo. Questa scelta ci ha permesso di essere operativi con il Covid dal giorno 2, lavorando da remoto”*. Un altro passaggio tecnologico chiave, come riferito dallo stesso F.A., è relativo all'infrastruttura server dell'organizzazione. Dal 2017 l'azienda decise di esternalizzare completamente la gestione dei Server, delegando la totalità della capacità informatica ad una Server Farm. Tale scelta, come sottolinea F.A., si è rivelata una scelta vincente, garantendo non solo un taglio di costi dato dal possesso del Hardware e dalla gestione dello stesso, considerando le risorse umane e le competenze da investire, ma aumentando esponenzialmente la flessibilità aziendale di adattarsi alle contrazioni o espansioni del mercato, o delle necessità operative aziendali, potendo modificare la capacità dei server tramite semplici richieste alla server farm e limitate variazioni di prezzo, sia in ottica espansiva che di contrazione. Riprendendo F.A. *“Anche questa super vincente come scelta perché ci ha permesso di essere molto*

flessibili”.

L'outsourcing delle competenze e delle risorse IT, come la decisione di delegare la gestione dei server a una server farm esterna, rientra in una strategia di business volta a ottimizzare le operazioni aziendali sfruttando la specializzazione e l'efficienza di fornitori esterni. Questo approccio consente all'azienda di concentrarsi sulle loro competenze chiave, riducendo al contempo i costi e aumentando la flessibilità operativa. L'adozione di soluzioni cloud, come Microsoft 365, ha ulteriormente trasformato il panorama IT, specialmente durante eventi imprevisti come la pandemia di COVID-19. Queste soluzioni cloud offrono la possibilità di lavorare in remoto, garantendo continuità operativa e accesso sicuro alle risorse aziendali, dimostrando come l'outsourcing e le tecnologie cloud siano diventati pilastri fondamentali nell'economia e nel management moderni, favorendo un ambiente di lavoro più agile e resiliente. L'adozione di strategie di outsourcing delle competenze IT e l'integrazione di una suite Microsoft 365 da parte di Eurotherm non solo segnala un impegno verso l'innovazione tecnologica ma anche evidenzia un approccio strategico mirato a consolidare un vantaggio competitivo significativo nel settore.

Queste decisioni riflettono una visione lungimirante dell'innovazione, non limitata al semplice raggiungimento di uno standard settoriale, ma orientata verso l'esplorazione di nuove opportunità di business e il miglioramento della resilienza organizzativa. In un'epoca caratterizzata da rapidi cambiamenti tecnologici e da un contesto globale incerto, accentuato dalla pandemia di COVID-19, Eurotherm ha dimostrato di interpretare l'innovazione tecnologica non solo come un requisito essenziale per rimanere competitivi ma come un'opportunità per distinguersi dai concorrenti. L'outsourcing di competenze specialistiche e l'adozione di soluzioni cloud avanzate come Microsoft 365 hanno permesso all'azienda di accelerare la digitalizzazione dei processi interni, migliorare la collaborazione a distanza e garantire una continuità operativa senza precedenti. Queste mosse strategiche hanno non solo ottimizzato le operazioni quotidiane ma hanno anche aperto nuove vie per l'agilità aziendale e l'innovazione di prodotto, sottolineando l'importanza di un ecosistema IT robusto e flessibile.

Eurotherm, attraverso queste scelte, sottolinea un'interpretazione dell'innovazione come leva per il vantaggio competitivo, riconoscendo che la capacità di adattarsi rapidamente e di adottare soluzioni tecnologiche avanzate è fondamentale per superare i concorrenti e soddisfare le crescenti aspettative dei clienti. In questo modo, l'azienda non solo si

posiziona come leader nel proprio settore per la capacità di innovare ma stabilisce anche nuovi standard per l'efficienza, la collaborazione e la gestione del cambiamento, testimoniando un approccio all'innovazione che va oltre la semplice conformità ai trend di mercato, per abbracciare pienamente le potenzialità offerte dalle nuove tecnologie.

4.3.2. Avvicinamento alla AI Generativa

L'avvicinamento di Eurotherm all'intelligenza artificiale riflette un percorso iniziato, come per molte aziende, dall'AI tradizionale. Tuttavia, ciò che distingue Eurotherm nel panorama industriale è la sua innata predisposizione all'innovazione e alla sperimentazione tecnologica. Sin dalle sue prime esplorazioni nell'ambito dell'AI, l'azienda ha scelto di sviluppare e implementare soluzioni interne, sfruttando le competenza e le capacità interne del proprio reparto IT per lo sviluppo di tutti gli Algoritmi necessari. Questa decisione ha permesso non solo di mantenere un alto grado di autonomia nella creazione di algoritmi su misura, ma anche di adattare con precisione le tecnologie alle specifiche esigenze aziendali, senza dipendere esclusivamente da soluzioni esterne o standardizzate. L'adozione della suite Microsoft 365 e l'utilizzo dei suoi strumenti avanzati per l'automazione dei processi evidenziano l'approccio proattivo di Eurotherm nel rimanere all'avanguardia nel campo tecnologico. L'avvicinamento verso l'AI generativa, rappresenta un ulteriore passo nell'evoluzione digitale dell'azienda. Questo movimento strategico non solo conferma la volontà di Eurotherm di sfruttare l'AI Generativo come leva per il vantaggio competitivo ma dimostra anche la capacità di adattarsi e innovare costantemente, mantenendo sempre un approccio orientato al futuro. La fondamenta costruita sull'AI tradizionale, con un focus sulla conoscenza interna e sullo sviluppo autonomo di soluzioni tecnologiche, pone le basi per esplorazioni più avanzate e potenzialmente rivoluzionarie nel campo dell'AI generativa.

L'irruzione di Chat GPT nel panorama tecnologico ha segnato un punto di svolta nella diffusione delle conoscenze relative all'intelligenza artificiale generativa, catalizzando l'attenzione globale e rendendo l'argomento ineludibile anche per il settore industriale. La sua capacità di generare contenuti informativi, spiegazioni e approfondimenti in maniera intuitiva e accessibile ha trasformato radicalmente il modo in cui le novità tecnologiche raggiungono un pubblico vasto, inclusi professionisti e appassionati al di fuori del campo strettamente tecnico. Per F.A., la risonanza di Chat GPT è stata una vera rivelazione; ha reso il flusso di informazioni non solo abbondante ma anche comprensibile, stimolando

una curiosità fino ad allora latente e trasformandola in un curioso desiderio di approfondimento. La facilità con cui ha potuto accedere a dettagli, esempi e casi d'uso di AI generativa attraverso questa piattaforma ha significativamente abbattuto le barriere all'apprendimento, invitandolo a esplorare con entusiasmo le potenzialità e le applicazioni pratiche di queste tecnologie emergenti.

Parallelamente, l'impulso dato dalle notizie e dagli articoli diffusi grazie a Chat GPT ha catturato anche l'interesse di Christian, che ha prontamente riconosciuto il valore strategico di rimanere al passo con i tempi in un'epoca caratterizzata da rapidi cambiamenti tecnologici. La decisione di inviare F.A. all'evento di Alberto Mattiello, noto per le sue competenze nel delineare i contorni del futuro digitale, è stata quindi non solo una risposta alla curiosità stimolata da Chat GPT ma anche un investimento consapevole verso l'innovazione aziendale.

In questo contesto, Chat GPT si è rivelato non semplicemente uno strumento di divulgazione, ma un vero e proprio catalizzatore di crescita personale e aziendale. Grazie alla spinta iniziale di Chat GPT, Eurotherm ha abbracciato un approccio proattivo nell'esplorazione dell'AI, vedendo in essa non solo una frontiera tecnologica da esplorare ma anche un formidabile strumento per rafforzare la propria competitività nel mercato.

4.3.3. Embrace Gen AI

L'approccio "Embrace Gen AI" di Eurotherm incarna una filosofia aziendale orientata all'innovazione e all'accettazione attiva delle nuove frontiere tecnologiche, in particolare dell'intelligenza artificiale generativa. Questa strategia si distingue per il suo impegno nella comprensione, sperimentazione e integrazione delle potenzialità dell'AI generativa nei processi aziendali, mirando non solo a mantenere la competitività ma anche a riconfigurare il panorama industriale. L'adozione proattiva di workshop come il "Hometa Talk" evidenzia la visione lungimirante di Eurotherm, che, attraverso l'istruzione e l'innovazione continua, aspira a trasformare i paradigmi operativi tradizionali, sfruttando l'AI per stimolare la creatività, ottimizzare le operazioni e sviluppare nuovi modelli di business.

Il workshop "Hometa Talk", rappresenta il momento cruciale nell'adozione e nell'integrazione dell'intelligenza artificiale generativa all'interno dell'ecosistema aziendale. Questa serie di incontri, articolati in quattro sessioni verticali dedicate all'AI,

ha segnato un passo decisivo verso la comprensione e l'applicazione pratica delle capacità rivoluzionarie offerte dalle tecnologie basate sull'intelligenza artificiale generativa.

L'obiettivo principale del workshop era duplice: da un lato, familiarizzare i partecipanti con i concetti fondamentali e le potenzialità dell'AI generativa, dall'altro, dimostrare concretamente come queste tecnologie possano essere sfruttate per ottimizzare i processi, innovare i servizi e creare valore aggiunto in ambiti diversi all'interno dell'azienda. Attraverso una serie di presentazioni interattive, casi studio approfonditi e sessioni di brainstorming, i partecipanti hanno avuto l'opportunità di esplorare le applicazioni dell'AI in scenari reali, valutandone l'impatto sulle strategie aziendali future. La decisione di Eurotherm e del Gruppo di impegnarsi in questi workshop verticali su AI nasce dalla convinzione che l'intelligenza artificiale rappresenti non solo una tendenza tecnologica emergente, ma un vero e proprio pilastro su cui costruire il futuro dell'azienda. L'approccio proattivo adottato del tipo "Embrace Gen AI" evidenzia la volontà dell'organizzazione di posizionarsi all'avanguardia nell'innovazione tecnologica, abbracciando l'AI generativa come strumento fondamentale per rimanere competitivi in un mercato in rapida evoluzione.

La forte volontà di Christian e della proprietà di svolgere tali workshop, trasuda l'identità e l'approccio verso l'innovazione che distingue l'azienda non solo all'interno del proprio settore, ma all'interno del contesto imprenditoriale Nazionale. Tale scelta, fa emergere la volontà di Eurotherm e del gruppo, di voler capire e utilizzare la nuova tecnologia di AI Generativa, trasmettendo un approccio "Embrace Gen AI".

"Embrace Gen AI" si basa su una comprensione approfondita e su un'accettazione entusiastica delle potenzialità offerte dall'AI generativa. Questo approccio implica un impegno costante nella ricerca e nello sviluppo, nella formazione del personale e nell'adozione di soluzioni AI che possono trasformare radicalmente i processi aziendali, dalla produzione al customer service, dalla gestione delle risorse umane all'innovazione di prodotto. I vantaggi di questo approccio sono molteplici. Innanzitutto, permette a Eurotherm di anticipare i cambiamenti del mercato e di adattarsi rapidamente alle nuove esigenze dei clienti, garantendo una competitività di lungo termine. Inoltre, l'integrazione dell'AI generativa nei processi aziendali ottimizza l'efficienza operativa, riduce i costi e favorisce la nascita di nuove idee e soluzioni creative, apportando un valore aggiunto inestimabile. La scelta di adottare un approccio "Embrace Gen AI" fa emergere la volontà

di Eurotherm di non solo comprendere e utilizzare la nuova tecnologia di AI Generativa, ma anche di assumere un ruolo attivo nella definizione del futuro tecnologico del settore. Questo posizionamento avanguardista distingue Eurotherm dalle altre aziende italiane, molte delle quali si avvicinano all'innovazione con maggiore cautela o seguendo modalità più tradizionali.

“Hometa Talk” diventa quindi l’input che permetterà a Eurotherm ed al gruppo di sviluppare 6 progetti pilota basati su Intelligenza Artificiale Generativa.

4.4. Aree della “Trasformazione Gen AI”

L'approccio di Eurotherm verso la Trasformazione Gen AI rappresenta un impegno all'avanguardia nell'esplorazione e nell'integrazione dell'intelligenza artificiale generativa nei loro processi operativi e strategici. Il workshop "Hometa Talk" su AI Generativa ha catalizzato questo slancio, trasformando l'entusiasmo e le idee emerse in progetti concreti. A partire dal Workshop, sull’onda non solo dell’entusiasmo creato, ma anche sulla consapevolezza del valore della tecnologia, si sono sviluppati diversi progetti pilota, 5 in totale, di cui uno, Chatbot Zeromax, attualmente operativo e 4 in fase di implementazione, inoltre tutta la produzione di immagini e di contenuti Marketing e pubblicitari è stata completamente sostituita con modelli generativi, come dice F.A. *“Il Marketing ha fatto suo l'utilizzo di AI per produrre i contenuti pubblicitari [...] è già diventata parte integrante del processo creativo”* .

Si deduce il grado di priorità associato all’implementazione di AI Generativa vedendo l’organizzazione avviare progetti pilota particolarmente innovativi nel 2023. Questo processo ha non solo aumentato l'entusiasmo e la consapevolezza tecnologica all'interno dell'azienda ma ha anche rafforzato una cultura aziendale già incline all'innovazione, spingendola verso un maggiore interesse per l'AI generativa.

Ognuno dei progetti partoriti dalla penetrazione dei workshop e dalla diffusione delle conoscenze in tema Gen AI impatta aree differenti all’interno ed all’esterno dei processi aziendale e mirano a risolvere problematiche esistenti, qualità dei lavori ma anche ad aumentare l’efficienza e la produttività della macchina aziendale.

4.4.1. Progetti pilota

L'implementazione della AI Generativa viene identificata in questi 6 progetti riassunti nella tabella sottostante:

PROGETTO	STATO	AREA	FUNZIONE	TECNOLOGIA
Generazione Immagini	Attivo	Marketing	Generazione immagini e contenuti grafici in sostituzione delle piattaforme classiche	AI Generativa TEXT to IMAGE: “MidJourney”
Traduzione Automatica	W.O.P.	Marketing - Amministrazione	Traduzione automatica video tutorial, pagine web e in futuro i documenti interni	AI Generativa TEXT to TEXT: “Lokalise”
Misurazione Cantieri	W.O.P.	Operation - Logistica	Automatizzazione creazione preventivo tramite misurazione automatica con cellulare svolta dal cliente di luoghi e cantieri su cui sviluppare installazione prodotti	LiDAR: “Magic Plan” AI Generativa IMAGE to TEXT
Chatbot Installatori	Beta	Operation - Formazione	Chatbot intelligente automatico per istruire installatori e collaboratori esterni sulle funzionalità e le tecniche dei prodotti e assisterli nella fase di installazione	AI Generativa Chatbot: “Mendable”

Chatbot post-vendita	Beta	Customer Service	Chatbot intelligente per l'assistenza clienti in fase di utilizzo del prodotto Eurotherm installato	AI Generativa Chatbot: "Mendable"
Chatbot Zeromax	Attivo	Customer Service - Sales	Chatbot intelligente che assiste i potenziali clienti nelle informazioni per la costruzione del preventivo e le funzionalità della linea Zeromax	AI Generativa Chatbot: "Mendable"

1. Generazione di Immagini e contenuti grafici per Marketing e pubblicità:

L'iniziativa pilota descritta introduce un approccio rivoluzionario alla creazione di contenuti visivi. Questo progetto esplora l'impiego strategico dell'Intelligenza Artificiale Generativa per rivitalizzare il processo di sviluppo di concept di design e contenuti grafici. La sua essenza risiede nella capacità di trasformare radicalmente la produzione creativa, accelerando significativamente il brainstorming e la prototipazione grazie alla generazione automatica di una pluralità di proposte visive e aumentando l'utilizzo del potenziale creativo delle risorse umane permettendo una soddisfazione delle richieste maggiore ed in tempi minori. Il fulcro dell'implementazione di AI Generativa a livello di processo all'interno del Marketing e della creazione di contenuti è guidato dal grado di efficienza ed aumento di produttività che porta all'intero reparto.

Gli aspetti principali di questo progetto sono vari e toccano diversi aspetti. La personalizzazione su vasta scala consente la realizzazione di contenuti altamente mirati per segmenti di pubblico definiti. Questo aspetto non solo potenzia l'efficacia delle campagne ma rappresenta anche un salto qualitativo nell'interazione tra marca e consumatore, offrendo esperienze visive su misura che rispondono in modo preciso alle esigenze e preferenze dei destinatari. L'adozione di questa tecnologia apre nuovi orizzonti

nell'esplorazione creativa, offrendo ai designer una palette di possibilità inedite. Gli algoritmi di Intelligenza Artificiale si rivelano preziosi alleati nel suggerire accostamenti di colori, layout e stili, ampliando il campo delle potenzialità creative ben oltre i confini tradizionalmente esplorati. Dal punto di vista economico, il progetto promette una significativa riduzione dei costi di produzione. La capacità dell'AI di elaborare autonomamente una varietà di soluzioni grafiche elimina la necessità di un impegno umano esteso, consentendo alle risorse creative di essere allocate in modo più efficiente. Inoltre, l'automazione garantita dall'Intelligenza Artificiale assicura una coerenza di brand senza precedenti, attraverso l'adesione rigorosa alle linee guida aziendali in ogni elemento visivo prodotto. Questa evoluzione tecnologica implica un ripensamento dei ruoli tradizionali nel settore del design. I professionisti del settore sono chiamati a reinterpretare la propria figura professionale, orientandosi sempre più verso un ruolo di supervisor capaci di valutare, selezionare e perfezionare i contenuti generati dall'AI per assicurare che rispecchino al meglio gli obiettivi comunicativi della marca. La capacità di rispondere con prontezza alle dinamiche di mercato rappresenta un ulteriore valore aggiunto. La velocità nella produzione di materiali grafici permette alle aziende di adeguarsi rapidamente a tendenze emergenti e occasioni attuali, mantenendo la propria comunicazione sempre pertinente.

2. Traduzione automatica dei contenuti verso l'esterno:

Il progetto in esame rappresenta un'iniziativa volta a superare le barriere linguistiche e culturali nell'ambito della comunicazione aziendale, con particolare attenzione alla traduzione di video tutorial, contenuti web e, in prospettiva arrivare a poter tradurre anche documenti interni e documenti legali, attraverso l'utilizzo dell'Intelligenza Artificiale Generativa. In un contesto post-pandemico, che ha visto l'accelerazione nella produzione di contenuti digitali quali webinar e visite virtuali in cantiere, l'esigenza di rendere tali materiali accessibili a un pubblico internazionale è diventata sempre più pressante. Originariamente venivano prodotti in italiano per rispondere alle necessità del mercato principale, ma questi contenuti richiedono ora un'adeguata traduzione per poter raggiungere efficacemente clienti e partner all'estero. L'obiettivo principale di questo progetto è dunque quello di implementare una soluzione tecnologica che consenta la traduzione e la localizzazione di contenuti multimediali e testuali in diverse lingue, garantendo al contempo un alto livello di precisione e naturalità. Si è quindi proceduto alla valutazione di strumenti di traduzione automatizzata, in grado di fornire una risposta

rapida e scalabile alle necessità di traduzione, pur incontrando sfide legate ai costi e alla necessità di adattamenti successivi per assicurare fedeltà e coerenza con il tono e lo stile dell'azienda. Tra le considerazioni cruciali del progetto vi sono diversi punti centrali. Il primo e più importante è legato al doppiaggio. La scelta tra l'aggiunta di sottotitoli tradotti o il doppiaggio dei video in lingue target è stata guidata da un bilanciamento tra costi, tempi di realizzazione e l'esperienza desiderata per l'utente finale. Mentre i sottotitoli rappresentano una soluzione più economica e rapida, il doppiaggio può significativamente migliorare l'engagement del pubblico non madrelingua. La localizzazione del Contenuto è un aspetto fondamentale del progetto, che va oltre la mera traduzione letterale, includendo l'adattamento di espressioni culturali, unità di misura e convenzioni locali, rendendo i contenuti pienamente risonanti con le diverse audience. La presenza di testi all'interno delle grafiche video richiede un intervento specifico per la traduzione e l'eventuale sostituzione delle immagini, con implicazioni sul fronte delle risorse grafiche necessarie.

Il progetto, quindi, non si limita alla semplice traduzione ma mira a una vera e propria trasformazione dei contenuti, per rendere l'azienda protagonista di un mercato globalizzato. L'adozione di tecnologie basate sull'Intelligenza Artificiale Generativa si configura come una strategia chiave per affrontare queste sfide, offrendo soluzioni innovative per una comunicazione efficace e inclusiva a livello internazionale.

3. Automatizzazione del processo di generazione del preventivo tramite misurazione a distanza dei cantieri:

Il secondo progetto pilota affrontato dall'azienda mira a rivoluzionare il processo di generazione del preventivo per un lavoro od un installazione di soluzioni raffreddanti riscaldante, essendo un processo che occupa tante risorse nonostante sia in una fase precedente ad un effettivo rapporto di lavoro. Nello specifico quando si parla di formulazione di preventivo nel settore in cui opera Eurotherm il principale collo di bottiglia è dato dalla definizione degli spazi di intervento in, ovvero i metri quadri in cui è necessario apportare le installazioni. Quindi l'azienda mira a eliminare questo ostacolo cercando di ottenere le misurazioni di cantieri e dell'interno delle case in ottica cost-saving. Tradizionalmente, la stima delle misure richiedeva o la visita fisica di un rappresentante aziendale oppure si basava sulle informazioni fornite dal cliente, spesso imprecise a causa di piantine obsolete o discrepanze con la realtà fisica degli spazi. Per

superare questi ostacoli sta vendendo valutata l'introduzione di uno strumento tecnologico avanzato capace di eseguire scansioni ambientali 3D e di fornire misurazioni accurate degli spazi in esame.

La tecnologia al centro di questo progetto si avvale di sistemi tecnologici di scansione all'avanguardia denominato LiDAR, ovvero radar ottico dall'inglese Light Detection and Ranging, che utilizza un fascio laser per la misurazione degli spazi. In senso stretto è un apparato che individua un bersaglio e ne determina la distanza per mezzo di un raggio di luce. Questa tecnologia attualmente è limitata a dispositivi specifici acquistabili ed ad alcuni telefoni cellulari di ultima generazione (iPhone 12 Pro e Pro Max, iPhone 13 Pro e Pro Max, iPhone 14 Pro e Pro Max). L'obiettivo di questo progetto pilota è l'automatizzazione di tutto il processo di misurazione delle aree di intervento, solitamente posizionata nella fase precedente alla stilatura del preventivo. Per ovviare a questo problema, Eurotherm sta testando alternative basate sull'utilizzo di Intelligenza Artificiale Generativa che sfruttano l'utilizzo di foto e video degli spazi, per determinare le misurazioni, utilizzando quindi di fotocamere standard, presenti in qualsiasi telefono cellulare. Questa tecnologia crea rappresentazioni tridimensionali dettagliate degli ambienti che poi si trasformano in misurazioni su cui è possibile costruire il preventivo. Questa soluzione tecnologica permette di ottenere dati misurati estremamente precisi, superando le limitazioni tipiche delle metodologie tradizionali che richiedono lo spostamento e l'impegno di personale, o limitazioni dovute dallo scarso tasso di diffusione della tecnologia meccanica necessaria, come il Lidar, e riducendo significativamente la necessità di sopralluoghi in loco. Uno degli aspetti più innovativi del progetto è la generazione automatica di preventivi. Attraverso l'uso di un'applicazione semplice e intuitiva, i clienti possono autonomamente effettuare le scansioni dei loro spazi. Queste informazioni vengono poi utilizzate per calcolare in tempo reale un preventivo dettagliato, offrendo un servizio di elevata precisione e personalizzazione, a tempo zero e costo zero. L'approccio proposto non solo migliora l'efficienza del processo di preventivazione ma apre anche la strada a una raccolta dati di qualità superiore, utile per analisi avanzate e per l'ottimizzazione dei progetti.

Questo progetto pilota rappresenta un passo significativo verso la digitalizzazione dei processi aziendali, con particolare attenzione alle esigenze di precisione, efficienza e soddisfazione del cliente. La riduzione della necessità di sopralluoghi fisici e la capacità di fornire preventivi accurati in tempo reale non solo ottimizzano le operazioni interne

ma migliorano anche l'esperienza complessiva del cliente, evidenziando l'impegno dell'azienda nell'adozione di soluzioni innovative per rispondere alle sfide del mercato contemporaneo.

4. Chatbot Guida per Installatori e Collaboratori Esterni

Il progetto pilota in questione punta a innovare il settore dell'assistenza tecnica attraverso lo sviluppo di un chatbot dedicato, progettato per supportare installatori e operatori nell'uso dei prodotti elettronici. Questi dispositivi, solitamente caratterizzati da un'interfaccia touch e da una necessità di installazione tecnica qualificata, rappresentano una sfida comune per gli elettricisti, che spesso si trovano a dover gestire installazioni complesse senza una guida dettagliata. La realizzazione di questo chatbot nasce dalla constatazione che le tradizionali modalità di assistenza, ovvero le telefonate i messaggi WhatsApp, email e persino alcuni tentativi di videochiamate con realtà aumentata, non sempre consentono una risoluzione efficace e tempestiva dei problemi riscontrati sul campo. Questo strumento di assistenza digitale è concepito per offrire una guida passo passo nella risoluzione di problemi comuni, dalla corretta connessione del dispositivo alla configurazione iniziale, garantendo così che tutte le operazioni siano eseguite correttamente al primo tentativo. Grazie alla sua capacità di consultare rapidamente i contenuti di diversi manuali tecnici, il chatbot permette agli operatori di accedere istantaneamente alle informazioni necessarie, ottimizzando i tempi di intervento e riducendo il margine di errore.

La prospettiva dell'azienda in questo progetto è orientata, con il solito spirito di innovazione che delinea Eurotherm ed il gruppo, alla futura integrazione di funzionalità di realtà aumentata, già sperimentate dall'azienda in precedenza con scarsi risultati. La potenziale integrazione di Realtà Aumentata o Realtà Mista e l'assistenza in tempo reale di un Chatbot onnisciente in materia di installazione prodotti Eurotherm, consentirebbe di fornire indicazioni visive precise durante le interazioni a distanza senza nessun bisogno di operatori terzi. Questo aspetto è particolarmente rilevante in situazioni in cui la comunicazione verbale non è sufficiente per guidare l'installatore nelle operazioni tecniche. Oltre a migliorare l'efficienza dell'assistenza tecnica, il chatbot è progettato per facilitare la condivisione delle conoscenze all'interno di un team di assistenza in crescita, assicurando che le informazioni critiche siano facilmente accessibili a tutti i membri del personale. Inoltre, il sistema contribuisce alla raccolta di dati sull'installazione e l'utilizzo

dei prodotti, migliorando la capacità dell'azienda di tracciare le prestazioni dei dispositivi e di intervenire prontamente in caso di problemi, il tutto nel rispetto della privacy degli utenti.

La realizzazione del chatbot mira infine a semplificare i processi di comunicazione interni ed esterni, riducendo la dipendenza da canali più tradizionali e soggetti a inefficienze. La visione di estendere l'uso del chatbot anche ai centri assistenza esterni apre la possibilità di una risposta rapida e accurata alle esigenze tecniche dei clienti, indipendentemente dalla loro ubicazione o dalla lingua parlata, garantendo un servizio assistenziale di alto livello e una customer experience ottimizzata.

5. Chatbot di Customer Service Post-vendita

Il progetto relativo al chatbot per l'assistenza clienti sui prodotti Eurotherm si configura come un'innovativa soluzione digitale volta a migliorare significativamente l'esperienza post-vendita degli utenti. Questo strumento è stato concepito con l'obiettivo di guidare il cliente finale nell'uso quotidiano dei prodotti, rispondendo in maniera immediata e intuitiva a quesiti riguardanti l'operatività dei dispositivi, come ad esempio la configurazione di un sistema di riscaldamento a pavimento smart. Differenziandosi per la sua focalizzazione sull'uso pratico dei prodotti dopo l'installazione, questo chatbot è una vera e propria guida interattiva, capace di fornire istruzioni dettagliate per l'impostazione delle funzionalità desiderate, ad esempio regolare la temperatura per specifiche fasce orarie o svolgere configurazioni tecniche più complesse. Ciò rappresenta una risorsa preziosa per l'utente, che spesso preferisce evitare la consultazione di manuali tecnici, percepiti come complessi e poco accessibili. L'idea è che con l'utilizzo del chatbot si possano andare ad eliminare completamente i manuali tecnici per iscritto, che rendono tedianti le consultazioni per trovare le informazioni d'interesse..

L'interfaccia utente del chatbot è sviluppata per essere quanto più amichevole e intuitiva possibile, incoraggiando così l'interazione diretta degli utenti con lo strumento, piuttosto che la ricerca di informazioni attraverso canali più tradizionali. Questo approccio non solo migliora l'esperienza complessiva del cliente ma contribuisce anche a personalizzare l'assistenza fornita, adattando le risposte alle specifiche esigenze e configurazioni dei prodotti in possesso degli utenti. Integrato direttamente nel sito web dell'azienda, il chatbot è facilmente accessibile a tutti i clienti che necessitano di supporto, offrendo una soluzione pratica e immediata per la risoluzione dei dubbi più comuni. Questa

caratteristica si traduce in una riduzione significativa del carico di lavoro per il supporto clienti umano, permettendo al personale di concentrarsi su problemi più complessi o richieste specifiche che richiedono un intervento diretto. Attraverso la raccolta di dati sulle domande più frequenti e sulle difficoltà incontrate dagli utenti, il chatbot offre inoltre preziosi spunti per il miglioramento continuo dei prodotti e delle relative guide utente. La possibilità di servire clienti in diverse lingue amplifica l'accessibilità e l'efficacia di questo strumento, rimuovendo barriere linguistiche e rendendo l'assistenza post-vendita un vero punto di forza per l'azienda nel suo impegno verso un'esperienza utente senza precedenti.

Tra i Sei progetti pilota partiti in azienda precedentemente descritti, eccetto l'implementazione di AI nella creazione di contenuti multimediali nell'area marketing, i restanti 4 sono ancora in fase di sviluppo e strutturazione. In particolare però sono in uno stato di sviluppo avanzato i due Chatbot, sia il chatbot guida per installatori e collaboratori esterni, sia il chatbot di Customer Service post-vendita per clienti, attualmente in una fase Beta, quindi a poche settimane o mesi dal lancio. Il sesto ed ultimo progetto pilota è in realtà l'unico progetto verso l'esterno che per ora è attualmente in funzione, esso è il Chatbot di Customer Service "Zeromax".

4.4.2. Chatbot di Customer Service Pre-vendita "Zeromax"

Il progetto pilota del chatbot Zeromax rappresenta un'innovazione significativa nel campo del Customer Service di Eurotherm, mirando a ottimizzare l'esperienza di pre-vendita per i clienti interessati alla linea di prodotti Zeromax. Questo strumento digitale, integrato nel portale Zeromax, è stato ideato per assistere i clienti in una fase cruciale del loro percorso d'acquisto, fornendo risposte immediate e personalizzate alle loro domande, anche in orari in cui il supporto tradizionale potrebbe non essere disponibile. La peculiarità di questo chatbot, oltre ad essere basato su un Foundation Model di AI Generativa, risiede nella sua capacità di offrire un servizio clienti efficiente e continuativo, 24 ore su 24, 7 giorni su 7, consentendo agli utenti di ottenere assistenza in qualsiasi momento, citando F.A. *"È interessante vedere come la presenza 24 su 24, 7 giorni su 7 sia accettata e anzi sfruttata dalle persone"*. Questa caratteristica è particolarmente apprezzata dai clienti, come dimostra l'interazione con il sistema in orari insoliti, inclusi i momenti festivi e notturni, evidenziando la crescente aspettativa di disponibilità costante del supporto clienti nell'era digitale, riprendendo F.A *"È interessante vedere come questi sistemi vanno a prendere*

l'attenzione delle persone in orari più assurdi [...] Addirittura uno all'una del primo di gennaio”.

Il chatbot Zeromax, inoltre, alleggerisce notevolmente il carico di lavoro del personale del servizio clienti, automatizzando le risposte alle domande più frequenti e consentendo agli operatori umani di dedicarsi a questioni più complesse o a casi che richiedono un intervento umano. Questo non solo migliora l'efficienza operativa ma contribuisce anche a ridurre i costi di gestione del servizio clienti, ottimizzando al contempo la soddisfazione del cliente grazie a una risoluzione rapida delle loro esigenze. Un aspetto fondamentale del chatbot è la sua funzione di guida ai prodotti e servizi offerti da Zeromax, facilitando la comprensione delle caratteristiche specifiche e aiutando i clienti a prendere decisioni informate. Attraverso un'interazione semplice e intuitiva, il chatbot è in grado di fornire dettagli su come utilizzare i prodotti, su come impostare determinate funzioni e su come ottenere il massimo dai servizi disponibili nella linea Zeromax, arricchendo così l'esperienza complessiva dell'utente. La raccolta di dati effettuata tramite le interazioni con il chatbot rappresenta un ulteriore vantaggio, offrendo all'azienda preziose informazioni sulle domande più frequenti, sulle esigenze degli utenti e sulle potenziali aree di miglioramento per prodotti e servizi. Questi dati possono essere analizzati per affinare ulteriormente le strategie di assistenza e di marketing, garantendo che l'offerta risponda sempre meglio alle aspettative dei clienti.

Il chatbot è stato introdotto come iniziativa pionieristica nell'applicazione dell'IA Generativa all'interno dell'azienda, scegliendo specificatamente il settore del Servizio Clienti Pre-vendita per la linea di prodotti Zeromax. Zeromax rappresenta una soluzione innovativa nel campo dei sistemi di riscaldamento e raffrescamento radiante a pavimento, progettata per offrire un elevato comfort abitativo, con una marcata attenzione verso l'efficienza energetica e la sostenibilità ambientale. Questo sistema si distingue per la sua capacità di garantire una distribuzione omogenea del calore o del fresco attraverso la superficie del pavimento, assicurando condizioni ottimali di benessere negli ambienti residenziali o commerciali.

La scelta di questa specifica area aziendale per l'introduzione del chatbot è stata motivata da considerazioni strategiche, tra cui:

- La necessità di limitare l'ambito di applicazione iniziale della nuova tecnologia a segmenti aziendali meno estesi, evitando di esporre l'intera organizzazione a

potenziali rischi derivanti dall'adozione di un sistema non ancora pienamente testato.

- L'incertezza relativa alle risorse necessarie per la manutenzione del chatbot suggeriva un approccio cauto, privilegiando un'implementazione graduale che partisse da linee di prodotto specifiche.
- La preesistenza di un portale per la generazione di preventivi online per la linea Zeromax, sviluppato nel 2022, rendeva il chatbot un'integrazione perfetta, mirando a ridurre ulteriormente la dipendenza da interventi umani in questo processo di vendita

Il percorso di sviluppo e integrazione di questa tecnologia innovativa si è articolato in diverse fasi operative, delineando un modello di approccio sistematico e strategico all'innovazione tecnologica.

La fase iniziale del progetto ha preso avvio con un Workshop AI Generativa, organizzato in collaborazione con Hometa talk, volto a esplorare le potenzialità dell'AI Generativa e a stimolare la formazione di idee innovative. Questo incontro ha permesso al team di Eurotherm di acquisire una comprensione approfondita delle applicazioni e delle implicazioni dell'intelligenza artificiale nel contesto del servizio clienti, ponendo le basi per le successive fasi di ricerca e valutazione. Successivamente, la Ricerca e Valutazione degli strumenti di chatbot disponibili sul mercato ha costituito un momento cruciale, orientato alla selezione di una soluzione tecnologica che rispondesse efficacemente alle necessità comunicative con i clienti, nonché alle specifiche tecniche richieste, quali l'integrazione con piattaforme esterne e la flessibilità nell'adozione di sorgenti testuali predefinite o personalizzabili. Questa fase ha visto La Selezione dello Strumento ha rappresentato il culmine di questo processo di indagine, con la scelta di un chatbot che offrisse funzionalità avanzate in termini di verifica delle domande degli utenti e integrazione con risorse esterne, rispondendo così in maniera ottimale alle esigenze del progetto Zeromax. La fase successiva ha visto il Test Pratico dello strumento selezionato, attraverso un approccio dinamico che ha favorito la rapida messa online del chatbot per una valutazione diretta delle sue prestazioni in condizioni reali. Questo test "hands-on" ha permesso di identificare eventuali aree di miglioramento, assicurando che lo strumento fosse perfettamente allineato con le aspettative del progetto. Prima del lancio ufficiale, è stata effettuata una Verifica Interna, durante la quale il personale addetto alle vendite ha simulato interazioni con il chatbot per testarne la capacità di gestire le richieste in modo

efficace. Questo step ha consentito di rifinire ulteriormente le funzionalità del chatbot, ottimizzando le risposte in base ai feedback ricevuti. Infine, il Lancio e l'Implementazione del chatbot sul sito di Zeromax hanno segnato la conclusione di questo percorso innovativo, con l'introduzione ufficiale dello strumento a fine Agosto. Questa ultima fase ha consolidato l'integrazione del chatbot nel sistema di assistenza al cliente, inaugurando una nuova era di interazione digitale all'interno di Eurotherm, caratterizzata da un'assistenza immediata, personalizzata e accessibile in ogni momento, grazie all'impiego strategico dell'intelligenza artificiale generativa.

Rivoluziona la tua esperienza con il nostro chatbot intelligente: risposte istantanee 24/7 a portata di clic.

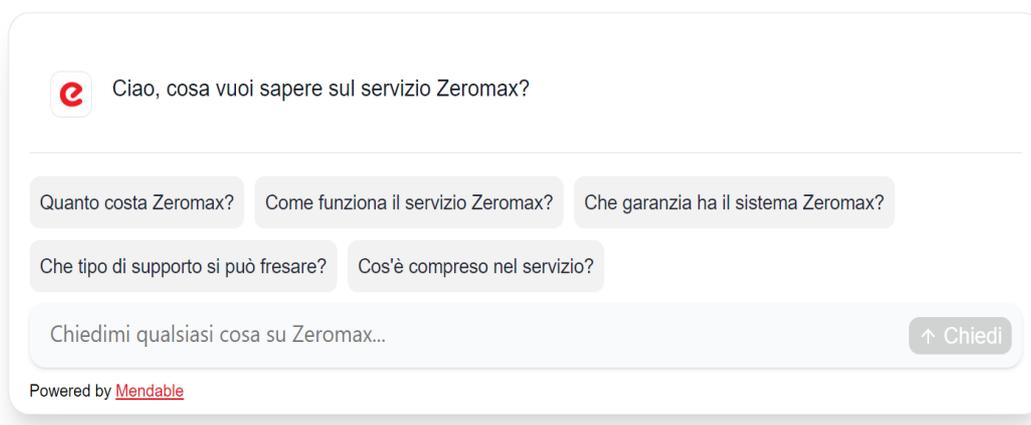


Figura 2: Chatbot "Zeromax", Eurotherm.it

Il Chatbot Zeromax, visibile in figura 2, strategicamente posizionato nella fase precedente alla costruzione del preventivo per la linea Zeromax nel sito web di Eurotherm, gioca un ruolo cruciale nell'ottimizzazione dell'esperienza utente e nell'efficienza del servizio clienti. Questo strumento digitale, collocato all'interno del portale che invita gli utenti a richiedere preventivi per i sistemi di riscaldamento e raffrescamento radiante Zeromax, funge da assistente virtuale avanzato, guidando i clienti attraverso le diverse fasi del processo informativo e decisionale. Il funzionamento tecnico-operativo del chatbot si articola in vari step operativi:

- Ricezione delle Query: Il chatbot inizia il suo lavoro raccogliendo le domande poste dagli utenti. Questo avviene tramite un'interfaccia di messaggistica intuitiva, situata nella sezione Zeromax del sito, che permette agli utenti di digitare liberamente le loro richieste.

- **Analisi e Interpretazione:** Una volta ricevuta una domanda, il sistema di intelligenza artificiale del chatbot procede con l'analisi del testo per interpretare correttamente l'intento dell'utente. Utilizza tecniche di elaborazione del linguaggio naturale (NLP) per decomporre la query in elementi comprensibili, identificando parole chiave e temi rilevanti.
- **Generazione della Risposta:** Sulla base dell'analisi effettuata, il chatbot seleziona o genera la risposta più adeguata alle domande poste. La capacità di generare risposte non si limita a un database statico di FAQ, ma include la possibilità di creare risposte personalizzate sfruttando vasti dataset di conoscenza interna ed esterna, integrati durante la fase di programmazione del chatbot.
- **Interazione Dinamica:** Il chatbot è progettato per gestire una conversazione fluida e dinamica, guidando l'utente verso le informazioni di cui ha bisogno o verso azioni specifiche, come la compilazione di un modulo per un preventivo. Può anche indirizzare l'utente verso risorse aggiuntive o consigliare specifiche pagine del sito per approfondimenti.
- **Feedback e Apprendimento:** Un aspetto fondamentale del funzionamento del chatbot Zeromax è la sua capacità di apprendere dalle interazioni. Grazie all'intelligenza artificiale generativa, il sistema analizza i feedback ricevuti per migliorare continuamente le proprie prestazioni, affinando la precisione delle risposte e la comprensione delle nuove query.

4.4.3. *Impatto Atteso vs Effettivo*

Le aree di applicazione primarie per i progetti di Intelligenza Artificiale Generativa avviati da Eurotherm si collocano sostanzialmente nei settori del Marketing e del Customer Service. Questa tendenza è in linea con le evidenze emerse dallo studio di McKinsey Sul potenziale economico delle AI Generative nelle aziende, che sottolinea come circa il 75% dei casi d'uso aziendali dell'AI generativa ottenga risultati significativamente positivi, in particolare nelle operazioni legate al supporto clienti, marketing e vendite, ingegneria del software, e ricerca e sviluppo, *The economic potential of generative AI (2023)*. Tale correlazione tra l'adozione dell'AI generativa e l'ottenimento di benefici tangibili in queste aree funzionali evidenzia il potenziale trasformativo di questa tecnologia nel rafforzare l'efficienza operativa e nel migliorare l'engagement dei clienti. Tra i vari progetti pilota avviati da Eurotherm per esplorare

l'applicazione delle intelligenze artificiali generative nel miglioramento delle operazioni aziendali e dell'interazione cliente, il Chatbot Zeromax rappresenta l'unico progetto che ha raggiunto la fase di completa implementazione e su cui, di conseguenza, si possono effettuare valutazioni concrete. L'unicità del suo dispiegamento fornisce una base solida per analizzare le prestazioni, l'efficacia e l'impatto generale sulle operazioni aziendali, distinguendolo dagli altri progetti pilota che sono ancora nelle fasi iniziali di valutazione o di sviluppo.

Analizzando il caso specifico del Chatbot "Zeromax", l'attenzione si concentra sul Customer Service nella fase di pre-vendita. Questo ambito si rivela particolarmente rilevante per diverse ragioni. Innanzitutto, il Customer Service pre-vendita funge da primo punto di contatto tra l'azienda e i potenziali clienti, giocando un ruolo cruciale nel formare la prima impressione dell'azienda e dei suoi prodotti. Attraverso interazioni tempestive e informative, si mira a rispondere efficacemente alle domande dei clienti, facilitando il processo decisionale e migliorando la percezione del brand. Inoltre, questa fase del servizio clienti permette di raccogliere dati preziosi sulle preferenze e sui bisogni dei consumatori, informazioni che possono essere utilizzate per ottimizzare l'offerta di prodotti e servizi e personalizzare le strategie di marketing.

L'obiettivo primario di questo progetto è di ottimizzare l'accessibilità delle informazioni per i potenziali clienti, fornendo risposte immediate e personalizzate alle loro domande, e di alleggerire il carico di lavoro del personale dedicato al servizio clienti, indirizzando in maniera autonoma le richieste più frequenti o di base. L'impatto atteso dall'introduzione del chatbot comprende un incremento significativo nella soddisfazione del cliente, una riduzione dei tempi di attesa per le risposte alle richieste di informazioni e un aumento dell'efficienza nelle operazioni di assistenza pre-vendita. Queste aspettative si basano sulla premessa che un supporto clienti efficace e tempestivo possa tradursi in un miglioramento dell'esperienza utente generale e in un aumento della propensione all'acquisto.

Strategicamente posizionato all'interno della pagina web dedicata alla linea di prodotti Zeromax, il chatbot si trova in un punto nevralgico del percorso del cliente, sulla pagina che invita gli utenti a procedere con la configurazione personalizzata del sistema desiderato. Questa scelta di design non è casuale ma mira a massimizzare l'efficacia dell'assistenza fornita nel momento più critico della customer journey, ovvero quando

L'utente mostra un interesse concreto verso il prodotto e necessita di ulteriori informazioni per procedere; “Tu cerchi sistema pavimento ristrutturazioni da Google, ti porta sul sito internet ti propone il sito zeromax che è specifico per fare preventivi dove ci sono delle informazioni e c'è anche il chatbot”. Il posizionamento del chatbot facilita un dialogo immediato e interattivo, consentendo agli utenti di chiarire dubbi o ottenere dettagli senza interruzioni nel loro percorso di scoperta e configurazione del prodotto. In questo modo, il chatbot diventa un elemento integrante dell'esperienza di acquisto, guidando il cliente attraverso le varie fasi del processo decisionale e contribuendo a creare un percorso fluido e informato verso la scelta finale del prodotto.

Le aspettative dell'azienda erano fin da subito mirate alla fascia di assistenza pre-vendita, con l'obiettivo principale di facilitare la creazione di preventivi ed invogliare i potenziali clienti a farsi fare il preventivo con la scusa di non dover avere a che fare con Call center o consulenze, F.A. riporta “volevamo andare lì, cioè intervenire sul canale prevendita, anche perché il chatbot è posizionato dove atterri per fare un preventivo, e lì siamo arrivati”.

L'aspetto più rivoluzionario e di impatto strategico risiede, oltre alla consulenza personalizzata, nella sua capacità di raccogliere dati in tempo reale sulle preferenze, comportamenti e interrogativi dei clienti. Questo approccio consente all'azienda non solo di comprendere in maniera approfondita ciò che "passa nella testa dei clienti", ma anche di anticipare le variazioni di domanda attraverso analisi predittive. La raccolta di dati attraverso interazioni guidate dal chatbot, infatti, apre la porta a un'analisi dettagliata dei bisogni emergenti, delle tendenze di consumo e delle potenziali aree di miglioramento dei prodotti o servizi offerti.

L'uso del Chatbot Zeromax per raccogliere dati sui clienti trasforma ogni interazione in un'opportunità di apprendimento per l'azienda. Dalle domande più frequenti alle richieste specifiche, ogni query racchiude preziose informazioni sulle aspettative e sulle esigenze dei clienti. Questo flusso costante di dati offre a Eurotherm una visione quasi in tempo reale delle dinamiche di mercato, consentendo un'analisi approfondita che va oltre la semplice transazione commerciale. Attraverso l'elaborazione e l'analisi dei dati raccolti, è possibile utilizzare modelli di analisi predittiva per identificare tendenze emergenti e potenziali cambiamenti nelle preferenze dei consumatori. L'abilità potenziale di poter anticipare queste variazioni consente a Eurotherm di adeguare in modo proattivo l'offerta

di prodotti, ottimizzare le strategie di marketing e, in ultima analisi, migliorare il posizionamento competitivo dell'azienda.

L'informazione raccolta tramite il Chatbot Zeromax gioca un ruolo fondamentale anche nella strategia di "risk reversal", ovvero nella riduzione del percepito rischio di acquisto da parte dei clienti. La conoscenza approfondita dei bisogni e delle perplessità dei consumatori permette di modellare offerte e garanzie che minimizzino le incertezze all'acquisto, migliorando la fiducia dei clienti nei confronti dell'azienda e dei suoi prodotti. Questo aspetto è particolarmente rilevante in un mercato competitivo, dove la capacità di distinguersi attraverso un forte orientamento al cliente può determinare il successo di un prodotto o servizio. In Conclusione, il Chatbot "Zeromax" ha aperto la strada a una serie di vantaggi aggiuntivi che hanno impattato l'azienda rispetto a quelli previsti. Vediamo ora di formulare un riassunto dei vantaggi attesi contro quelli Generati.

Impatto Atteso:

- Personalizzazione e Qualità delle Risposte Aumentata: Grazie alle capacità di apprendimento automatico e alla sofisticata elaborazione del linguaggio naturale, il Chatbot Zeromax è in grado di fornire risposte altamente personalizzate e di qualità superiore. Questa capacità di adattamento permette non solo di soddisfare le richieste specifiche dei clienti ma anche di offrire una comunicazione più umana e diretta, migliorando l'esperienza complessiva del cliente e incrementando la fiducia nei servizi dell'azienda.
- Disponibilità 24/7 per Richieste Specifiche: Il Chatbot Zeromax assicura una disponibilità ininterrotta, garantendo assistenza ai clienti in ogni momento del giorno e della notte. Questa caratteristica è particolarmente preziosa per rispondere a richieste specifiche che richiedono immediata attenzione, offrendo un livello di servizio clienti che va oltre le capacità operative umane, sia in termini di orario che di rapidità di risposta.
- Riduzione del Carico Umano su Customer Service: Automatizzando la gestione delle richieste più routinarie o frequenti, il chatbot alleggerisce significativamente il carico di lavoro sul personale dedicato al customer service. Questo consente di redistribuire le risorse umane verso attività che richiedono un maggiore intervento umano e una più profonda competenza tecnica o emotiva, ottimizzando l'uso delle competenze all'interno dell'organizzazione.

Riduzione dei Costi nel Customer Service: L'implementazione del Chatbot Zeromax contribuisce alla riduzione dei costi associati al customer service, minimizzando la necessità di un ampio staff dedicato al supporto clienti e riducendo i tempi di gestione delle richieste. Questa efficienza operativa si traduce in un significativo risparmio per l'azienda, permettendo di reinvestire le risorse in altre aree strategiche di sviluppo e innovazione.

Impatto Effettivo:

- **Raccolta Dati per Risk Reversal:** La raccolta di dati tramite il Chatbot Zeromax gioca un ruolo fondamentale nella strategia di risk reversal, un approccio volto a minimizzare le percezioni di rischio da parte dei clienti durante il processo di acquisto. Attraverso l'interazione diretta e personalizzata con i potenziali clienti, il chatbot raccoglie informazioni preziose sulle preoccupazioni, domande e obiezioni che questi ultimi possono avere riguardo ai prodotti o servizi offerti. Analizzando questi dati, Eurotherm è in grado di identificare modelli e tendenze comuni nelle preoccupazioni dei clienti, consentendo all'azienda di sviluppare strategie mirate per affrontare e mitigare tali rischi.
- **Raccolta Dati per Analisi Predittive:** La raccolta di dati operata dal Chatbot Zeromax si rivela essenziale per lo sviluppo di analisi predittive. Attraverso l'aggregazione e l'elaborazione dei dati raccolti dalle interazioni con i clienti, l'azienda può applicare tecniche di machine learning e modelli statistici per prevedere tendenze future del mercato, cambiamenti nelle preferenze dei consumatori o potenziali nuove aree di domanda. Queste analisi predittive permettono a Eurotherm di anticipare le mosse del mercato, adeguando proattivamente l'offerta di prodotti e servizi per soddisfare le esigenze future dei clienti prima che diventino evidenti.
- **Miglioramento Continuo del Servizio:** Uno dei vantaggi chiave del Chatbot Zeromax è la sua capacità di apprendimento continuo. Attraverso l'analisi delle interazioni con gli utenti, il chatbot si perfeziona costantemente, migliorando la qualità delle risposte e l'efficacia della comunicazione. Questo processo di apprendimento automatico assicura che il servizio offerto ai clienti sia sempre all'avanguardia, aumentando la soddisfazione e riducendo il tasso di errore nelle risposte fornite.

- Incremento dell'Engagement dei Clienti: Il Chatbot Zeromax, essendo disponibile 24/7, offre ai clienti un punto di contatto costante con l'azienda. Questa disponibilità continua incrementa l'engagement dei clienti, mantenendo aperto il dialogo con l'azienda e offrendo assistenza immediata in qualsiasi momento. Questo livello di interazione può migliorare significativamente la percezione del brand e rafforzare la lealtà dei clienti.
- Ampliamento dell'Accessibilità: Con la sua capacità di fornire assistenza in varie lingue e di adattarsi a diversi livelli di complessità delle richieste, il chatbot rende i servizi di Eurotherm accessibili a un pubblico più ampio. Questo ampliamento dell'accessibilità non solo apre nuovi mercati, ma assicura anche che tutti i clienti, indipendentemente dalla loro ubicazione geografica o capacità linguistica, possano ricevere supporto di alta qualità.
- Insights per lo Sviluppo di Prodotto: Le interazioni gestite dal Chatbot Zeromax forniscono una miniera di informazioni non solo sulle esigenze dei clienti ma anche sulle loro reazioni ai prodotti esistenti. Questi insights possono essere utilizzati per guidare lo sviluppo di nuovi prodotti o il miglioramento di quelli esistenti, assicurando che l'offerta di Eurotherm sia sempre in linea con le aspettative del mercato.
- L'impiego del Chatbot Zeromax nel pre-vendita non rappresenta solo un miglioramento nell'efficienza del servizio clienti, ma si configura come una strategia aziendale complessiva per comprendere meglio e anticipare le esigenze del mercato. La raccolta e l'analisi dei dati attraverso interazioni AI-enable diventano così un pilastro fondamentale per l'innovazione di prodotto, la personalizzazione del servizio e, in definitiva, per il mantenimento di un vantaggio strategico competitivo sostenibile nel tempo.

4.5. Risultati e Commenti finali

4.5.1. Risultati

L'implementazione del chatbot “Zeromax” è avvenuta a fine Agosto 2023. La fase di implementazione porta con sé un aspetto interessante. L'integrazione di chatbot basati su Intelligenza Artificiale Generativa, come il Chatbot Zeromax, nel proprio sito web

offrono una flessibilità e una facilità di integrazione notevoli, grazie all'uso di API (Application Programming Interfaces) che consentono una connessione semplice e diretta con le infrastrutture digitali esistenti. Il processo di fine tuning dei modelli di AI generativa, inclusi quelli che alimentano il Chatbot Zeromax, è altrettanto semplificato e accessibile. Le aziende hanno la possibilità di personalizzare il comportamento del chatbot integrando una propria libreria di documenti cartacei o digitali, come pagine web, manuali di prodotto o database di conoscenza. Attraverso tecniche di scraping e l'elaborazione del linguaggio naturale (NLP), il chatbot è in grado di analizzare questi documenti per estrarre le informazioni pertinenti su cui basare le sue risposte. Questo permette di adattare il chatbot alle specificità dell'offerta aziendale e alle domande più frequenti dei clienti, migliorando significativamente la rilevanza e l'accuratezza delle interazioni. L'implementazione, completata è avvenuta in un mese, con un impegno legato alla gestione ed alla manutenzione del contenuto e delle risposte vicino alle 2 ore al giorno, dalla terza settimana in poi l'impegno è passato a 2 ore a settimana, riflettendo un modello di efficienza operativa che ha ottimizzato le risorse senza compromettere la qualità del servizio.

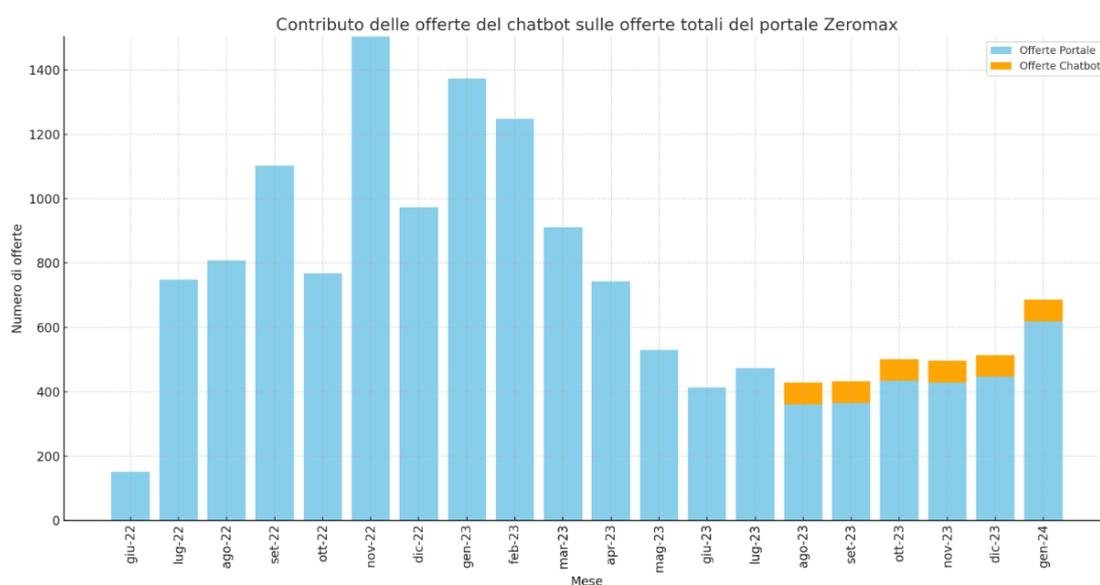


Figura 3: Richieste di preventivo portale Zeromax da Giugno 2022 a Gennaio 2024

L'analisi dei dati di performance del portale Zeromax, come si vede in figura 4, rivela un'interessante traiettoria di crescita e adattamento in risposta alle dinamiche del mercato. Lanciato a giugno 2022, il configuratore di preventivo nel portale Zeromax ha sfruttato con successo l'impulso dato dal trend delle ristrutturazioni e dai bonus statali, supportato

da un investimento strategico in pubblicità online e SEO. Dopo aver sperimentato picchi di richiesta nell'autunno 2022, si osserva un calo a partire da marzo 2023.

Dall'attivazione ad Agosto fino ai primi di Febbraio 2024, il chatbot ha gestito 4.614 messaggi, con una media di 769 messaggi al mese, dimostrando un'elevata capacità di coinvolgimento e interazione. Questa attività ha generato 403 richieste di preventivo, evidenziando un effettivo apporto del Chatbot alle richieste di preventivo in entrata. Sebbene il volume delle richieste abbia subito un calo dopo i picchi dell'autunno 2022, l'adozione di questa tecnologia AI ha effettivamente stabilizzato la tendenza, evitando un declino più marcato. Questo strumento di interazione automatizzato non solo ha rinvigorito il flusso di richieste, ma ha anche dimostrato la capacità di mantenere un livello costante di engagement, sottolineando l'importanza di integrare soluzioni innovative per adattarsi alle mutevoli condizioni del mercato. Dai dati forniti possiamo elaborare alcuni indici utili a capire meglio l'impatto del Chatbot nella creazione dei preventivi:

- Tasso di conversione: Il tasso di conversione del chatbot è calcolato dividendo il numero totale di messaggi scambiati tramite chatbot per il numero di richieste di preventivo generate attraverso il chatbot. La formula è:

$$\text{Tasso di conversione del chatbot} = \frac{\text{Nr. totale messaggi scambiati tramite chatbot}}{\text{Nr. richieste di preventivo tramite chatbot}}$$

Per il periodo specificato, da Agosto 2023 a Gennaio 2024 abbiamo un totale di 4614 messaggi scambiati e 403 richieste di preventivo, quindi in media, sono necessari circa 11,45 messaggi per convertire una conversazione iniziata in una richiesta di preventivo. Questo indica che, per ogni 11-12 interazioni con il chatbot, si genera una richiesta di preventivo, evidenziando un'efficacia relativamente alta del chatbot nel coinvolgere gli utenti fino a portarli a compiere un'azione concreta.

- Incidenza Chatbot sui preventivi totali: L'Incidenza del chatbot sulle richieste totali è il rapporto tra il numero di offerte generate dal chatbot e il numero totale di offerte. Questa proporzione si calcola come:

$$\text{Proporzione delle offerte del chatbot} = \frac{\text{Offerte generate dal chatbot}}{\text{Offerte totali}} \times 100\%$$

Applicando i dati raccolti nel periodo Agosto 23 – Gennaio 24, abbiamo identificato che 403 offerte sono state generate direttamente tramite interazioni con il chatbot su un totale di 3057 offerte complessive, risultando in una proporzione totale del 13.18%. Questo dato quantifica l'impatto diretto del chatbot sulle attività di vendita, dimostrando che più di un decimo delle offerte totali sono state facilitate dalla sua implementazione.

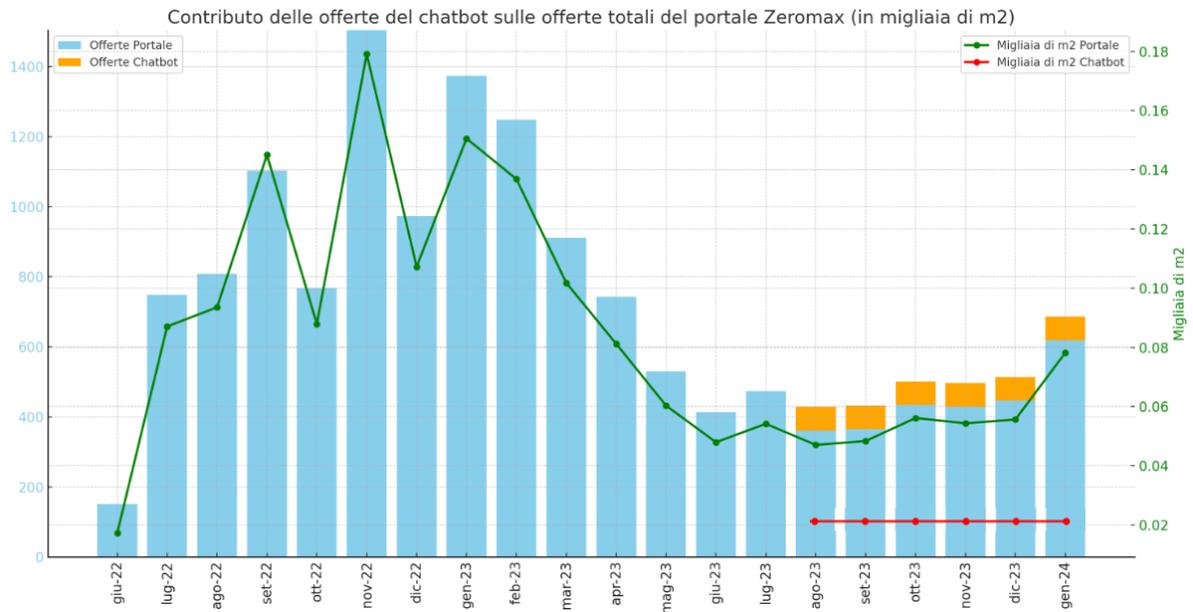


Figura 4: Richieste di preventivo e metratura portale Zeromax da Giugno 2022 a Gennaio 2024

La superficie complessiva preventivata attraverso i contatti generati dal chatbot ammonta a 47.372 metri quadrati, un indicatore dell'ampio volume di potenziali affari facilitati da questa soluzione tecnologica. L'investimento iniziale e i costi operativi del chatbot, descritti come non eccessivi e gestiti con un canone, si rivelano relativamente proporzionati rispetto ai benefici ottenuti.

- Media della superficie per richiesta di preventivo: La formula per calcolare la superficie media preventivata per ciascuna richiesta di preventivo attraverso l'interazione con il chatbot è la seguente

$$\text{Superficie media per richiesta di preventivo} = \frac{\text{Superficie totale preventivata dal chatbot}}{\text{Numero totale di richieste di preventivo generate dal chatbot}}$$

La superficie media preventivata per ciascuna richiesta è di circa 117,55 metri quadrati. Questo valore fornisce un'indicazione dell'ampiezza media dei progetti

per i quali gli utenti richiedono preventivi, offrendo un'idea del volume di affari potenziale generato attraverso l'interazione con il chatbot.

- Incidenza dal Chatbot sulla metratura totale: La proporzione della superficie preventivata dal chatbot è il rapporto tra la superficie totale preventivata dalle interazioni con il chatbot e la superficie totale preventivata in generale. La formula è

$$\text{Proporzione della superficie preventivata dal chatbot} = \left(\frac{\text{Superficie preventivata dal chatbot}}{\text{Superficie totali preventivate}} \right) \times 100\%$$

Analogamente all'incidenza sui preventivi, abbiamo determinato che il chatbot è stato responsabile per la generazione di preventivi che coprono 47.372 metri quadrati su un totale di 339.927 metri quadrati preventivati, corrispondente a una proporzione del 13.95%. Questo evidenzia l'efficacia del chatbot non solo nel generare un volume significativo di offerte ma anche nel contribuire a offerte di sostanziale valore e dimensione.

Questi risultati dimostrano non solo l'efficacia del chatbot Zeromax nel facilitare un'interazione efficace e tempestiva con i potenziali clienti ma anche il suo significativo contributo nell'aumentare il volume delle richieste di preventivo, ottimizzando le risorse aziendali e migliorando l'efficienza operativa. La capacità del chatbot di gestire un volume elevato di interazioni con una minima supervisione umana sottolinea il suo valore come strumento di automazione nel contesto aziendale di Eurotherm, offrendo spunti preziosi per future strategie di digitalizzazione e miglioramento del servizio clienti.

4.5.2. Sfide e Prospettive

L'azienda Eurotherm si trova al crocevia di un'innovazione tecnologica significativa con l'adozione di strumenti di intelligenza artificiale generativa che stanno trasformando il modo di lavorare quotidiano. Lo sviluppo di progetti pilota di AI Generativa e la prevista implementazione di CoPilot di Microsoft 365, una piattaforma che integra l'AI nei programmi Microsoft e che permette agli utenti di interagire in linguaggio naturale per eseguire compiti specifici come l'analisi di dati in Excel o la generazione di contenuti in Word e PowerPoint, testimonia l'impegno dell'azienda verso l'esplorazione e l'adozione di tecnologie di Intelligenza Artificiale generativa per migliorare l'efficienza operativa e la produttività interna.

La visione di Eurotherm guarda a un futuro in cui l'intelligenza artificiale è pienamente integrata nelle attività quotidiane di tutti i dipendenti, migliorando non solo le operazioni interne ma anche l'interazione con i clienti e la gestione dei servizi. L'obiettivo è sfruttare il potenziale dell'AI per mantenere un vantaggio competitivo nel mercato, anticipando le mosse dei concorrenti e rispondendo in modo più agile alle esigenze del cliente. Dai progetti pilota partiti, al futuro inserimento di CoPilot 365 si incarna l'ambizione di Eurotherm di adottare soluzioni che facilitino un ambiente di lavoro più efficiente e innovativo. L'azienda è consapevole che l'adozione di queste tecnologie non è fine a sé stessa ma rappresenta un mezzo per potenziare le capacità umane, accelerare la produzione di valore e migliorare la qualità del servizio offerto ai clienti. La visione di Eurotherm riflette un equilibrio tra innovazione tecnologica e miglioramento umano, con l'obiettivo di trasformare le sfide future in opportunità di crescita e sviluppo.

5. CONCLUSIONI

Nel contesto attuale del progresso tecnologico, l'Intelligenza Artificiale Generativa si posiziona all'avanguardia, delineando un confine evolutivo tra le sfide intrinseche e le opportunità emergenti. Questo ambito rivoluzionario della tecnologia, caratterizzato da una dinamica tensione tra limitazioni etiche, limitazioni tecniche e potenzialità innovative, offrendo un fertile terreno di indagine per la ricerca accademica avanzata.

Da una parte, le restrizioni tecnologiche, che includono una critica dipendenza dalla qualità e dalla quantità dei dati necessari per l'addestramento dei modelli e la problematica della scalabilità dei sistemi, costituiscono ostacoli non trascurabili. Tali limitazioni non solo incidono direttamente sulle prestazioni e sull'efficacia dei modelli generativi ma sollevano anche questioni di natura pratica riguardanti l'accessibilità e la sostenibilità su larga scala delle tecnologie AI. L'esigenza di ampi dataset di elevata qualità per un efficace addestramento dei modelli, unita alla necessità di infrastrutture computazionali avanzate per gestire una complessità in costante aumento, mette in luce la dipendenza da risorse frequentemente onerose e tecnicamente impegnative

D'altro canto, l'AI Generativa apre scenari precedentemente inimmaginabili, con la sua capacità di generare nuovi contenuti, dall'elaborazione creativa di testi alla produzione di immagini e musica, fino alla personalizzazione dei servizi in modo univoco per ciascun utente. Queste opportunità, senza precedenti, trovano applicazione in diversi settori, quali il marketing, l'intrattenimento, il Customer Service e molti altri, promettendo di trasformare radicalmente i processi produttivi e operativi attraverso un'automazione avanzata e una personalizzazione su misura, con conseguenti significativi incrementi dell'efficienza e riduzioni dei costi. Inoltre, l'impiego dell'AI Generativa nel campo della formazione e dello sviluppo professionale si prefigge di rivoluzionare le metodologie didattiche, rendendo l'apprendimento più interattivo, accessibile e su misura per l'individuo.

Nonostante le sfide esistenti, l'evoluzione dell'AI Generativa continua a estendere i limiti di ciò che è tecnologicamente realizzabile, portando con sé la promessa di reinventare le modalità di interazione tra macchine e umani, nonché le interazioni interpersonali. La capacità di navigare in questo panorama incerto, equilibrando le limitazioni tecniche con le potenzialità trasformative, giocherà un ruolo cruciale nel modellare il futuro delle AI

Generative e il loro impatto sulla società. Il percorso da intraprendere è intriso di complessità, presentando sfide tecniche, etiche e normative, ma è altresì caratterizzato dalle innumerevoli possibilità di progresso e innovazione intrinseche all'AI Generativa.

L'introduzione dell'Intelligenza Artificiale Generativa nel panorama imprenditoriale sta innescando una profonda trasformazione nel tessuto lavorativo delle organizzazioni moderne, segnando un punto di svolta nella concezione stessa del lavoro. Questo cambiamento non si limita semplicemente a un'evoluzione delle metodologie operative, ma rappresenta una vera e propria rivoluzione nel modo in cui viene concepita l'attività lavorativa. In un'era dominata dall'efficienza e dall'innovazione, la AI Generativa si presenta come il catalizzatore di una nuova era lavorativa, in cui il valore aggiunto dell'individuo si misura non solo in termini di ore lavorative prestata o do output prodotto, ma anche nella capacità di interagire in modo sinergetico con le tecnologie avanzate, integrandole nella routine per ottimizzare i processi e stimolare la crescita aziendale.

Nel settore del Customer Service, questo cambiamento si manifesta con un impatto ancora più marcato. L'inevitabile destino dei call center tradizionali e del servizio clienti umano sembra orientato verso l'obsolescenza, spinto dai costi elevati e dalle complessità logistiche legate alla gestione del personale umano ed alla loro gestione e formazione. La AI Generativa offre una soluzione a queste sfide, promettendo di rivoluzionare il modo in cui le aziende interagiscono con i loro clienti. Con chatbot e sistemi AI capaci di gestire un'ampia varietà di richieste clienti in modo puntuale, efficiente ed in tempo reale, il bisogno di intervento umano si riduce notevolmente. Tuttavia, è importante sottolineare che, nonostante i progressi dell'AI generativa nel Customer Service, il valore dell'elemento umano rimane insostituibile in determinate situazioni, specialmente quelle che richiedono un elevato livello di empatia, comprensione e giudizio critico. La sfida per le aziende, quindi, sarà quella di trovare il giusto equilibrio tra efficienza tecnologica e tocco umano, garantendo che l'introduzione dell'AI Generative non comprometta la qualità dell'interazione con i clienti, ma anzi la arricchisca e la personalizzi ulteriormente.

La capacità delle Generative AI di automatizzare processi e compiti precedentemente affidati all'intervento umano, invita quindi ad una riflessione sul ruolo del lavoro nell'identità professionale e personale degli individui. La tradizionale equazione che lega il lavoro al compimento di attività operative e ripetitive viene sostituita da un nuovo paradigma in cui la creatività, la strategia e l'innovazione diventano gli elementi centrali

della produttività. In questo contesto, il lavoratore assume il ruolo di supervisore e stratega, delegando all'AI compiti di natura operativa per concentrarsi su obiettivi di più alto livello, quali lo sviluppo di nuovi progetti, l'ottimizzazione dei flussi lavorativi e la creazione di valore aggiunto per l'azienda e per il cliente, il grado di diffusione di questa tecnologia è ambivalente, poiché la sua flessibilità non esclude un'adozione top-down come bottom-up, ridefinendo così il contesto lavorativo per come lo intendiamo, noi, oggi.

Questa ridefinizione del contesto lavorativo comporta importanti implicazioni sia a livello individuale che organizzativo. Per i dipendenti, l'acquisizione di competenze digitali avanzate e la capacità di collaborare efficacemente con le Generative AI diventano requisiti imprescindibili, sollecitando le aziende ad investire in programmi di formazione mirati. A livello organizzativo, si assiste a una ristrutturazione dei modelli di business, con una maggiore enfasi sulla flessibilità, sull'agilità operativa e sulla capacità di adattamento alle mutevoli esigenze del mercato.

In conclusione, l'introduzione della Generative AI nelle aziende non rappresenta soltanto un avanzamento tecnologico, ma segna una svolta nella concezione del lavoro, trasformandolo da attività prevalentemente esecutiva a processo intellettuale e creativo. In questo scenario, "lavorare" assume un significato più ampio, che abbraccia la gestione dell'innovazione, la leadership strategica e la collaborazione sinergica con le Intelligenze Artificiali Generative. Tale evoluzione richiede un approccio professionale più sofisticato, in cui la capacità di integrare le competenze umane con le potenzialità offerte dalla tecnologia diventa la chiave per generare successo e competitività nel panorama aziendale del futuro.

BIBLIOGRAFIA

- Alan Turing. (2004). *Computing Machinery and Intelligence* (1950).
- Alex Krizhevsky, Ilya Sutskever, & Geoffrey Hinton. (2012). *ImageNet Classification with Deep Convolutional Neural Networks*.
- Anantrasirichai, N., & Bull, D. (2020). *Artificial Intelligence in the Creative Industries: A Review*.
- Andrew L. Samuel. (1959). *Some Studies in Machine Learning Using the Game of Checkers*.
- Ben-Haim MS, Williams P, Howard Z, Mama Y, Eidels A, Algom D. (2016). *The Emotional Stroop Task: Assessing Cognitive Performance under Exposure to Emotional Content*.
- Bender, Emily & Gebru, Timnit & McMillan-Major, Angelina & Shmitchell, Shmargaret. (2021). *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?*.
- Bolton, R.N., McColl-Kennedy, J.R., Cheung, L., Gallan, A., Orsingher, C., Witell, L., and Zaki, M. (2018), "Customer experience challenges: bringing together digital, physical and social realms".
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., Hesse, C., . . . Amodei, D. (2020). *Language Models are Few-Shot Learners*.
- Caporusso, Nicholas & Zhang, Kelei & Carlson, Gordon & Jachetta, Daniel & Patchin, Devon & Romeiser, Spencer & Vaughn, Noah & Walters, Angela. (2020). *User Discrimination of Content Produced by Generative Adversarial Networks*.
- Costa, Victor & Lourenço, Nuno & Correia, João & Machado, Penousal. (2020). *Neuroevolution of Generative Adversarial Networks*.

D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. Chen, T. Lillicrap, F. Hui, L. Sifre, T. Graepel, & D. Hassabis. (2017). Mastering the game of Go without human knowledge.

Dean, Jeffrey & Corrado, G.s & Monga, Rajat & Chen, Kai & Devin, Matthieu & Le, Quoc & Mao, Mark & Ranzato, Aurelio & Senior, Andrew & Tucker, Paul & Yang, Ke & Ng, Andrew. (2012). Large Scale Distributed Deep Networks.

Donald Nguyen & Andrew Lenharth & Keshav Pingali. (2014). Deterministic Galois: On-demand, Portable and Parameterless.

Dresler T, Mériau K, Heekeren HR, van der Meer E. (2009). Emotional Stroop task: effect of word arousal and subject anxiety on emotional interference.

Frederick P. Brooks, Jr. (1995). The Mythical Man-Month: Essays on Software Engineering.

Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014) Generative Adversarial Nets.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning.

Hastie, Trevor & Tibshirani, Robert & Friedman, Jerome. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction.

Heinonen, Kristina & Michelsson, Thomas. (2010). The use of digital channels to create customer relationships.

Homburg, C., Jozić, D., & Kuehnl, C. (2017). Customer Experience Management: Toward Implementing an Evolving Marketing Concept.

Homburg, Christian & Jozić, Danijel & Kuehnl, Christina. (2015). Customer experience management: Toward implementing an evolving marketing concept.

Howard, J., & Ruder, S. (2018). Universal Language Model Fine-tuning for Text Classification.

Jain, Varsha & Sheth, Jagdish & Mogaji, Emmanuel & Ambika, Anupama. (2023). Artificial Intelligence in Customer Service: An Introduction to the Next Frontier to Personalized Engagement.

Jordan, M. I., & Mitchell, T. M. (2015). *Machine learning: Trends, perspectives, and prospects*.

J. McCarthy, M.L. Minsky, N. Rochester, and C.E. Shannon. (2006) "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955".

Klopfenstein, Lorenz & Delpriori, Saverio & Malatini, Silvia & Bogliolo, Alessandro. (2017). *The Rise of Bots: A Survey of Conversational Interfaces, Patterns, and Paradigms*.

Kumar, Anuj & Telang, Rahul. (2011). *Product Customization and Customer Service Costs: An Empirical Analysis*.

Kumar, V. & Reinartz, Werner. (2016). *Creating Enduring Customer Value*.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). *Deep learning*.

Lemon, Katherine & Verhoef, Peter. (2016). *Understanding Customer Experience Throughout the Customer Journey*.

Mctear, Michael & Callejas, Zoraida & Griol, David. (2016). *The Conversational Interface*.

Nie, D., Trullo, R., Petitjean, C., Ruan, S., & Shen, D. (2017). "Medical Image Synthesis with Context-Aware Generative Adversarial Networks."

Payne, Adrian & Frow, Pennie. (2005). *A Strategic Framework for Customer Relationship Management*.

Plekhanov, D., Franke, H., & Netland, T. H. (2023). *Digital transformation: A review and research agenda*.

Radford, Alec, Jeff Wu, Rewon Child, David Luan, Dario Amodei and Ilya Sutskever. "Language Models are Unsupervised Multitask Learners." (2019).

Russell, S. & Norvig, P. (2016). *Artificial Intelligence: A Modern Approach*.

Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., & Chen, X. (2016). *Improved Techniques for Training GANs*.

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). *Sequence to Sequence Learning with Neural Networks*.

Thorat, Sandeep A. and Jadhav, Vishakha, A Review on Implementation Issues of Rule-based Chatbot Systems (April 2, 2020).

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention Is All You Need.

Valdez-de-Leon, Omar. (2016). A Digital Maturity Model for Telecommunications Service Providers.

Warner, Karl & Wäger, Maximilian. (2019). Building dynamic capabilities for digital transformation: An ongoing process of strategic renewal.

Webb, T., Holyoak, K. J., & Lu, H. (2022). Emergent Analogical Reasoning in Large Language Models.

Zeithaml, V.A., Parasuraman, A. and Berry, L.L. (1990) Delivering Quality Service: Balancing Customer Perceptions and Expectations.

Gomber, Peter & Koch, Jascha-Alexander & Siering, Michael. (2017). Digital Finance and FinTech: current research and future research directions.

Hoy, Matthew. (2018). Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants.

Jurafsky, Daniel & Martin, James. (2008). Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition.

Kingma, D. P., & Welling, M. (2013). Auto-Encoding Variational Bayes.

Lin, Chien-Chang & Huang, Anna & Yang, Stephen. (2023). A Review of AI-Driven Conversational Chatbots Implementation Methodologies and Challenges (1999–2022).

SITOGRAFIA

Excitements and Concerns in the Post-ChatGPT Era: Deciphering Public Perception of AI through Social Media Analysis, Jinsheng Pan, Hanjia Lyu, Jiebo Luo, Department of Computer Science, University of Rochester, 2023. <https://arxiv.labs.arxiv.org/html/2307.05809>

OpenAI, GPT-4 Technical Report, 2023. <https://cdn.openai.com/papers/gpt-4.pdf>

ELIZA Chatbot, Wikipedia. [https://it.wikipedia.org/wiki/ELIZA_\(chatterbot\)](https://it.wikipedia.org/wiki/ELIZA_(chatterbot))

The Evolution of Chatbots: From Customer Service to Sales and Marketing, Tantibts, 2023, <https://www.linkedin.com/pulse/evolution-chatbots-from-customer-service-sales-marketing-tanbits/>

The evolution of customer service technology, whoson, 2022. <https://www.whoson.com/customer-service/the-evolution-of-customer-service-technology>

Exploding Topics, ChatGPT Users. <https://explodingtopics.com/blog/chatgpt-users>

The economic potential of generative AI: The next productivity frontier, 2023, McKinsey e Company, 2023. <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier>

APPENDICE

Lista Domande Intervista Caso studio Eurotherm. Intervistato Fabio Aprea:

- Qual è il tuo ruolo in azienda e quali sono stati i principali milestones a tema tecnologia ed innovazione da quando sei in azienda?
- C'è stato un avvicinamento particolare alla AI Generativa?
- Cos'ha portato l'azienda ad adottare un approccio "Embrace Generative AI" ?
- Hai notato cambiamenti a livello di cultura aziendale pre e post Generative AI ?
Se no, ci sono almeno stati dei cambiamenti nei comportamenti?
- Come sta venendo implementata l'AI Generativa in azienda?
- Quali sono i progetti pilota di AI generativa attualmente in corso?
- Quali sono le aspettative sui progetti di AI generativa? Si stanno effettivamente manifestando?
- E' stato complesso sviluppare ed implementare i progetti di AI Generativa?
- Quali risultati avete raccolto dall'implementazione del Chatbot "Zeromax"?
- Che impatto sta avendo il chatbot? corrisponde a quelli previsti?
- Come ti aspetti sarà il futuro dell'AI Generativa in azienda ma anche in generale?