

UNIVERSITÀ POLITECNICA DELLE MARCHE
FACOLTÀ DI INGEGNERIA

Dipartimento di Ingegneria dell'Informazione
Corso di Laurea in Ingegneria Informatica e dell'Automazione



TESI DI LAUREA

**Esperienze nel contesto dell'Intelligenza Artificiale: riconoscimento
di immagini, trascrizione speech-to-text e text-to-speech**

**Experiences in the context of Artificial Intelligence: image
recognition, speech-to-text and text-to-speech transcription**

Relatore

Prof. Domenico Ursino

Candidata

Tosca Pierro

ANNO ACCADEMICO 2022-2023

*Non è fantastica l'arte della deduzione?
Abbiamo individuato tutti i termini del problema,
tranne la soluzione*

Isaac Asimov, Io robot

Sommario

L'Intelligenza Artificiale (IA) ha compiuto passi da gigante negli ultimi decenni, trasformando il panorama tecnologico e ridefinendo i limiti delle capacità delle macchine. Questa crescita esponenziale ha innescato una rivoluzione in svariati settori, sfidando le prospettive convenzionali sulla tecnologia e aprendo la strada a un futuro di opportunità e innovazioni senza precedenti. Questa tesi si propone di esplorare in dettaglio l'IA, partendo dalle sue radici storiche fino alla sua evoluzione attuale. Si esamineranno, anche, le sfide etiche e le implicazioni sociali legate all'adozione dell'IA. Questo contesto critico offre una visione completa delle complessità che accompagnano lo sviluppo e l'implementazione dell'IA nella nostra società. Verranno, anche, analizzate in prima persona delle tecnologie dell'Intelligenza Artificiale e verranno svolti dei test.

Keyword: Intelligenza Artificiale, Machine Learning, Deep Learning, Natural Language Processing, Cloud computing, Reti neurali, Visione artificiale, Speech-to-text, Text-to-speech

Introduzione	1
1 Introduzione all'Intelligenza Artificiale	3
1.1 Definizione di Intelligenza Artificiale	3
1.1.1 Breve storia dell'Intelligenza Artificiale	4
1.2 Apprendimento automatico (Machine Learning)	5
1.2.1 Reti neurali artificiali	6
1.2.2 I modelli di apprendimento delle reti neurali	7
1.2.3 Applicazioni delle reti neurali artificiali	8
1.2.4 Algoritmi genetici	8
1.2.5 Logica fuzzy	10
1.3 Natural Language Processing	10
1.3.1 Come funziona	10
1.3.2 Ambiti applicativi	11
1.4 Implicazioni etiche dell'Intelligenza Artificiale	11
2 Amazon AWS	13
2.1 Cos'è Amazon Web Services	13
2.1.1 AWS e il Cloud	14
2.1.2 Tipi di cloud computing	14
2.1.3 Servizi Cloud	15
2.2 Servizi di AWS	16
2.2.1 Machine Learning	16
2.2.2 Amazon EC2	17
2.2.3 Amazon S3	17
2.2.4 Amazon DynamoDB	17
2.2.5 Lambda	17
2.3 Sicurezza	18
2.3.1 Calcolo crittografico in AWS	18
2.3.2 Protezione dei dati	18
3 Riconoscimento di immagini	19
3.1 Cos'è Rekognition	19
3.1.1 La visione artificiale	20
3.1.2 Servizi di Amazon Rekognition	20

3.2	Casi d'uso usati	21
3.2.1	Analisi delle etichette	21
3.2.2	Confronto facciale	23
3.2.3	Analisi facciale	24
4	Trascrizione speech-to-text	25
4.1	Cos'è Transcribe	25
4.1.1	Come funziona	25
4.2	Tipi di analisi	26
4.2.1	Amazon Transcribe Toxicity Detection	26
4.2.2	Amazon Transcribe Medical	27
4.3	Test svolti	27
4.3.1	Programma televisivo	27
4.3.2	Trascrizione a più lingue	28
4.3.3	Trascrizione dei dialetti	28
5	Trascrizione text-to-speech	30
5.1	Cos'è Polly	30
5.1.1	Funzionalità	30
5.1.2	Tipi di analisi	31
5.2	Test di trascrizione text-to-speech	31
5.2.1	Utilizzo del tag <lang>	32
5.2.2	Utilizzo tag <say-as>	32
6	Discussione	35
6.1	Discussione sul riconoscimento delle immagini	35
6.2	Discussione sulla trascrizione del parlato	36
6.3	Discussione sulla riproduzione vocale di un testo	36
	Conclusioni	37
	Bibliografia	38
	Ringraziamenti	40

Elenco delle figure

1.1	Modalità di funzionamento del Machine Learning	5
1.2	Esempio del funzionamento delle reti neurali	6
1.3	Esempio di un algoritmo generico	9
1.4	Logica Fuzzy: esempio del bicchiere	10
1.5	L'etica	11
2.1	I servizi offerti da AWS	13
2.2	Servizi di Cloud Computing	15
3.1	Immagine illusoria data in input ad Amazon Rekognition	21
3.2	Risultati forniti da Amazon Rekognition avendo in input l'immagine della Figura 3.1	21
3.3	Impressione calar del sole immagine data in input ad Amazon Rekognition	22
3.4	Risultati avendo in input l'immagine della Figura 3.3	22
3.5	Immagine delle Gemelle Olsen date come input ad Amazon Rekognition per il confronto facciale	23
3.6	Immagini della stessa persona, con trucco e senza, data in input ad Amanon Rekognition per il confronto facciale	23
3.7	Risultato dell'analisi facciale fornito da Amazon Rekognition avendo in input la persona truccata della figura 3.6	23
3.8	Risultato dell'analisi facciale fornito da Amazon Rekognition avendo in input la persona senza trucco della figura 3.6	24
3.9	Risultati dell'analisi facciale dell'immagine data in input di Albert Einstein	24
4.1	Parte del programma di Rete4 in cui vediamo i sottotitoli separati in base agli interlocutori	27
4.2	Esempio di trascrizione in tempo reale di due lingue	28
4.3	Spezzone del video in cui avviene lo scambio di lingue	28
4.4	Trascrizione della Divina Commedia	29
4.5	Parte della Divina Commedia originale	29

L'Intelligenza Artificiale (IA) è un ramo dell'informatica che si propone di dotare le macchine di capacità intellettuali, rendendole, così, in grado di apprendere, ragionare e adattarsi autonomamente. Questa disciplina ha attraversato un percorso di evoluzione straordinario, passando da un ambito di studio teorico a una realtà che occupa ogni aspetto della nostra vita quotidiana.

L'origine dell'IA può essere fatta risalire ai primi esperimenti concettuali nel campo dell'automazione e della teoria dei giochi del XX secolo. Tuttavia, è negli ultimi decenni che l'IA ha raggiunto il culmine della sua influenza, grazie a progressi significativi in campi chiave come l'apprendimento automatico, le reti neurali e l'elaborazione del linguaggio naturale.

L'evoluzione dell'IA ha seguito un percorso in cui le macchine sono passate da eseguire compiti specifici a dimostrare capacità di apprendimento e adattamento intelligente. Dai sistemi basati su regole rigide alle sofisticate reti neurali artificiali, ispirate al funzionamento del cervello umano, l'IA ha abbracciato un paradigma di apprendimento più autonomo e flessibile.

Questa evoluzione ha trovato applicazioni in una vasta gamma di settori, come la medicina, l'automazione industriale, la finanza e la guida autonoma. Oggi, ci troviamo di fronte a un panorama in cui l'IA è parte integrante della nostra vita quotidiana, dalle raccomandazioni personalizzate sui social media alle risposte intuitive degli assistenti vocali.

Attraverso questa tesi vedremo alcune tappe dell'evoluzione dell'IA esaminando le principali componenti dell'Intelligenza Artificiale. Infatti, dalle reti neurali al Machine Learning, analizzeremo anche le sfide etiche e sociali e le opportunità emergenti da questo rapido progresso tecnologico. Inoltre, affronteremo il modo in cui l'IA sta trasformando il nostro mondo cercando di bilanciare l'entusiasmo dell'innovazione con una consapevolezza critica delle sue implicazioni.

Nel particolare, parleremo anche della piattaforma di Intelligenza Artificiale Amazon Web Service (AWS) e dei servizi da essa offerti. Il cloud computing, di cui AWS è un esempio chiave, ha trasformato il modo in cui le risorse informatiche vengono fornite e consumate.

Fornendo accesso a server, storage, database e altro attraverso Internet, il cloud computing consente alle organizzazioni di ridurre i costi, aumentare l'agilità e concentrarsi sulle proprie attività principali anziché sulla gestione dell'infrastruttura.

In conclusione, testeremo in prima persona questa tecnologia e successivamente analizzeremo i risultati ottenuti.

I tool di AWS che analizzeremo saranno i seguenti:

- Rekognition;

- Transcribe;
- Polly.

Rekognition è un servizio di riconoscimento di immagini. Noi lo testeremo per osservare la tipologia delle sue risposte.

Transcribe è un servizio di riconoscimento vocale automatico. La trascrizione automatica può essere utile in molte situazioni, dalla creazione di sottotitoli per video alla registrazione di incontri e interviste.

Polly è un servizio che permette la sintesi vocale, con una trascrizione del tipo text-to-speech. Testeremo tale tool per vedere come l'uso di tag particolari (attraverso il linguaggio SSML) può influenzare la pronuncia e garantire in questo modo una maggiore flessibilità nella generazione della voce.

Riassumendo il tutto, la presente tesi è composta da sette capitoli strutturati come di seguito specificati:

- Nel Capitolo 1 verrà introdotto cos'è l'IA e la sua evoluzione. Verrà descritto cos'è una rete neurale e le sue varie tipologie ed, infine, si discuterà anche di che cos'è il Natural Language Processing.
- Nel Capitolo 2 verrà spiegato cos'è il cloud e, in particolare, il cloud Amazon Web Service. Verranno spiegati brevemente quali sono i servizi di AWS e che tipo di sicurezza può garantire.
- Nel Capitolo 3 verrà utilizzato uno strumento di Machine Learning per il riconoscimento delle immagini, ovvero Amazon Rekognition, e verranno eseguiti dei test con esso.
- Nel Capitolo 4 sarà utilizzato un altro strumento di Machine Learning, chiamato Amazon Transcribe, per la trascrizione di speech-to-text e di sottotitoli sia di interi video che di dirette streaming.
- Nel Capitolo 5 sarà utilizzato uno strumento per la trascrizione text-to-speech.
- Nel Capitolo 6 verranno riportati i risultati ottenuti dai test effettuati nei capitoli precedenti.
- Nelle Conclusioni saranno tratte tutte le osservazioni finali relative alla ricerca e verranno delineati alcuni possibili sviluppi futuri.

Introduzione all'Intelligenza Artificiale

In questo capitolo discuteremo di cos'è l'Intelligenza Artificiale, della sua evoluzione e dell'etica che ne consegue, parlando del suo sviluppo attraverso le reti neurali e i suoi vari metodi di apprendimento. Parleremo anche dell'elaborazione del linguaggio naturale e degli ambiti in cui è applicata.

1.1 Definizione di Intelligenza Artificiale

L'intelligenza rappresenta un complesso di facoltà psichiche e mentali che, mediante processi cognitivi come l'apprendimento, la riflessione e la comprensione, permette di assimilare concetti e organizzare il comportamento sia nel dominio delle idee sia in quello delle attività pratiche, al fine di risolvere problemi e raggiungere obiettivi.

Di conseguenza, l'Intelligenza Artificiale (IA) è un ambito dell'informatica dedicato alla creazione di sistemi o macchine in grado di simulare o replicare alcune delle capacità cognitive umane. Queste capacità includono l'apprendimento, il ragionamento, la percezione, il linguaggio naturale e la risoluzione di problemi.

L'IA mira a sviluppare sistemi in grado di eseguire compiti complessi in modo autonomo, apprendendo dai dati e adattandosi alle varie situazioni, emulando la versatilità e la flessibilità dell'intelligenza umana.

L'IA è in grado di sviluppare algoritmi, modelli e sistemi che permettono alle macchine di eseguire compiti in modo intelligente, apprendere dagli errori, adattarsi all'ambiente e migliorare le proprie prestazioni nel tempo, il tutto per fare le cose in modo "razionale".

L'IA può essere divisa in due categorie principali: l'IA debole (o ristretta) e l'IA forte (o generale). L'IA debole si riferisce a sistemi progettati per svolgere compiti specifici o problemi limitati, come il riconoscimento vocale, il riconoscimento di immagini, il gioco degli scacchi o l'elaborazione del linguaggio naturale. L'IA forte, d'altra parte, si riferisce a sistemi in grado di affrontare compiti complessi e generali, avvicinandosi o superando le capacità cognitive umane in diversi ambiti.

L'Intelligenza Artificiale utilizza una varietà di approcci e tecniche, avremo il machine learning (apprendimento automatico), le reti neurali artificiali, il deep learning (apprendimento profondo), l'elaborazione del linguaggio naturale, la visione artificiale, la logica fuzzy, l'ottimizzazione, gli algoritmi genetici, e molti altri ancora. L'IA trova applicazioni in settori come l'assistenza sanitaria, la finanza, la produzione, l'automazione, i veicoli autonomi, l'e-commerce, i giochi e l'assistenza virtuale.

L'obiettivo finale dell'IA è quello di sviluppare sistemi che siano in grado di comprendere, adattarsi e interagire con il mondo reale in modo simile ad un essere umano, portando a benefici significativi nella società, nell'economia e nella qualità della vita delle persone.

1.1.1 Breve storia dell'Intelligenza Artificiale

L'Intelligenza Artificiale (IA) ha una storia che risale almeno agli anni '50 del XX secolo, un periodo di grande fermento scientifico sullo studio del calcolatore e il suo utilizzo per sistemi intelligenti:

L'IA è stata ufficialmente considerata un campo di ricerca quando il matematico e informatico Alan Turing pubblica il suo articolo "Computing Machinery and Intelligence" nella seconda metà del 1900, che introduce il famoso "Test di Turing" per valutare l'intelligenza delle macchine.

Gli studi sono proseguiti negli anni fino ad arrivare al giorno d'oggi con il Deep Learning, una branca del machine learning basata su reti neurali profonde. Queste ultime mostrano risultati sorprendenti in diversi compiti, come il riconoscimento delle immagini e la traduzione automatica.

L'IA diventa un argomento di grande rilevanza pubblica e si sperimentano sviluppi avanzati come veicoli autonomi, assistenti virtuali, robotica avanzata, e altro ancora.

Gli sviluppi futuri dell'IA promettono di superare le sfide attuali e aprono la strada a nuove possibilità che possono trasformare radicalmente la nostra società, l'economia e la stessa definizione di intelligenza.

L'Intelligenza Artificiale Generale (AGI) è un termine che si riferisce a un tipo di Intelligenza Artificiale che possiede la capacità di comprendere, apprendere e applicare la conoscenza in modo generale, al pari dell'intelligenza umana.

A differenza di altri tipi di Intelligenza Artificiale specializzata, che sono progettati per compiti specifici, l'AGI mira a sviluppare un'Intelligenza Artificiale che possa raggiungere un livello di intelligenza generale comparabile a quello umano.

L'obiettivo dell'AGI è quello di creare un'Intelligenza Artificiale che possa eseguire con successo una vasta gamma di compiti complessi e adattarsi a nuove situazioni e sfide in modo simile a come farebbe un essere umano. Ciò richiede una comprensione profonda della cognizione umana, dell'apprendimento automatico, del ragionamento, del problem solving e di altre abilità cognitive.

L'AGI rappresenta una sfida tecnologica e scientifica significativa, poiché richiede lo sviluppo di algoritmi, modelli e infrastrutture che possano consentire a un sistema artificiale di acquisire e utilizzare la conoscenza in modo autonomo.

La ricerca sull'AGI è incentrata sull'elaborazione di approcci e metodologie che consentano alla macchina di raggiungere un'intelligenza generale.

L'impiego diffuso dell'AGI potrebbe avere un impatto significativo sull'occupazione, sull'economia, sulla privacy e su altre sfere della società. Pertanto, è necessario condurre ricerche interdisciplinari e adottare una prospettiva olistica per comprendere e affrontare le implicazioni dell'AGI.

Tuttavia, è importante sottolineare che l'AGI è ancora un obiettivo futuro e che attualmente non esiste un'Intelligenza Artificiale che soddisfi completamente i requisiti dell'AGI.

La superintelligenza è un concetto che si riferisce ad un'Intelligenza Artificiale (IA) che supera in modo significativo l'intelligenza umana in tutte le sfere cognitive, mentre l'Intelligenza Artificiale Generale (AGI) mira a creare un'Intelligenza Artificiale con un livello di intelligenza simile a quello umano.

La superintelligenza potrebbe essere in grado di risolvere compiti complessi in modo molto più rapido ed efficiente rispetto alle capacità umane. Potrebbe avere una conoscenza

vasta e dettagliata in un'ampia gamma di discipline, superando anche gli esperti umani in molti settori.

Inoltre, la superintelligenza potrebbe migliorare continuamente le sue capacità di apprendimento e di risoluzione dei problemi, superando costantemente i limiti umani.

1.2 Apprendimento automatico (Machine Learning)

L'apprendimento automatico, noto anche come Machine Learning, è una branca dell'IA che si occupa dello sviluppo di algoritmi e modelli che consentono alle macchine di imparare dai dati e migliorare le proprie prestazioni nel tempo senza essere esplicitamente programmate.

L'obiettivo principale dell'apprendimento automatico è quello di permettere alle macchine di acquisire conoscenze, prendere decisioni o effettuare predizioni basate sui dati disponibili.

L'apprendimento automatico richiede una quantità significativa di dati di addestramento. Questi dati contengono esempi e informazioni su un determinato problema o dominio.

L'accuratezza e la rappresentatività dei dati di addestramento svolgono un ruolo cruciale nel successo del modello di apprendimento automatico.

Il Machine Learning si basa sull'utilizzo di algoritmi capaci di estrarre conoscenze dai dati di addestramento. Questi algoritmi possono essere suddivisi in diverse categorie, tra cui apprendimento supervisionato, apprendimento non supervisionato, apprendimento per rinforzo e apprendimento semi-supervisionato. Ogni categoria si basa su principi e tecniche specifiche per affrontare diverse tipologie di problemi. Esamineremo queste categorie nel seguito.

I modelli di apprendimento automatico vengono valutati utilizzando metriche di prestazione appropriate. Queste metriche misurano l'accuratezza, la precisione, il recall e altri parametri per valutare quanto bene il modello si comporta rispetto ai dati di test.

Attraverso il feedback fornito dalle metriche di valutazione, è possibile apportare miglioramenti al modello, come l'ottimizzazione dei parametri o l'utilizzo di diverse tecniche di pre-elaborazione dei dati.

Nella Figura 1.1 mostriamo graficamente le modalità di funzionamento del Machine Learning.

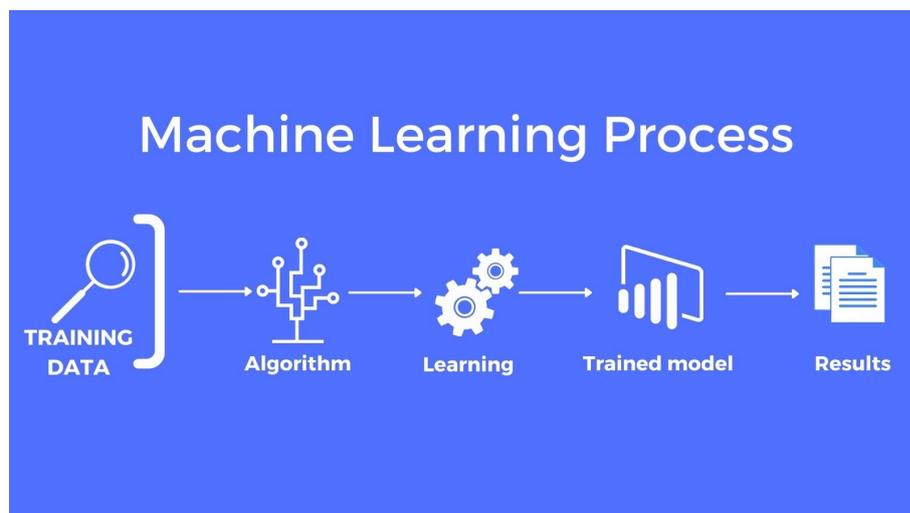


Figura 1.1: Modalità di funzionamento del Machine Learning

L'apprendimento automatico trova applicazione in una vasta gamma di settori e problemi, tra cui:

- *Riconoscimento di pattern e classificazione*: ad esempio riconoscimento di immagini, riconoscimento vocale, analisi dei testi, rilevamento di frodi, diagnosi medica.
- *Previsione*: ad esempio previsione del mercato azionario, previsione del tempo, previsione della domanda di prodotti.
- *Raccomandazioni personalizzate*: ad esempio sistemi di raccomandazione di film, musica o prodotti, personalizzazione dell'esperienza utente.
- *Automazione e robotica*: ad esempio veicoli autonomi, robotica industriale, automazione dei processi.
- *Elaborazione del linguaggio naturale*: ad esempio traduzione automatica, assistenti virtuali, analisi del sentiment, chatbot.

1.2.1 Reti neurali artificiali

Le reti neurali artificiali note anche come ANN (Artificial Neural Network) o SNN (Simulated Neural Network); sono un concetto chiave nell'Intelligenza Artificiale (IA) e nell'apprendimento automatico.

Si ispirano al funzionamento del cervello umano e sono modelli matematici composti da interconnessioni di unità di elaborazione chiamate neuroni artificiali.

Volendo dare una definizione più dettagliata potremmo dire che le reti neurali sono modelli di calcolo matematico-informatici basati sul funzionamento delle reti neurali biologiche, ossia modelli costituiti da interconnessioni di informazioni. Tali interconnessioni derivano da neuroni artificiali e processi di calcolo basati sul modello delle scienze cognitive chiamato "connessionismo", vale a dire basati su PDP – Parallel Distributed Processing, elaborazione a parallelismo distribuito delle informazioni. Secondo tale principio il cervello umano elabora le informazioni dei vari sensi in modo parallelo e distribuisce le informazioni in tutti i vari nodi della rete, non in una memoria centrale.

Facendo il paragone con l'informatica tradizionale, i calcoli avvengono in modo seriale e non in parallelo e i dati vengono memorizzati in una memoria centrale, anche se, come vedremo, negli ultimi anni si sono fatti enormi balzi in avanti sia sul fronte delle memorie sia dal punto di vista del calcolo computazionale che dei processori hardware. La Figura 1.2 mostra un esempio di funzionamento delle reti neurali.

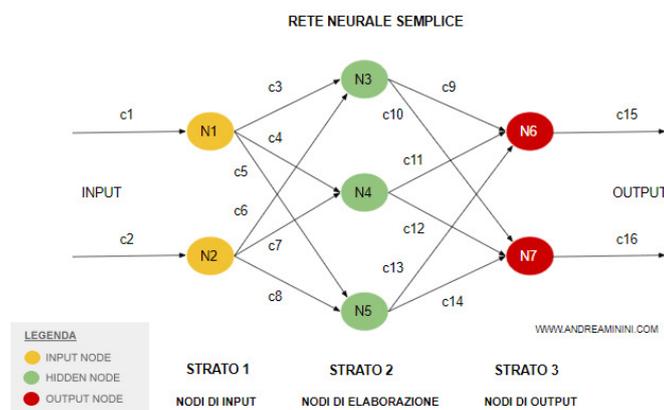


Figura 1.2: Esempio del funzionamento delle reti neurali

Una rete neurale di fatto si presenta come un sistema “adattivo” in grado di modificare la sua struttura (i nodi e le interconnessioni) basandosi sia su dati esterni sia su informazioni interne che si connettono e passano attraverso la rete durante la fase di apprendimento e ragionamento.

Le reti neurali artificiali sono strutture non-lineari di dati statistici organizzate come strumenti di modellazione; esse ricevono segnali esterni su uno strato di nodi (che rappresenta l’unità di elaborazione, il processore); ognuno di questi “nodi d’ingresso” è collegato a svariati nodi interni della rete che, tipicamente, sono organizzati a più livelli in modo che ogni singolo nodo possa elaborare i segnali ricevuti trasmettendo ai livelli successivi il risultato delle sue elaborazioni (quindi delle informazioni più evolute, dettagliate).

In linea di massima, le reti neurali sono formate da tre tipologie (che, però, possono coinvolgere migliaia di neuroni e decine di migliaia di connessioni):

- lo strato degli ingressi (I – Input): è quello che ha il compito di ricevere ed elaborare i segnali in ingresso adattandoli alle richieste dei neuroni della rete;
- il cosiddetto strato nascosto (H – hidden): è quello che ha in carica il processo di elaborazione vero e proprio (e può anche essere strutturato con più colonne-livelli di neuroni);
- lo strato di uscita (O – Output): qui vengono raccolti i risultati dall’elaborazione dello strato H che vengono adattati alle richieste del successivo livello-blocco della rete neurale.

1.2.2 I modelli di apprendimento delle reti neurali

Affinché questo processo risulti performante è necessario “addestrare” le reti neurali, ossia fare in modo che esse apprendano come comportarsi nel momento in cui andrà risolto un problema ingegneristico, come, per esempio, il riconoscimento di un essere umano analizzando delle immagini (tecnologia del riconoscimento facciale).

Il tema dell’apprendimento è collegato al Machine Learning, inteso come insieme di algoritmi che utilizzano metodi matematico-computazionali per apprendere informazioni dall’esperienza (quindi in modo automatico e adattivo). I modelli principali in uso oggi sono i seguenti:

- *Supervised Learning*: all’algoritmo vengono forniti sia set di dati come input sia le informazioni relative ai risultati desiderati con l’obiettivo che la rete identifichi una regola generale che colleghi i dati in ingresso con quelli in uscita; in altre parole vengono forniti degli esempi di input e di output in modo che il sistema impari il nesso tra essi e ne estrapoli una regola riutilizzabile per altri compiti simili.
- *Unsupervised Learning*: al sistema vengono forniti solo set di dati senza alcuna indicazione del risultato desiderato. Lo scopo di questo secondo metodo di apprendimento è “risalire” a schemi e modelli nascosti, ossia identificare negli input una struttura logica senza che questi siano preventivamente etichettati.
- *Apprendimento per rinforzo*: in questo caso, il sistema deve interagire con un ambiente dinamico che gli consente di avere i dati di input e perseguire un obiettivo al raggiungimento del quale riceve una ricompensa, imparando anche dagli errori identificati mediante “punizioni”. Il comportamento e le prestazioni del sistema sono determinati da una routine di apprendimento basata su ricompensa e punizione.

- *Apprendimento semi-supervisionato*: è un modello “ibrido” dove al computer viene fornito un set di dati incompleti per l’allenamento/apprendimento. Alcuni di questi input sono “dotati” dei rispettivi esempi di output (come nell’apprendimento supervisionato); altri, invece, ne sono privi (come nell’apprendimento non supervisionato). L’obiettivo di fondo è sempre lo stesso: identificare regole e funzioni per la risoluzione dei problemi, nonché modelli e strutture di dati utili a raggiungere determinati obiettivi.

1.2.3 Applicazioni delle reti neurali artificiali

Le reti neurali artificiali sono state utilizzate con successo in molte applicazioni dell’Intelligenza Artificiale. Altre tipologie di reti neurali sono le seguenti:

1. *Le reti neurali feedforward*, o perceptron multilivello (multi-layer perceptron, MLP), formate da un livello di input, uno o più livelli nascosti e un livello di output. Sebbene si faccia anche comunemente riferimento a queste reti neurali come MLP, è importante notare che esse sono effettivamente formate da neuroni sigmoidei, non perceptron, poiché la maggior parte dei problemi del mondo reale è non lineare. I dati vengono di norma passati in questi modelli per essere addestrate e rappresentano la base per la visione artificiale, l’NLP (Natural Language Processing) e altre reti neurali. Sono caratterizzate da un passaggio di informazioni che va in un’unica direzione.
2. *Le reti neurali convoluzionali* (Convolutional Neural Network, CNN), ampiamente utilizzate per il riconoscimento e la classificazione di immagini e oggetti nel contesto della computer vision. Queste reti sfruttano i principi dall’algebra lineare, in particolare la moltiplicazione della matrice, per identificare i modelli all’interno di un’immagine.
3. *Le reti neurali ricorrenti* (Recurrent Neural Network, RNN) e *le reti neurali trasformazionali* (Transformer), utilizzate per compiti come il riconoscimento del parlato, la traduzione automatica, la generazione di testi, l’analisi del sentimento e le serie temporali, come le previsioni meteorologiche o quelle finanziarie. Vengono applicate, inoltre, al controllo di sistemi complessi, come il controllo di robot o di processi industriali, e vengono utilizzate per creare sistemi di raccomandazione personalizzati, come quelli utilizzati dalle piattaforme di streaming video o di e-commerce.

1.2.4 Algoritmi genetici

Gli algoritmi genetici sono una tecnica di ottimizzazione e ricerca ispirata dal processo di evoluzione naturale spesso utilizzata nell’Intelligenza Artificiale per risolvere problemi complessi, migliorare la performance dei modelli e trovare soluzioni ottimali in modo efficiente.

Gli algoritmi genetici partono da un certo numero di possibili soluzioni (individui), chiamate popolazione, che sono rappresentate da cromosomi che possono essere codificati come sequenze di geni o bit.

Durante ciascuna iterazione l’algoritmo effettua una selezione degli elementi della popolazione e li combina per creare nuovi elementi (nuova generazione) della popolazione stessa.

La combinazione di due soluzioni crea una terza soluzione (nuova generazione) che eredita dalle precedenti alcune caratteristiche e le combina in un nuovo patrimonio “genetico” in modo simile a quanto accade in natura.

Oltre alla combinazione delle caratteristiche delle soluzioni discendenti l’algoritmo genetico può introdurre delle mutazioni casuali che introducono nuove caratteristiche che si aggiungono a quelle originarie incrementando il numero delle soluzioni.

Come accade per la selezione naturale, i nuovi elementi più efficienti/efficaci (elementi forti) si sostituiscono a quelli meno efficienti/efficaci (elementi deboli).

La successione delle generazioni determina l'evoluzione verso la soluzione ottimale.

Gli elementi fondamentali di un algoritmo genetico sono i seguenti:

- *La funzione di fitness*, la quale consente di calcolare la probabilità di riproduzione per ciascun elemento della popolazione misurando il livello di adattabilità di un individuo della popolazione (una soluzione al problema) all'ambiente.
- *La selezione naturale*, che sceglie in base alle probabilità di riproduzione gli elementi per la formazione delle nuove coppie. Generalmente, la selezione è l'accoppiamento delle soluzioni migliori, ovvero quelle con un valore più alto nella funzione di fitness, per generare altre soluzioni.
- *Il cross-over*, il quale determina la modalità di combinazione del patrimonio genetico degli elementi genitori negli elementi figli (nuova generazione). Il cross-over (o crossover) rappresenta l'accoppiamento di due soluzioni con lo scopo di generarne altre.
- *La mutazione casuale*, che avviene quando l'algoritmo genetico inserisce alcune variazioni casuali nel patrimonio genetico degli elementi figli per ottenere un'altra soluzione migliore o peggiore della precedente. La mutazione è una modifica casuale di una variabile binaria all'interno di una soluzione.

Nella Figura 1.3 viene riportato un esempio di funzionamento di un algoritmo genetico.

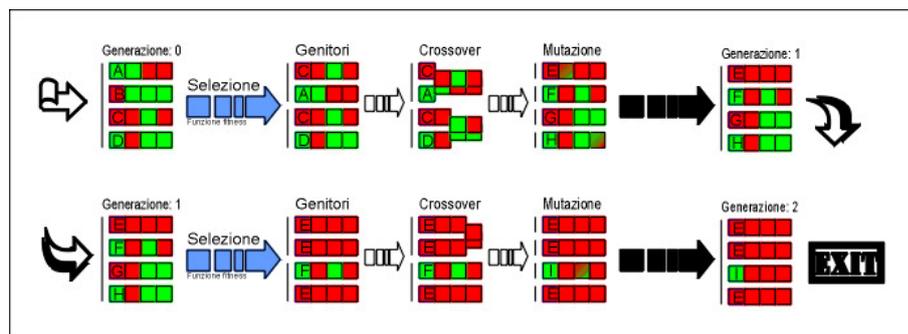


Figura 1.3: Esempio di un algoritmo genetico

Il processo di selezione, crossover e mutazione viene ripetuto per diverse generazioni, evolvendo in questo modo la popolazione, fino a raggiungere una soluzione ottimale o un criterio di terminazione predefinito.

Gli algoritmi genetici sono ampiamente utilizzati nell'IA per affrontare problemi complessi di ottimizzazione, adattamento e ricerca e, per questo, vengono utilizzati per trovare la configurazione ottimale dei parametri di un modello o di un algoritmo di apprendimento automatico.

Ad esempio, possono essere applicati per ottimizzare i pesi di una rete neurale nonché nella progettazione di circuiti dove è necessario trovare una combinazione ottimale di componenti e collegamenti.

Servono, anche, per trovare un percorso ottimale in un grafo, come nel caso del problema del commesso viaggiatore e per la progettazione di strutture complesse, come ponti o edifici, ottimizzando la disposizione dei materiali e la resistenza strutturale.

In conclusione, gli algoritmi genetici vengono impiegati principalmente per selezionare le migliori funzioni o caratteristiche da utilizzare in un modello di apprendimento automatico, riducendo la dimensionalità dei dati e migliorando la performance del modello.

1.2.5 Logica fuzzy

La logica fuzzy è una forma di logica che si occupa di gestire l'incertezza e l'approssimazione. A differenza della logica classica, che utilizza solo valori di verità "vero" o "falso", la logica fuzzy consente l'assegnazione di valori di verità parziali compresi tra 0 e 1.

Nella logica fuzzy un oggetto è contemporaneamente sia vero che falso ma con diversi valori di verità. Questo permette di rappresentare e manipolare concetti che sono sfumati o ambigui.

La logica fuzzy si basa sul concetto di insiemi fuzzy; questi sono insiemi la cui appartenenza degli elementi può variare in modo graduale.

Invece di definire una regola rigida per l'appartenenza o la non appartenenza di un elemento a un insieme, la logica fuzzy assegna gradi di verità in base a quanto un elemento soddisfa i criteri dell'insieme. Ad esempio, invece di valutare un bicchiere solo pieno (A) o vuoto (B), la logica fuzzy consente di assegnare valori intermedi, come 0.5 o 0.3, per riflettere il livello di aderenza alle caratteristiche di un bicchiere pieno (Figura 1.4).

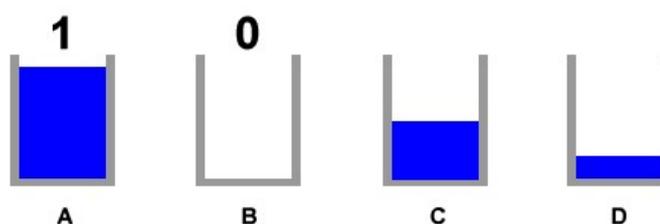


Figura 1.4: Logica Fuzzy: esempio del bicchiere

La logica fuzzy trova applicazione in diversi campi, come il controllo dei sistemi, l'Intelligenza Artificiale, la gestione dell'incertezza e il riconoscimento di modelli.

Viene utilizzata per affrontare problemi complessi in cui le condizioni non sono definite in modo preciso o in cui esistono livelli di verità sfumati.

Grazie alla sua capacità di modellare e ragionare su informazioni approssimate, la logica fuzzy offre un'alternativa potente alla logica classica per affrontare la complessità del mondo reale.

1.3 Natural Language Processing

Il Natural Language Processing (NLP), o elaborazione del linguaggio naturale, è un settore specifico dell'Intelligenza Artificiale che si occupa di fornire ai computer la capacità di comprendere il testo e le parole pronunciate allo stesso modo in cui lo fanno gli esseri umani.

1.3.1 Come funziona

Il Natural Language Processing funziona attraverso degli algoritmi che servono a distinguere i diversi aspetti del dialogo tra uomo e macchina, quali la fonetica, la morfologia, la semantica e i diversi contesti in cui il discorso può avvenire. Questi algoritmi sono principalmente tre: tokenizzazione, stemming e lemmatizzazione. La tokenizzazione ha la capacità di dividere il parlato in frasi e singole parole, distinguendo, così, quali parole appartengono ad una frase principale e quali appartengono alle frasi secondarie. Lo stemming è il processo di riduzione di una forma flessa al suo tema, cioè la parte della parola che rimane dopo aver tolto la desinenza. La lemmatizzazione è il processo che permette di arrivare al suo lemma, quindi dalla forma flessa a quella canonica.

1.3.2 Ambiti applicativi

Gli ambiti applicativi dei sistemi di elaborazione del linguaggio naturale per il business possono essere riassunti nei seguenti quattro:

- *Il riconoscimento vocale, chiamato anche speech-to-text*, in grado di convertire in modo affidabile i dati vocali in dati di testo. Il riconoscimento vocale è necessario per qualsiasi applicazione che esegue comandi vocali o risponde a domande orali.
- *La generazione del linguaggio naturale text-to-speech*, la quale inserisce informazioni strutturate in linguaggio umano.
- *La disambiguazione del senso delle parole*, che seleziona il significato corretto di una parola che possiede molteplici significati attraverso un processo di analisi semantica determinando il significato che ha più senso nel contesto dato.
- *L'analisi del sentiment*, la quale tenta di estrarre qualità soggettive - atteggiamenti, emozioni, sarcasmo, confusione, sospetto - dal testo fornito.

1.4 Implicazioni etiche dell'Intelligenza Artificiale

L'implementazione dell'Intelligenza Artificiale (IA) ha avuto un impatto significativo sull'occupazione umana, portando a cambiamenti nel modo in cui lavoriamo e nelle competenze richieste (Figura 1.5).



Figura 1.5: L'etica

Ecco alcune considerazioni relative all'impatto dell'IA sull'occupazione:

- *Automazione di compiti ripetitivi*: l'IA ha reso possibile l'automazione di compiti ripetitivi e routine in diversi settori. Ciò ha portato a una riduzione della necessità di lavoratori per eseguire tali attività, poiché possono essere svolte in modo più efficiente e accurato da sistemi di Intelligenza Artificiale.
- *Riduzione di posti di lavoro manuali*: l'IA e la robotica hanno avuto un impatto significativo sui lavori manuali, come nell'industria manifatturiera. I robot e gli algoritmi intelligenti possono svolgere compiti fisici che richiedono forza e ripetitività, portando alla riduzione di alcuni lavori manuali tradizionali.
- *Creazione di nuovi ruoli e professioni*: l'IA ha anche aperto la strada alla creazione di nuovi ruoli e professioni nel campo dell'IA stessa. Sono emersi nuovi lavori come scienziati dei dati, ingegneri dell'apprendimento automatico e specialisti dell'IA, che richiedono competenze specifiche per sviluppare, implementare e gestire sistemi di Intelligenza Artificiale.

- *Automazione dei processi decisionali*: l'IA ha anche impattato i lavori che coinvolgono processi decisionali. Gli algoritmi di IA possono analizzare grandi quantità di dati e fornire raccomandazioni o prendere decisioni in base a tali analisi. Ciò ha portato a una riduzione del coinvolgimento umano in alcune decisioni e a una revisione del ruolo umano nella supervisione e nell'interpretazione delle decisioni prese dai sistemi di IA.
- *Collaborazione uomo-macchina*: mentre l'IA può automatizzare alcuni compiti, ci sono anche molti scenari in cui si sta sviluppando la collaborazione tra umani e macchine. L'IA può migliorare le capacità umane, fornendo assistenza, analisi e supporto nella presa di decisioni complesse. Ciò significa che molte occupazioni stanno sperimentando una trasformazione in cui gli umani lavorano in tandem con sistemi di IA per ottenere risultati migliori.
- *Necessità di nuove competenze*: l'avanzamento dell'IA richiede nuove competenze e adattabilità da parte dei lavoratori. Competenze come la comprensione dell'IA, l'interpretazione dei risultati degli algoritmi, la capacità di lavorare con dati e algoritmi e le competenze di problem-solving e pensiero critico sono sempre più importanti per rimanere competitivi nel mercato del lavoro.

In questo capitolo vedremo cos'è Amazon Web Service e i suoi servizi soffermandoci anche, su cos'è il cloud e nella sicurezza offerta da Amazon Web Service.

2.1 Cos'è Amazon Web Services

Amazon Web Services (AWS) è una piattaforma di cloud computing, messa a disposizione da Amazon. Si caratterizza per una vasta gamma di servizi (schematizzati in Figura 2.1) messi a disposizione sia alle aziende che ai singoli utenti.



Figura 2.1: I servizi offerti da AWS

2.1.1 AWS e il Cloud

Il cloud computing consiste nella fornitura di servizi di computing, quali software, database, server, reti e altre risorse IT, tramite una piattaforma di servizi cloud via Internet con tariffe del tipo pay-as-you-go. Ciò significa che gli utenti finali sono in grado di accedere a software e applicazioni ovunque si trovino.

Con il cloud computing non è necessario effettuare ingenti investimenti iniziali in hardware e dedicare tempo alla sua gestione. Ora è possibile fornire esattamente il tipo e le dimensioni di risorse di elaborazione necessarie per realizzare nuove idee o gestire il reparto IT. È possibile accedere rapidamente a tutte le risorse necessarie e pagare solo per ciò che effettivamente si utilizza.

Avendo le nuove risorse IT a portata di click, si riduce il tempo necessario per rendere le risorse disponibili agli sviluppatori da settimane a pochi minuti, quindi abbiamo un aumento della velocità. Ciò si traduce in un notevole aumento dell'agilità per l'organizzazione, poiché il costo e il tempo necessari per sperimentare e sviluppare sono notevolmente inferiori.

Il cloud computing è reso possibile grazie alla virtualizzazione che consente di creare un computer "virtuale" simulato ed esclusivamente digitale che si comporta come se fosse un computer fisico con hardware proprio.

È molto più facile distribuire la propria applicazione in più regioni del mondo in poco tempo, fornendo così una latenza inferiore e un'esperienza migliore ai clienti.

2.1.2 Tipi di cloud computing

In passato, per fare una distinzione tra cloud pubblico, privato, ibrido e multicloud bastava fare riferimento alla proprietà e al luogo in cui si trovavano, le macchine. Ma oggi queste differenze non sono più così evidenti.

I cloud pubblici sono ambienti cloud basati su un'infrastruttura IT, che solitamente, non appartiene all'utente finale. I provider di cloud pubblico più importanti includono Alibaba Cloud, Amazon Web Services (AWS), Google Cloud, IBM Cloud e Microsoft Azure.

I cloud privati sono ambienti cloud riservati a un singolo utente finale o a una singola azienda che operano dietro il firewall di tale utente o azienda. Quando l'infrastruttura IT sottostante è dedicata a un singolo cliente, con un accesso completamente isolato, tutti i cloud diventano privati.

I cloud privati gestiti sono un'alternativa adatta alle piccole aziende o con personale IT non specializzato, poiché offrono servizi di cloud privato e infrastrutture migliori ai clienti.

Un cloud ibrido appare come un singolo ambiente IT creato a partire da più ambienti connessi tramite reti LAN (Local Area Network), WAN (Wide Area Network) o VPN (Virtual Private Network) e/o API. Ogni sistema IT si trasforma in cloud ibrido quando le app che vi risiedono possono muoversi tra più ambienti, separati ma connessi.

Un ambiente multicloud è formato da più di un servizio cloud e da più di un fornitore di servizi cloud, pubblici o privati. Tutti i cloud ibridi sono multicloud, ma ciò non implica che tutti i multicloud sono cloud ibridi.

2.1.3 Servizi Cloud

I servizi cloud sono costituiti da infrastrutture, piattaforme o software ospitati da fornitori esterni e resi accessibili agli utenti tramite Internet. Questi servizi si suddividono principalmente in tre categorie as-a-Service: IaaS (Infrastructure as a Service), PaaS (Platform as a Service) e SaaS (Software as a Service), mostrati nella Figura 2.2. In ciascuna di queste categorie, i dati dell'utente fluiscono tra i dispositivi client front-end e i sistemi del fornitore di servizi cloud attraverso Internet, offrendo una gamma diversificata di soluzioni.

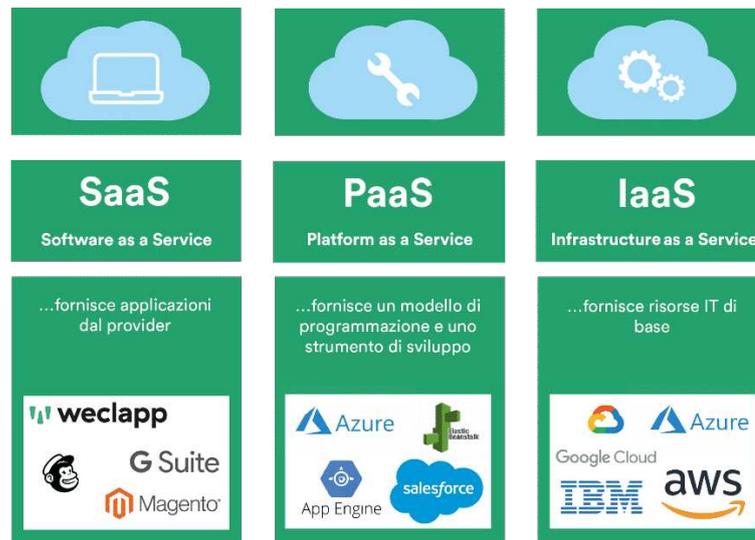


Figura 2.2: Servizi di Cloud Computing

Nel contesto di un'implementazione IaaS, il fornitore di servizi cloud si occupa della gestione dell'infrastruttura, comprendente server fisici, reti, virtualizzazione e archiviazione dei dati, per conto del cliente attraverso una connessione Internet. L'utente ottiene l'accesso a questa infrastruttura, fondamentalmente come un servizio noleggiato, utilizzando un'API o un cruscotto, e gestisce aspetti come il sistema operativo, le applicazioni e il middleware.

L'utente non ha bisogno di gestire o controllare l'infrastruttura del cloud, può scegliere quali sistemi operativi e applicazioni eseguire e può avere la possibilità di controllare l'accesso di rete (ad esempio, configurare un firewall). Il provider gestisce, invece, tutti gli aspetti hardware, inclusi connettività di rete, dischi rigidi, archiviazione dei dati e server. Esso si assume la responsabilità per la gestione di guasti, riparazioni e problemi hardware. Questo rappresenta il modello di implementazione tipico per i servizi di storage basati su cloud forniti dai provider.

In una soluzione PaaS, l'hardware e la piattaforma di software applicativo sono forniti e gestiti da un provider di servizi cloud esterno. L'infrastruttura di base include molte delle strutture IaaS (server, strutture di archiviazione, OS database e connessioni di rete).

Le strutture per lo sviluppo e la distribuzione del software includono compilatori, middleware, librerie di programmi, sistemi in esecuzione (ad esempio, Java runtime e .NET runtime) e servizi che ospitano le applicazioni di un cliente.

L'utente, tuttavia, gestisce le applicazioni in esecuzione sulla piattaforma e i dati da esse utilizzati. Progettato principalmente per sviluppatori e programmatori, il servizio PaaS offre agli utenti una piattaforma cloud condivisa per lo sviluppo e la gestione delle applicazioni, svolgendo un ruolo significativo nella metodologia DevOps. Questo approccio consente agli utenti di concentrarsi sullo sviluppo e la gestione delle applicazioni senza la necessità di creare e gestire l'infrastruttura tipicamente associata a questo processo.

Un'opzione SaaS offre agli utenti un'applicazione software gestita dal provider di servizi cloud. Queste applicazioni, che di solito sono applicazioni web o per dispositivi mobili, sono accessibili dagli utenti attraverso un browser. L'utente, che si collega alle applicazioni cloud tramite un'API o un cruscotto, è responsabile degli aggiornamenti software, della risoluzione dei bug e di altre attività di manutenzione generica del software. Inoltre, il servizio SaaS elimina la necessità di installare un'applicazione specifica sui singoli computer degli utenti, offrendo diverse modalità di accesso al software a livello di team o gruppo.

Quando un utente accede, l'applicazione viene eseguita su un server in un data center cloud anziché sul computer dell'utente. I file dell'utente vengono archiviati nel data center cloud anziché sul dispositivo locale dell'utente. Un esempio è dovuto a Office 365 di Microsoft e dai suoi servizi.

I vantaggi sono i seguenti:

1. *Accesso universale*: il software SaaS può essere accessibile in qualsiasi momento da qualsiasi dispositivo.
2. *Sincronizzazione garantita*: l'utente non dovrà mai sincronizzare manualmente le copie su più dispositivi.
3. *Alta disponibilità*: ci sono sistemi elaborati di backup. Alcuni sistemi SaaS utilizzano un browser web come app di accesso.

2.2 Servizi di AWS

AWS è costituito da molti servizi cloud che si possono utilizzare in combinazioni personalizzate in base alle proprie esigenze aziendali o organizzative. Per accedere ai servizi, si possono utilizzare i AWS Management Console, e gli AWS CLI Software Development Kit (SDK).

2.2.1 Machine Learning

Il Machine Learning è progettato per fornire strumenti e servizi digitali per imparare dai dati, identificare modelli, fare previsioni e, quindi, agire sulla base di queste ultime. Esso utilizza grandi quantità di dati per creare e convalidare la logica decisionale che costituisce la base del modello di Intelligenza Artificiale.

Un sottoinsieme del Machine Learning in rapida crescita è l'IA generativa, che si basa su modelli di grandi dimensioni preaddestrati su un vasto set di dati, comunemente denominati modelli di fondazione (FM).

I modelli di Machine Learning utilizzati dipendono fortemente dalla qualità dei dati di input per generare risultati accurati. Poiché i dati sono soggetti a modifiche nel tempo, è necessario un monitoraggio continuo per rilevare, correggere e mitigare eventuali problemi legati all'accuratezza e alle prestazioni. Questa fase di monitoraggio può richiedere il re-training del modello nel tempo, utilizzando i dati più recenti e raffinati.

A differenza dei carichi di lavoro delle applicazioni, che seguono istruzioni passo-passo per risolvere un problema, i carichi di lavoro di Machine Learning permettono agli algoritmi di apprendere dai dati in un ciclo iterativo e continuo. Il ML Lens, integrandosi con il Framework Well-Architected, affronta e amplia questa differenza, fornendo indicazioni specifiche per i carichi di lavoro di Machine Learning.

2.2.2 Amazon EC2

Amazon Elastic Compute Cloud (Amazon EC2) offre una capacità di elaborazione flessibile e on-demand all'interno dell'ecosistema di cloud Amazon Web Services (AWS). Sfruttando Amazon EC2, è possibile ridurre i costi hardware, consentendo un sviluppo e una distribuzione più veloci delle applicazioni.

Questa piattaforma consente di avviare facilmente i server virtuali necessari, configurare parametri di sicurezza e rete, nonché gestire lo storage. In situazioni di carichi di lavoro intensi, come processi mensili o annuali o improvvisi picchi di traffico sul sito web, è possibile aumentare la capacità. Al contrario, in periodi di minor utilizzo, è possibile ridimensionare la capacità, adattandola alle esigenze del momento.

Amazon EC2 offre il supporto per l'elaborazione, la memorizzazione e la trasmissione dei dati delle carte di credito da parte di commercianti o fornitori di servizi, ed è stata convalidata come conforme allo standard di sicurezza dei dati delle carte di pagamento (PCI DSS - Payment Card Industry Data Security Standard).

2.2.3 Amazon S3

Amazon Simple Storage Service (Amazon S3) è un servizio di archiviazione di oggetti che offre una combinazione eccellente di scalabilità, disponibilità dei dati, sicurezza e prestazioni. Aziende di varie dimensioni e settori possono sfruttare Amazon S3 per archiviare e proteggere qualsiasi quantità di dati, adattandosi a una vasta gamma di casi d'uso, tra cui data lake, siti web, applicazioni mobili, backup e ripristino, archivio, applicazioni enterprise, dispositivi IoT e analisi di big data. Amazon S3 mette a disposizione funzionalità di gestione che consentono di ottimizzare, organizzare e configurare l'accesso ai dati, rispondendo così ai requisiti specifici delle aziende, delle organizzazioni e delle normative di conformità.

2.2.4 Amazon DynamoDB

Amazon DynamoDB è un Database Management System NoSQL completamente gestito da AWS, progettato per fornire prestazioni veloci e prevedibili con una scalabilità continua. DynamoDB semplifica la gestione e l'espansione di database distribuiti, eliminando la necessità di occuparsi della fornitura di hardware, della configurazione e installazione, della replica, delle patch software e della scalabilità del cluster. Il servizio include anche la crittografia dei dati memorizzati, semplificando la protezione di informazioni sensibili.

Con DynamoDB, è possibile creare tabelle di database in grado di archiviare e recuperare qualsiasi quantità di dati, gestendo contemporaneamente qualsiasi livello di traffico delle richieste. La flessibilità del servizio consente di regolare la quantità di dati delle tabelle trasmessi in una unità di tempo senza dover affrontare tempi di inattività o riduzioni delle prestazioni. La console di gestione di AWS offre strumenti per monitorare l'utilizzo delle risorse e le metriche delle prestazioni.

2.2.5 Lambda

AWS Lambda è un servizio di calcolo che consente l'esecuzione di codice senza la necessità di gestire server.

Lambda esegue il codice su un'infrastruttura di calcolo altamente disponibile, gestendo integralmente le risorse di calcolo. Questa gestione include la manutenzione dei server e del sistema operativo, la fornitura di capacità, la scalabilità automatica e la registrazione dei log.

2.3 Sicurezza

I servizi e gli strumenti di crittografia di AWS utilizzano un'ampia gamma di tecnologie di archiviazione e crittografia che aiutano a proteggere i dati archiviati e in transito. Tradizionalmente, per poter essere utilizzati in un'operazione di calcolo, i dati dovevano essere decrittografati. Il calcolo crittografico è una tecnologia che opera direttamente sui dati protetti da crittografia, in modo che i dati sensibili non risultino mai esposti.

Nel cloud la protezione di dati è una responsabilità condivisa tra AWS e il cliente. Tra i clienti ci possono essere fornitori di servizi finanziari, aziende del settore sanitario e agenzie governative che si fidano di AWS per le loro informazioni sensibili.

2.3.1 Calcolo crittografico in AWS

Il calcolo crittografico si riferisce a un'ampia gamma di tecniche per la salvaguardia della privacy, tra cui il calcolo multilaterale sicuro, la crittografia omomorfa, l'apprendimento federato nella salvaguardia della privacy e la crittografia ricercabile. AWS sta sviluppando servizi e strumenti di calcolo crittografico per aiutare a raggiungere i propri obiettivi di sicurezza e conformità, consentendo di sfruttare la flessibilità, la scalabilità, le prestazioni e la facilità d'uso offerte.

2.3.2 Protezione dei dati

La protezione della privacy avviene con servizi e funzionalità che consentono di implementare i controlli sulla privacy, inclusi l'accesso avanzato, la crittografia e la funzionalità di logging, semplificando la crittografia dei dati in transito e inattivi avvalendosi di chiavi gestite da AWS o completamente gestite dal cliente.

Il pilastro della sicurezza di AWS Well-Architected Framework aiuta gli architetti del cloud nella progettazione di infrastrutture sicure, ad alte prestazioni ed efficienti, adatte a una vasta gamma di applicazioni e carichi di lavoro. Fondato su sei principi fondamentali (eccellenza operativa, sicurezza, affidabilità, efficienza delle prestazioni, ottimizzazione dei costi e sostenibilità), AWS Well-Architected offre un approccio ottimizzato per la valutazione delle architetture e l'implementazione di progetti facilmente scalabili sia per i clienti che per i partner.

AWS ha ottenuto certificazioni e accreditamenti riconosciuti a livello internazionale, dimostrando la conformità a rigorosi standard internazionali, quali ISO 27017 per la sicurezza nel cloud, ISO 27701 per la gestione delle informazioni sulla privacy e ISO 27018 per la privacy nel cloud. Non vengono usati i dati del cliente né eventuali informazioni derivate da essi a scopo pubblicitario o di marketing.

Riconoscimento di immagini

In questo capitolo vedremo cos'è Rekognition e le sue funzioni, per poi spiegare dei test svolti su questa piattaforma.

3.1 Cos'è Rekognition

Amazon Rekognition è un servizio di riconoscimento ad autoapprendimento di AWS che può essere usato per identificare oggetti o persone.

Esso include un'API semplice e facile da usare in grado di analizzare rapidamente qualsiasi file di immagine o video archiviato in Amazon S3. Il software impara continuamente da nuovi dati e aggiunge continuamente nuove etichette e funzionalità di confronto facciale al servizio.

L'API Amazon Rekognition è il servizio che può:

- identificare le etichette (oggetti, concetti, persone, scene e attività) e il testo;
- identificare le proprietà delle immagini;
- rilevare automaticamente contenuti espliciti od osceni nelle immagini e fornire punteggi di affidabilità;
- fornire funzionalità di analisi facciale, confronto dei volti e ricerca dei volti estremamente accurate.

Con le API di riconoscimento facciale di Amazon Rekognition, è possibile rilevare, analizzare e confrontare i volti per un'ampia varietà di casi d'uso, tra cui la verifica degli utenti, la catalogazione, il conteggio delle persone e la sicurezza pubblica.

Amazon Rekognition si basa sulla stessa tecnologia di deep learning sviluppata dagli esperti di visione artificiale di Amazon per analizzare miliardi di immagini e video ogni giorno. Non richiede competenze di Machine Learning per essere utilizzata.

Rekognition Video è un servizio di riconoscimento video che rileva le attività, interpreta i movimenti delle persone in un filmato e riconosce oggetti, personaggi famosi e contenuti inappropriati nei video archiviati in Amazon S3 e nei flussi in diretta. Rekognition Video rileva le persone e le segue nel video anche quando i loro volti non sono visibili e quando la persona esce o entra nell'inquadratura. Ad esempio, può essere utile in un'applicazione che invia notifiche in tempo reale quando qualcuno consegna un pacco alla porta. Rekognition Video

consente, anche, di indicizzare i metadati, come oggetti, attività, scene, punti di riferimento, personaggi famosi e volti, agevolando la ricerca nel video.

3.1.1 La visione artificiale

La visione artificiale è una tecnologia che sfrutta le reti neurali convoluzionali e ricorrenti, che le macchine utilizzano per riconoscere automaticamente le immagini e descriverle in modo accurato ed efficiente. Oggi i sistemi informatici hanno accesso a un grande volume di immagini e dati video provenienti o creati da smartphone, telecamere del traffico, sistemi di sicurezza e altri dispositivi. Le applicazioni di visione artificiale utilizzano l'Intelligenza Artificiale e il Machine Learning per elaborare accuratamente questi dati, in modo da identificare gli oggetti e riconoscere i volti, nonché per la classificazione, la raccomandazione, il monitoraggio e il rilevamento.

3.1.2 Servizi di Amazon Rekognition

I servizi offerti da Amazon Rekognition sono i seguenti:

- rilevamento delle etichette;
- proprietà dell'immagine;
- moderazione delle immagini;
- analisi facciale;
- confronto facciale;
- riconoscimento di volti celebri;
- testo nell'immagine
- rilevamento dei PPE

Amazon Rekognition consente l'accesso gratuito a tali servizi per realizzare dei piccoli dimostratori. Qualora si vogliano realizzare dei sistemi commerciali sarà necessario pagare.

Nell'ambito della presente tesi noi utilizzeremo: l'analisi delle etichette, l'analisi facciale e il confronto facciale. Per quanto riguarda l'analisi delle etichette, abbiamo che un'etichetta può essere un oggetto, una scena o un concetto rilevato in un'immagine in base ai relativi contenuti. Ad esempio, la foto di alcune persone su una spiaggia tropicale può contenere etichette come "Persona", "Acqua", "Sabbia", "Palma" e "Costume" (oggetti), "Spiaggia" (scena) e "Esterno" (concetto).

Rekognition supporta migliaia di etichette appartenenti a categorie comuni, come, ad esempio persone, animali, piante, eventi, trasporti ed elettronica.

L'analisi facciale consiste nel rilevamento di un volto all'interno di un'immagine e nell'estrazione dei relativi attributi. Amazon Rekognition Image restituisce la cornice per ogni volto rilevato in un'immagine, insieme ad attributi, come sesso, presenza di occhiali da sole e punti di riferimento del volto.

Il confronto facciale è il processo che consente di confrontare un volto con uno o più altri volti per valutarne la somiglianza. Utilizzando l'API CompareFaces, Amazon Rekognition Image consente di misurare la probabilità che i volti presenti in due immagini diverse appartengano alla stessa persona. L'API confronta un volto individuato nell'immagine originale con ogni volto rilevato nell'immagine di destinazione e restituisce un punteggio di somiglianza per ogni confronto. Per ogni faccia rilevata si avranno, anche, una cornice e un punteggio di affidabilità. Il confronto facciale può essere utilizzato anche per verificare l'identità di una persona confrontandola con foto presenti in archivio quasi in tempo reale.

3.2 Casi d'uso usati

Le demo di Amazon Rekognition consentono di effettuare test su immagini, esplorando diversi casi d'uso presi in considerazione. Come indicato in precedenza, in questo capitolo utilizzeremo l'analisi delle etichette, il confronto facciale e l'analisi facciale.

3.2.1 Analisi delle etichette

Testiamo la Figura 3.1 che è una tipica immagine illusoria in cui possiamo vedere vari scenari: o due visi di due persone anziane o tre persone. L'IA, però, non è in grado di vedere entrambe le cose, come infatti risulta dalla sua analisi nella Figura 3.2.

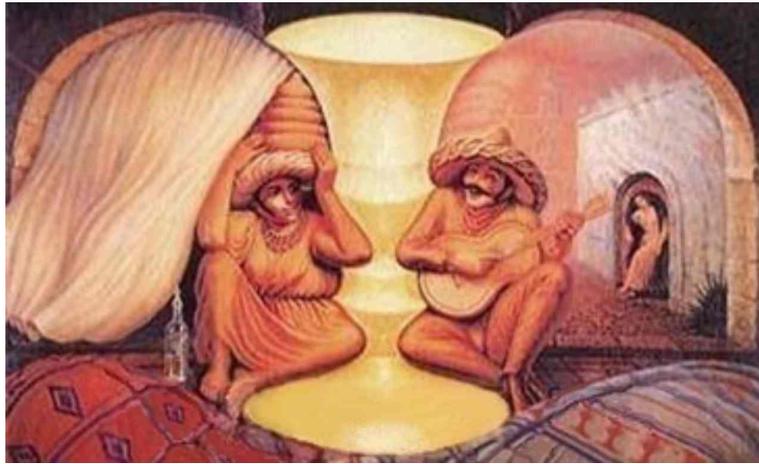


Figura 3.1: Immagine illusoria data in input ad Amazon Rekognition

▼ Risultati	
Art	98.6 %
Painting	97.1 %
Adult	93.7 %
Female	93.7 %
Person	93.7 %
Woman	93.7 %
Face	83 %
Head	83 %
Alien	82.2 %
Drawing	74.7 %
Beverage	65.3 %
Coffee	65.3 %

Coffee Cup	65.3 %
Angel	57.8 %
Modern Art	57.4 %
Architecture	55.6 %
Building	55.6 %
Monastery	55.6 %
Lamp	55.4 %

Mostra meno

Figura 3.2: Risultati forniti da Amazon Rekognition avendo in input l'immagine della Figura 3.1

In particolare, la Figura 3.2 ci dice che Amazon Rekognition ha individuato delle donne e delle facce, che sono risultati non del tutto sbagliati. Inoltre, l'IA, nonostante abbia delle percentuali di precisione abbastanza alte, ha individuato anche degli alieni, del caffè e degli angeli, cose visivamente sbagliate.

Abbiamo, poi, analizzato un altro dipinto, questa volta un dipinto impressionistico di Monet (Figura 3.3).

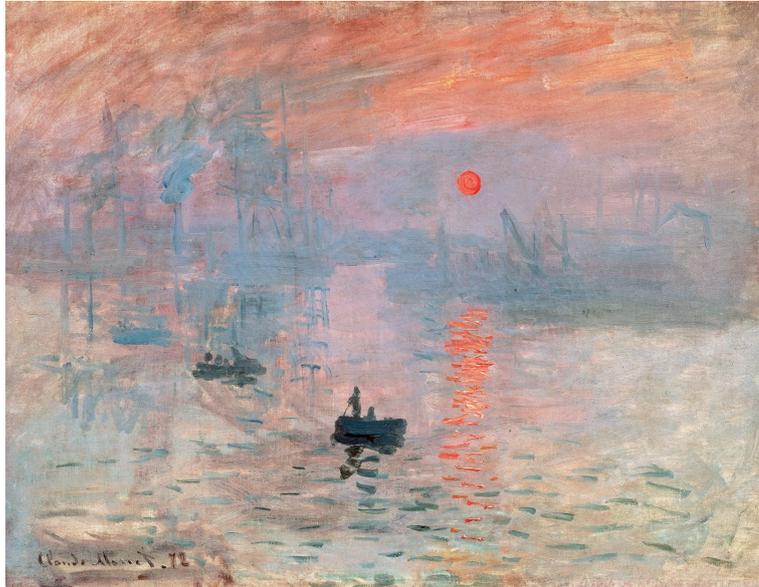


Figura 3.3: Impressione calar del sole immagine data in input ad Amazon Rekognition

Analizzando l'immagine (Figura 3.4) notiamo che ci sono molti elementi che non si percepiscono, come per esempio tutte le barche sul fondo. Come se non bastasse, l'IA ci dà anche delle informazioni sbagliate, come ad esempio la presenza di uccelli. Nonostante tutto, l'IA riesce ad identificare la Figura 3.4 come un dipinto raffigurante una scena all'aperto con al centro un veicolo.

▼ Risultati	
Art	99.9 %
Painting	99.9 %
Boat	81 %
Sailboat	81 %
Transportation	81 %
Vehicle	81 %
Outdoors	77 %
Nature	70.4 %

Animal	62.8 %
Bird	62.8 %
Watercraft	56.4 %
Cruiser	55.9 %
Military	55.9 %
Navy	55.9 %
Ship	55.9 %
Modern Art	55.7 %

Mostra meno

Figura 3.4: Risultati avendo in input l'immagine della Figura 3.3

3.2.2 Confronto facciale

Nel confronto facciale abbiamo analizzato due immagini, una in cui abbiamo due gemelle e vogliamo vedere se l'IA le riconosce come diverse. Abbiamo preso ad esempio le gemelle Olsen. Nella figura 3.5 notiamo che l'IA le identifica come la stessa persona al 99.9%.



Figura 3.5: Immagine delle Gemelle Olsen date come input ad Amazon Rekognition per il confronto facciale

Nella Figura 3.6 vediamo una drag queen da truccata e da struccata; l'IA la riconosce come la stessa persona al 92.4%.

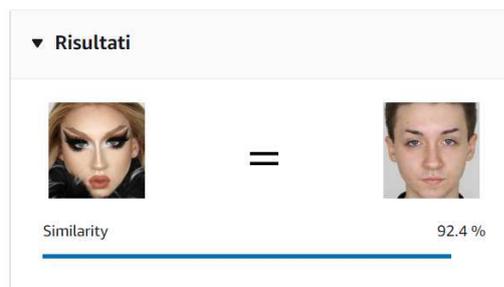


Figura 3.6: Immagini della stessa persona, con trucco e senza, data in input ad Amazon Rekognition per il confronto facciale

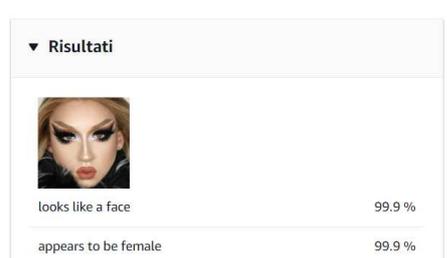


Figura 3.7: Risultato dell'analisi facciale fornito da Amazon Rekognition avendo in input la persona truccata della figura 3.6

La cosa più curiosa che possiamo osservare è che, nel caso in cui le immagini fossero analizzate singolarmente, l'IA risponderebbe affermando che la persona truccata nella Figura 3.6 è una donna (Figura 3.7), mentre, la stessa persona struccata verrebbe riconosciuta come uomo (Figura 3.8). Quindi, possiamo concludere che l'IA nonostante sia in grado di riconoscere la somiglianza tra due immagini, è stata allenata a distinguere il sesso maschile e femminile attraverso l'assenza/presenza di trucco.

▼ Risultati	
	
looks like a face	99.9 %
appears to be male	84.9 %

Figura 3.8: Risultato dell'analisi facciale fornito da Amazon Rekognition avendo in input la persona senza trucco della figura 3.6

3.2.3 Analisi facciale

Per quanto riguarda l'analisi facciale, abbiamo provato a vedere se riconosce le emozioni anche quando abbiamo delle espressioni particolari. Un esempio è quello di Albert Einstein al suo compleanno mentre fa una linguaccia (Figura 3.9).

	
looks like a face	99.7 %
appears to be male	95.9 %
age range	59 - 67 years old
not smiling	73.1 %
appears to be surprised	61.2 %
not wearing glasses	88.2 %
not wearing sunglasses	99.8 %
eyes are open	86 %
mouth is open	62.4 %
does not have a mustache	95.5 %
has a beard	91 %
face is not occluded	98.4 %

Figura 3.9: Risultati dell'analisi facciale dell'immagine data in input di Albert Einstein

Esaminando la Figura 3.9 possiamo notare, infatti, che Amazon Rekognition percepisce la sua espressione come confusa.

Come curiosità possiamo notare che l'IA non riconosce la presenza dei baffi (con una sicurezza del 95.5%).

Trascrizione speech-to-text

In questo capitolo vedremo cos'è Amazon Transcribe e come funziona. Vedremo, anche, i vari tipi di analisi tra cui l'analisi della tossicità del linguaggio e l'analisi della trascrizione medica. Infine proporremo delle analisi su alcuni test interessanti fatti tramite questo tool.

4.1 Cos'è Transcribe

Amazon Transcribe è un servizio di riconoscimento vocale automatico che permette di aggiungere a qualsiasi applicazione funzionalità di conversione da voce in testo per questo si chiama trascrizione speech-to-text. Le funzionalità di Transcribe consentono di inserire input audio, di migliorare la precisione nella trascrizione con la personalizzazione, di analizzare i contenuti nell'audio multicanale, suddividere il discorso dei singoli oratori. Esso, inoltre, analizza le chiamate dei call center e di filtra i contenuti per garantire la privacy degli utenti.

Si possono trascrivere i contenuti multimediali in tempo reale (streaming) oppure si possono trascrivere i file multimediali che si trovano in un bucket Amazon S3 (batch).

4.1.1 Come funziona

Amazon Transcribe utilizza modelli di machine learning per convertire la voce in testo. I metodi di trascrizione possono essere suddivisi in due categorie principali:

- *Trascrizioni batch*: trascrive i file multimediali che sono stati caricati in un bucket Amazon S3. Si può utilizzare AWS CLI, Console di gestione AWS e vari SDK AWS per le trascrizioni batch.
- *Trascrizioni in streaming*: trascrive i flussi multimediali in tempo reale. Si può utilizzare la Console di gestione AWS, WebSocket e vari SDK AWS per le trascrizioni in streaming.

Amazon Transcribe aggiunge automaticamente la punteggiatura a tutte le lingue supportate e mette in maiuscolo le parole in modo appropriato per le lingue che utilizzano la distinzione tra maiuscole e minuscole nei loro sistemi di scrittura. Le trascrizioni contengono dati sul contenuto trascritto, inclusi punteggi di confidenza e timestamp per ogni parola.

Per la maggior parte delle lingue, i numeri vengono trascritti nelle loro forme verbali. Tuttavia, se i propri contenuti multimediali sono in inglese o tedesco, Amazon Transcribe tratta i numeri in modo diverso a seconda del contesto in cui vengono utilizzati.

4.2 Tipi di analisi

Amazon Transcribe effettua diverse analisi, tra cui quella del linguaggio tossico e quella del linguaggio medico specifico. Di seguito analizzeremo entrambi nel dettaglio.

4.2.1 Amazon Transcribe Toxicity Detection

Amazon Transcribe Toxicity Detection sfrutta segnali audio e testuali per identificare e classificare contenuti tossici basati sulla voce in sette categorie, tra cui molestie sessuali, incitamento all'odio, minacce, abusi, volgarità, insulti ed immagini. Oltre al testo, Amazon Transcribe Toxicity Detection utilizza segnali vocali, come toni e intonazione, per individuare l'intento tossico nel parlato. Si tratta di un miglioramento rispetto ai sistemi di moderazione dei contenuti standard progettati per concentrarsi solo su termini specifici, senza tenere conto delle intenzioni.

Il rilevamento del parlato tossico è progettato per aiutare a moderare le piattaforme di social media che coinvolgono il dialogo peer-to-peer, come nei giochi online e nelle piattaforme di chat sociale. L'uso di discorsi tossici può essere profondamente dannoso per gli individui, nei gruppi di pari e nelle comunità. La segnalazione di linguaggio dannoso aiuta le organizzazioni a mantenere le conversazioni civili e a mantenere un ambiente online sicuro e inclusivo in cui gli utenti possano creare, condividere e partecipare liberamente.

Amazon Transcribe segnala e classifica i discorsi tossici riducendo al minimo il volume di dati che devono essere elaborati manualmente. Ciò consente ai moderatori dei contenuti di gestire il discorso in modo rapido ed efficiente sulle piattaforme.

Le categorie del discorso tossico includono:

- *Linguaggio volgare*: discorso che contiene parole, frasi o acronimi scortesi, volgari o offensivi.
- *Incitamento all'odio*: discorso che critica, insulta, denuncia o disumanizza una persona o un gruppo sulla base di un'identità (come razza, etnia, genere, religione, orientamento sessuale, abilità e origine nazionale).
- *Sessuale*: discorso che indica interesse, attività o eccitazione sessuale utilizzando riferimenti diretti o indiretti a parti del corpo, tratti fisici o sesso.
- *Insulti*: discorso che include un linguaggio umiliante, umiliante, beffardo, offensivo o sminuente. Questo tipo di linguaggio è anche etichettato come bullismo.
- *Violenza o minaccia*: discorso che include minacce che cercano di infliggere dolore, lesioni o ostilità verso una persona o un gruppo.
- *Grafica*: discorso che utilizza immagini visivamente descrittive e spiacevolmente vivide. Questo tipo di linguaggio è spesso intenzionalmente prolisso per amplificare il disagio del destinatario.
- *Molestie o abusi*: discorsi destinati a pregiudicare il benessere psicologico del destinatario, compresi termini umilianti e oggettivanti. Questo tipo di linguaggio è anche etichettato come molestia.

Il rilevamento della tossicità analizza i segmenti del discorso (il discorso tra pause naturali) e assegna punteggi di confidenza a questi segmenti. I punteggi di confidenza sono valori compresi tra 0 e 1. Un punteggio di confidenza più elevato indica una maggiore probabilità che la parola venga trascritta correttamente.

4.2.2 Amazon Transcribe Medical

Amazon Transcribe Medical opera secondo un modello di responsabilità condivisa, in base al quale AWS è responsabile della protezione dell'infrastruttura che esegue Amazon Transcribe Medical e l'utente è responsabile della gestione dei propri dati.

Amazon Transcribe Medical è un servizio di riconoscimento vocale automatico (ASR), disponibile solo in lingua inglese, progettato per i professionisti medici che desiderano trascrivere discorsi di carattere medico, come appunti dettati dal medico, monitoraggio della sicurezza dei farmaci, appuntamenti di telemedicina o conversazioni medico-paziente. Amazon Transcribe Medical è disponibile tramite streaming in tempo reale (tramite microfono) o trascrizione di un file caricato (batch).

4.3 Test svolti

Facendo dei test di trascrizione in tempo reale abbiamo analizzato alcuni ambiti, abbiamo provato a far trascrivere un semplice dialogo preso da un programma televisivo, abbiamo provato a far trascrivere un discorso in più lingue sia nello streaming che nella creazione di sottotitoli e, infine, abbiamo provato a far trascrivere il dettato della Divina Commedia per vedere se riuscisse a trascrivere anche il "volgare".

4.3.1 Programma televisivo

Per fare il test su un video di un programma televisivo, abbiamo usato un file audio salvato su Bucket S3 di AWS. In particolare, abbiamo creato un percorso per la trascrizione dei sottotitoli in streaming e una volta terminato il processo, abbiamo inserito il file .SRT contenente i sottotitoli nel video prodotto in VLC Media. Il video che abbiamo usato per svolgere questo test è stato trovato su YouTube; è un programma televisivo, che va in onda su Rete4, si chiama "Stasera Italia". Vediamo che nel video non ci sono tanti errori nei sottotitoli; notiamo, però, che mentre parlano più persone l'IA riesce a distinguere i vari interlocutori. Tuttavia, quando le varie persone del video parlano in maniera confusa e le voci si sovrappongono, l'IA si trova in difficoltà nel trascrivere (Figura 4.1), e alcune volte possiamo notare l'assenza dei sottotitoli.



Figura 4.1: Parte del programma di Rete4 in cui vediamo i sottotitoli separati in base agli interlocutori

4.3.2 Trascrizione a più lingue

In questa sezione testiamo l'IA per vedere se riesce a trascrivere correttamente quando abbiamo un discorso fatto in più lingue, in questo caso in italiano e in inglese. Per studiare quest'ambito abbiamo fatto una registrazione in cui viene descritto in italiano cosa sta avvenendo nel video, che rappresenta una scena di Harry Potter e la pietra Filosofale, preso su YouTube in lingua inglese.

Utilizziamo anche questa volta Amazon S3 per produrre i sottotitoli al video; prendiamo il file .SRT e lo inseriamo nel lettore multimediale VLC e notiamo che i sottotitoli sono precisi e non abbiamo nessun errore. Proviamo, poi, a far riprodurre il video per avere una trascrizione in tempo reale, e qui incominciamo a notare degli errori fatti dall'IA quando abbiamo il passaggio dalla lingua inglese a quella italiana. Come possiamo vedere in Figura 4.2, l'IA non riesce a percepire subito che la lingua è diversa, e quindi scrive delle parole casuali in inglese.

Nella Figura 4.3 possiamo notare il sottotitolo in italiano, che è corretto, e nella Figura 4.2, abbiamo lo stesso spezzone del discorso; nella riga evidenziata notiamo che la trascrizione è completamente diversa.

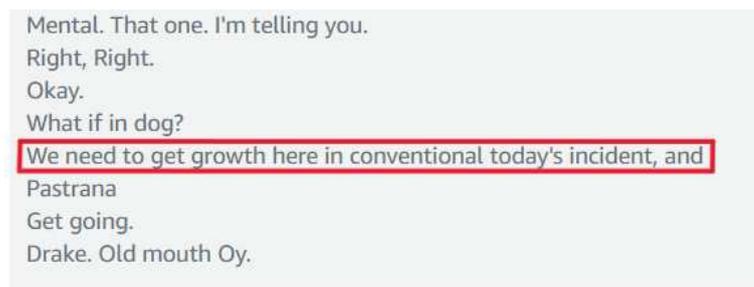


Figura 4.2: Esempio di trascrizione in tempo reale di due lingue



Figura 4.3: Spezzone del video in cui avviene lo scambio di lingue

4.3.3 Trascrizione dei dialetti

Come ultimo test abbiamo considerato la lettura del Canto 1 dell'Inferno della Divina Commedia di Dante Alighieri effettuata da Gassman; tale video è stato ricavato da YouTube.

Output della trascrizione

Nel mezzo del cammin di nostra vita.
Mi ritrovai
Per una silva oscura.
Che la diritta via
Era solo la verità. Hai quanto qual era la cosa dura?
Questa sera il vaso in faccia e ha fornite.
Che nel pensier rinnova la paura.
Tant'è amara che poco è più morte.
Ma per trattare del ben chi vi trovai
Diro' dell'altro le cose chivo scorte

Figura 4.4: Trascrizione della Divina Commedia

Nel mezzo del cammin di nostra vita
mi ritrovai per una selva oscura,
ché la diritta via era smarrita.

Ahi quanto a dir qual era è cosa dura
esta selva selvaggia e aspra e forte
che nel pensier rinova la paura!

Tant'è amara che poco è più morte;
ma per trattar del ben ch'i' vi trovai,
dirò de l'altre cose ch'i' v' ho scorte.

Figura 4.5: Parte della Divina Commedia originale

Analizzando le Figure 4.4 e 4.5 possiamo notare che nella Figura 4.4 ci sono degli errori, anche se di poca importanza.

Notiamo, anche, che l'IA ha sbagliato la punteggiatura e qualche termine che è molto diverso dall'italiano di oggi. Per il resto possiamo dire che la trascrizione fatta dall'IA non è andata molto lontana dalla realtà.

Trascrizione text-to-speech

In questo capitolo vedremo cos'è Amazon Polly e le sue funzionalità. Successivamente, esporremo il linguaggio SSML con i suoi tag più particolari.

5.1 Cos'è Polly

Amazon Polly fornisce un'API che permette di integrare rapidamente la sintesi vocale in un'applicazione. È sufficiente inviare alla relativa API il testo da convertire in voce e Amazon Polly restituirà immediatamente all'applicazione un flusso audio da riprodurre direttamente o da memorizzare in formato audio standard, ad esempio MP3.

Amazon Polly include decine di voci naturali e supporta diverse lingue, per permettere di selezionare la voce più adatta e distribuire le proprie applicazioni con risposte vocali in diversi paesi. Oltre alle voci TTS e NTTS (Neural Text-to-Speech), Amazon Polly offre voci in forma lunga che migliorano la qualità del discorso e sono più naturali e simili a quelle umane.

5.1.1 Funzionalità

Amazon Polly supporta lo Speech Synthesis Markup Language (SSML), un linguaggio di markup W3C standard, basato su XML, per la sintesi vocale. I tag di SSML permettono di variare progressione, enfasi e intonazione. I tag SSML personalizzati di Amazon offrono opzioni uniche, come la possibilità di far parlare certe voci nello stile di conversazione Newscaster. Grazie a questo livello di flessibilità, è possibile creare voci estremamente verosimili in grado di attirare e mantenere l'attenzione degli utenti.

È possibile utilizzare Amazon Polly per generare parlato da testo semplice o da documenti SSML. L'utilizzo del testo ottimizzato SSML offre un controllo aggiuntivo sul modo in cui Amazon Polly genera il parlato dal testo fornito.

Questo linguaggio include varie opzioni, ad esempio:

- includere una lunga pausa nel testo;
- modificare la velocità o il tono della voce;
- enfatizzare parole o frasi specifiche;
- utilizzare la pronuncia fonetica;

- sussurrare;
- inserire suoni respiratori;
- utilizzare lo stile di conversazione del giornalista.

Quando si utilizza SSML, sono presenti diversi caratteri riservati che richiedono un trattamento speciale. Questo perché SSML utilizza questi caratteri come parte del suo codice.

5.1.2 Tipi di analisi

Con i dizionari personalizzati di Amazon Polly, può essere modificata la pronuncia di singole parole, ad esempio acronimi, presi da altre lingue e neologismi (ad esempio "ROTFL" oppure "C'est la vie" in lingue diverse dal francese). Per personalizzare queste pronunce, è sufficiente caricare un file XML con le voci lessicali.

È possibile creare voci bilingue accentate utilizzando qualsiasi voce Amazon Polly, ma solo quando si utilizzano tag SSML.

Normalmente, tutte le parole nel testo immesso vengono pronunciate nella lingua predefinita della voce specificata che si sta utilizzando.

Amazon Polly dispone di un sistema Neural TTS (NTTS) in grado di produrre voci di qualità ancora superiore rispetto alle voci standard. Il sistema NTTS produce le voci di sintesi vocale più naturali e simili a quelle umane possibili.

Le voci TTS standard utilizzano la sintesi concatenativa, questo metodo mette insieme i fonemi del parlato registrato, producendo un parlato sintetizzato dal suono molto naturale. Tuttavia, le inevitabili variazioni nel parlato e le tecniche utilizzate per segmentare le forme d'onda limitano la qualità complessiva che si riesce a raggiungere.

Il sistema TTS neurale di Amazon Polly non utilizza la sintesi concatenativa standard per produrre il parlato. Esso ha due parti:

- *Una rete neurale* che converte una sequenza di fonemi, le unità linguistiche più basilari, in una sequenza di spettrogrammi, che sono istantanee dei livelli di energia in diverse bande di frequenza.
- *Un vocoder*, che converte gli spettrogrammi in un segnale audio continuo.

Il primo componente del sistema TTS neurale è un modello sequenza-sequenza. Questo modello non crea i suoi risultati esclusivamente dall'input corrispondente, ma considera, anche, la sequenza degli elementi dell'input. Il modello sceglie gli spettrogrammi emessi in modo che le loro bande di frequenza enfatizzino le caratteristiche acustiche che il cervello umano utilizza durante l'elaborazione del parlato.

L'output di questo modello passa, quindi, a un vocoder neurale. Ciò converte gli spettrogrammi in forme d'onda del parlato. Se addestrato sui grandi set di dati utilizzati per costruire sistemi di sintesi concatenativa generici, questo approccio sequenza per sequenza produrrà voci di qualità superiore e dal suono più naturale.

5.2 Test di trascrizione text-to-speech

Effettuando dei test nella demo di Amazon Polly, esploriamo e analizziamo alcuni tag particolari, come il cambio della pronuncia avendo sempre la voce selezionata su Bianca (italiano) per vedere il modo in cui vengono pronunciati dei tipi speciali di parole.

5.2.1 Utilizzo del tag <lang>

In questa sezione utilizziamo il tag <lang> di SSML per fare degli esempi. Quando si utilizza il tag <lang> in SSML, è importante notare che esso specifica la lingua della porzione di testo racchiusa al suo interno, utilizzando l'attributo `xml:lang`, ma non modifica la lingua della voce utilizzata.

Quindi, se si utilizza il tag <lang> per indicare una lingua diversa, ma la voce selezionata è configurata per un'altra lingua, potrebbe verificarsi un problema di pronuncia.

Per prima cosa selezioniamo Bianca (parla italiano) come voce da utilizzare.

Se non utilizziamo il tag <lang> e scriviamo il testo in una lingua diversa da quella della voce selezionata, la voce pronuncerà le parole seguendo la lingua della voce impostata.

In altre parole, la lingua della voce (in questo caso utilizziamo Bianca che parla italiano) guiderà la pronuncia del testo, indipendentemente dalla lingua effettiva del testo scritto.

```
<say-as>  
  Oggi esploreremo diversi argomenti, inizieremo parlando  
  'en español', poi passeremo a discutere di alcuni  
  concetti 'auf Deutsch'.  
</say-as>
```

Se la voce selezionata è configurata per l'italiano, pronuncerà il testo seguendo le regole di pronuncia italiane, anche se le parole "an español" e "auf Deutsch" sono in francese e in tedesco, rispettivamente.

Procediamo ora con l'utilizzo del tag <lang>:

```
<say-as>  
  Nel romanzo, i personaggi si trovano in una città multiculturale,  
  mentre camminano per le strade, sentono lingue diverse.  
  <lang xml:lang="en-US">And there will also be a conversation in  
  English</lang>,  
  <lang xml:lang="fr-FR">Et il y aura aussi une conversation en  
  français</lang>.  
  La varietà di lingue riflette l'atmosfera cosmopolita della città.  
</say-as>
```

La voce selezionata per parlare l'italiano legge le frasi "And there will also be a conversation in English" e "Et il y aura aussi une conversation en français" nella lingua specificata dal tag ma con l'accento italiano, poiché la lingua della voce rimane quella impostata nell'identificativo della voce.

5.2.2 Utilizzo tag <say-as>

Con il tag <say-as>, siamo in grado di plasmare in modo più dettagliato la pronuncia di Amazon Polly. Questo tag, ad eccezione dell'opzione "characters", è compatibile con una varietà di formati, tra cui quelli a lungo termine TTS, neurali e standard. Tuttavia, è essenziale notare una sottigliezza: se, durante l'esecuzione, Amazon Polly rileva il tag <say-as> con l'opzione "characters" mentre utilizza una voce neurale, la frase coinvolta verrà sintetizzata con la voce standard associata.

L'attributo chiave qui è "interpret-as", il quale, abbinato al tag <say-as>, ci consente di istruire Amazon Polly su come affrontare caratteri, parole e numeri specifici. Questa funzionalità è particolarmente utile per fornire un contesto aggiuntivo che elimina dubbi su come Amazon Polly dovrebbe modulare il testo, assicurando una pronuncia più accurata e personalizzata.

Il tag `<say-as>` si basa sull'utilizzo dell'attributo `"interpret-as"`, il quale offre una varietà di valori possibili, ciascuno seguendo la stessa sintassi. La sintassi è:

```
<say-as interpret-as="value">[text to be interpreted]</say-as>
```

fornendo un modo versatile per modulare il comportamento di Amazon Polly nella pronuncia del testo.

I diversi valori disponibili dell'attributo `"interpret-as"`, che consentono di personalizzare la pronuncia del testo in vari modi, sono i seguenti:

- `characters` o `spell-out`: ogni lettera del testo viene pronunciata individualmente; ciò crea un effetto simile a quando si recitano le lettere dell'alfabeto come "a-b-c";
- `cardinal` o `number`: il testo numerico viene interpretato come un numero cardinale, ad esempio, "1.234";
- `ordinal`: il testo numerico viene interpretato come un numero ordinale, come in "1.234esimo";
- `digits`: ciascuna cifra del testo numerico viene pronunciata individualmente, creando una sequenza come "1-2-3-4";
- `fraction`: il testo numerico viene interpretato come una frazione, la sintassi utilizzata è:

```
cardinal number/cardinal number
```

Questa azione supporta sia frazioni comuni, come "3/20", che frazioni miste, come "2 ½";

- `unit`: il testo numerico viene interpretato come una misura; ciò consente di esprimere quantità seguite da un'unità, ad esempio "1/2inch" o "1meter";
- `date`: il testo viene interpretato come una data, il cui formato è specificato tramite l'attributo `"format"`;
- `time`: il testo numerico viene interpretato come una durata in minuti e secondi, come ad esempio, "1'21";
- `address`: il testo viene interpretato come parte di un indirizzo stradale;
- `expletive`: viene emesso un segnale acustico per il contenuto incluso nel tag;
- `telephone`: il testo numerico viene interpretato come un numero telefonico, sia di 7 o 10 cifre, e può gestire anche interni telefonici, come, ad esempio, "2025551212x345".

Tra i vari valori sopraelencati vediamo uno particolare che è il 'date'. Quando imponiamo `"interpret-as"` su `date` dobbiamo specificare prima il formato della data. La sintassi è la seguente:

```
<say-as interpret-as="date" format="format">[date]</say-as>
```

I formati che si possono utilizzare sono i seguenti:

- `mdy`: mese, giorno e anno;

- dmy: giorno, mese e anno;
- ymd: anno, mese e giorno;
- md: mese e giorno;
- dm: giorno e mese;
- ym: anno e mese;
- my: mese e anno;
- d: giorno;
- m: mese;
- y: anno;
- `yyyymmdd`: anno, mese e giorno. Se si utilizza tale formato, possiamo fare in modo che Amazon Polly salti parti della data utilizzando i punti interrogativi e 'Format' non è necessario.

Nei successivi esempi vediamo dei tipici utilizzi del valore "date":

```
<say-as interpret-as="date" format="dmy">12-09-2020</say-as>
</speak>
```

Amazon Polly leggerà questa data come 12 Settembre 2020.

Se inserissimo dei punti interrogativi all'interno della data può essere evitata la specifica del suo formato:

```
<say-as interpret-as="date">????0922</say-as>
```

Questa volta l'IA leggerà questa data come 22 Settembre; in questo esempio abbiamo aggiunto dei punti interrogativi al posto dell'anno; l'IA non è entrata in confusione e ha letto la restante data con mese e giorno.

In questa sezione parleremo dei tool utilizzati dalla piattaforma di Amazon Web Service, descrivendo la loro utilità.

6.1 Discussione sul riconoscimento delle immagini

L'avvento dell'Intelligenza Artificiale ha portato con sé un notevole progresso nel campo del riconoscimento delle immagini, rivoluzionando molteplici settori e influenzando la nostra vita quotidiana. Questa tecnologia avanzata ha reso possibile l'analisi e l'interpretazione di contenuti visivi con una precisione sempre maggiore.

Il riconoscimento delle immagini attraverso l'IA si presenta come un alleato potente nell'automazione di compiti complessi; la capacità di analizzare grandi quantità di dati visivi facilita la ricerca, l'analisi e l'estrazione di informazioni significative.

L'applicazione dell'IA in settori come la medicina, la sicurezza e l'automazione veicolare promette di migliorare la nostra vita quotidiana, rendendola più sicura ed efficiente.

Tuttavia, questa potente tecnologia non è esente da preoccupazioni etiche. Il rischio di violazioni della privacy emerge come una delle principali preoccupazioni, poiché il riconoscimento delle immagini potrebbe essere utilizzato senza il consenso delle persone coinvolte. Questo solleva interrogativi fondamentali sulla gestione responsabile dei dati visivi e sulla necessità di definire chiaramente i limiti etici di tali sistemi.

Parallelamente, il riconoscimento facciale è emerso come una componente distintiva, suscitando dibattiti accesi su sicurezza, controllo degli accessi e praticità. La capacità di identificare un individuo sulla base delle sue caratteristiche facciali trova applicazione in settori quali la sicurezza pubblica, il controllo degli accessi e l'autenticazione digitale. Ciò ha portato ad un miglioramento nella gestione della sicurezza, ma, al contempo, ha generato discussioni sulla sorveglianza di massa e sulla potenziale minaccia alla privacy individuale.

In questo scenario, è imperativo bilanciare l'innovazione tecnologica con un approccio etico e responsabile.

Con l'utilizzo di Amazon Rekognition attraverso la piattaforma Amazon Web Service abbiamo visto come vengono identificati e analizzati i volti nelle immagini, rilevando caratteristiche facciali e consentendo l'identificazione di individui.

Abbiamo notato come questa demo sia molto utile ed efficace, tenendo conto anche di alcuni errori che questa IA può commettere.

6.2 Discussione sulla trascrizione del parlato

La trascrizione automatica di grandi quantità di contenuti audio, grazie all'IA, è diventata una risorsa inestimabile in vari contesti.

Conferenze, seminari, riunioni aziendali e altri eventi possono essere documentati in modo efficiente, consentendo una conservazione dettagliata delle conversazioni. Ciò si traduce in un notevole risparmio di tempo e sforzi, rendendo più accessibili e gestibili grandi volumi di informazioni vocali.

L'integrazione di queste soluzioni con altre applicazioni e servizi ha ulteriormente ampliato le possibilità. L'automazione dei flussi di lavoro, la facilità di condivisione e archiviazione dei testi trascritti e la ricerca rapida all'interno di vasti archivi audio rappresentano solo alcune delle sfaccettature di questa sinergia tecnologica.

Tuttavia, nonostante i benefici evidenti, sorgono importanti questioni etiche e pratiche. La privacy diventa un tema cruciale, poiché le conversazioni e i discorsi vengono trascritti e archiviati. La necessità di regolamentazioni chiare e pratiche per la gestione responsabile dei dati vocali è imperativa per garantire che i vantaggi dell'IA nella trascrizione del parlato non vengano compromessi da preoccupazioni legate alla sicurezza e alla riservatezza.

Con l'utilizzo di Amazon Transcribe, attraverso la piattaforma di Amazon Web Service, abbiamo visto l'utilità di questa tecnologia per convertire in modo automatico registrazioni audio e video in testo scritto.

La creazione automatica di sottotitoli per contenuti multimediali rappresenta un'altra utilità significativa di Amazon Transcribe. Questa funzionalità mostra la sua utilità con la presenza di vari interlocutori o quando abbiamo un dialogo contenente più di una lingua parlata, ma ha delle difficoltà quando abbiamo più interlocutori che si sovrappongono fra loro nei dialoghi.

In conclusione, però, questa IA non solo rende i contenuti più accessibili a una vasta audience, ma migliora anche l'esperienza utente per coloro che preferiscono o necessitano di sottotitoli per comprendere appieno il contenuto dei video e dei materiali multimediali.

6.3 Discussione sulla riproduzione vocale di un testo

La sintesi vocale, comunemente conosciuta come Text-to-Speech (TTS), è una tecnologia all'avanguardia che ha il potere di trasformare il testo scritto in un flusso di parole fluide e naturali, aprendo le porte a una vasta gamma di possibilità e cambiamenti significativi nei modi in cui consumiamo le informazioni e interagiamo con esse.

Al centro di questa rivoluzione c'è l'obiettivo di creare voci sintetiche che non solo riproducano in modo preciso il contenuto testuale, ma che siano anche in grado di catturare la naturalezza e l'intonazione delle voci umane.

Le applicazioni pratiche di questa tecnologia sono idealmente infinite. Dai sistemi di navigazione vocale, che guidano i viaggiatori attraverso strade sconosciute, dagli assistenti virtuali, che ci forniscono informazioni in tempo reale, fino ai lettori di schermo che rendono accessibili i contenuti scritti alle persone con disabilità visive, l'IA Text-to-Speech ha reso possibile un mondo di nuove interazioni ed opportunità.

Analizzando Amazon Polly, attraverso Amazon Web Service, abbiamo scoperto che, grazie al linguaggio SSML, esso è in grado di variare in maniera parlata il contenuto che viene scritto con pause ed enfasi opportune.

In conclusione, possiamo affermare che la continua evoluzione di questa tecnologia porta all'incessante ridefinizione della frontiera tra l'essere artificiale e il naturale.

Nel corso di questa tesi, abbiamo condotto un'analisi approfondita della tecnologia dell'Intelligenza Artificiale (IA), che ha rivoluzionato il panorama tecnologico, ridefinendo fondamentalmente il modo in cui concepiamo l'automazione, l'apprendimento e l'interazione. Ci siamo soffermati sulla definizione stessa di Intelligenza Artificiale, esplorando le sue capacità di risolvere problemi complessi e cercando di emulare le funzioni del cervello umano. La storia dell'evoluzione dell'IA, dai suoi albori fino ai giorni nostri, ci ha mostrato come questa disciplina si è sviluppata e ha maturato le sue potenzialità.

Abbiamo dedicato particolare attenzione alla capacità dell'IA di apprendere in modo automatico dai dati e migliorare costantemente le sue prestazioni. Questo processo di apprendimento automatico costituisce uno dei pilastri fondamentali che ha contribuito a rendere l'IA una forza motrice nell'innovazione tecnologica.

Successivamente, il nostro focus si è spostato sulla piattaforma di cloud computing di Amazon Web Service (AWS), che si è rivelata molto importante per l'applicazione pratica dell'IA. Attraverso i numerosi servizi dedicati all'IA offerti da AWS abbiamo esplorato le possibilità che questa piattaforma offre, dall'elaborazione di dati all'implementazione di modelli avanzati di Machine Learning.

Ci siamo soffermati sull'analisi di tre specifici strumenti di AWS: Rekognition, Transcribe e Polly. Abbiamo constatato che ognuno di essi fornisce funzionalità specializzate che possono essere integrate in diversi contesti applicativi, aprendo anche nuove prospettive per l'innovazione in settori come l'analisi dei dati, la sicurezza informatica e l'automazione dei processi aziendali.

Tuttavia, nonostante i successi e le applicazioni concrete raggiunte fino ad oggi, abbiamo sottolineato che l'evoluzione dell'IA è un percorso in costante sviluppo. Infatti, guardando al futuro, abbiamo identificato l'integrazione di etica e responsabilità come elementi chiave nell'implementazione dell'IA. L'attenzione all'impatto sociale, la trasparenza e la protezione della privacy sono diventate, e continueranno a essere, aspetti cruciali per garantire una coesistenza armoniosa tra l'IA e la società.

In conclusione, prevediamo che l'IA continuerà a estendersi in nuovi settori, offrendo soluzioni sempre più sofisticate e personalizzate e che la ricerca di un equilibrio tra innovazione tecnologica e considerazioni etiche rimarrà al centro del suo futuro sviluppo.

- ASIMOV, I. (1950), *Io, robot*, Mondadori.
- BOSTROM, N. (2023), *Superintelligenza. Tendenze, pericoli, strategie*, Bollati Boringhieri.
- FLORIDI, L. (2022a), *Etica dell'intelligenza artificiale. Sviluppi, opportunità, sfide*, Cortina Raffaello.
- FLORIDI, L. (2022b), *Irriducibile. La coscienza, la vita. i computer e la nostra natura*, Cortina Raffaello.
- MARMO, R. (2021), *Algoritmi per l'intelligenza artificiale. Progettazione dell'algoritmo, dati e machine learning, neural network, deep learning*, Hoepli.
- PETER NORVIG, S. R. (2021), *Artificial Intelligence: A Modern Approach, Global Edition*, Athenaem Uitgeverij.
- RUSSELL, S. (2020), *Human Compatible: Artificial Intelligence and the Problem of Control*, Penguin.
- SERVICES, A. W. (2020), *Amazon Web Services A Complete Guide - 2021 Edition*.
- STUART J. RUSSELL, P. N. (2021), *Intelligenza artificiale. Un approccio moderno*, Pearson.
- TEGMARK, M. (2018), *Life 3.0: Being Human in the Age of Artificial Intelligence*, Penguin.
- WHITEPAPERS, A. (2016), *Cloudera Enterprise Data Hub on AWS*.

Siti web consultati

- Google Cloud – <https://cloud.google.com/?hl=it>
- AWS Documentation – <https://docs.aws.amazon.com/index.html>
- Wikipedia – www.wikipedia.org
- OpenAI – <https://openai.com>
- Microsoft Learn – <https://learn.microsoft.com>

- AI4Business – <https://www.ai4business.it>
- Andrea Minini – <https://www.andreaminini.com>
- Microsoft Azure – <http://azure.microsoft.com>
- IBM – www.ibm.com
- Treccani – www.treccani.it

Ringraziamenti

Il primo ringraziamento va al Prof Ursino per la sua guida, il suo sostegno e la sua completa disponibilità. I suoi consigli e la pazienza dimostrata durante questo percorso di ricerca sono stati fondamentali per il successo di questo lavoro.

Devo ringraziare anche, in particolar modo, i miei genitori, mio fratello e Martina per il loro sostegno. Il vostro incoraggiamento e la vostra costante fiducia sono stati, per me, fondamentali per questo percorso accademico.

Durante questo viaggio universitario ho avuto la fortuna di essere circondata da amici straordinari, i quali hanno reso il mio percorso meno aspro.

Ringrazio, quindi, tutti i miei amici di 'Beethoven' con i quali mi sono divertita tra una lezione e l'altra, con le competizioni di briscola e con varie sessioni di gossip. Un ringraziamento speciale va ad Amir, Luca, Ardu, Leonardo, Lorenzo, Matteo, Michele, Nicola, Riccardo, Alan, Francesco, Ale e Piccia. Ognuno di voi è stata una persona meravigliosa, e sono felice di aver avuto la fortuna di incontrarvi.

Ringrazio le mie coinquiline: Jing, Michela, Erika, Alice e Chiara, per aver reso accogliente una casa che non lo era.

Jing per essere stata come una sorella maggiore per me e, soprattutto, per avermi insegnato a preparare i ravioli cinesi.

Michela ed Erika per aver contribuito a rendere molti mercoledì davvero speciali e unici.

Ringrazio tutti gli altri amici che ho incontrato durante questi anni e con cui ho ricordi speciali: Claudio e Laura Carli.

Vorrei ringraziare anche la famiglia del mio fidanzato per avermi accolto ed essersi comportati come se fossi stata parte di loro.

Per ultimo lascio il ringraziamento più speciale al mio ragazzo Gabriel, con cui ho passato i momenti migliori e i più felici di questi ultimi anni, che mi ha aiutato a rialzarmi in momenti di grande difficoltà e che mi ha sempre dato tanta forza per andare avanti. A Gabriel devo dire: "Grazie di cuore per tutto, senza di te non ce l'avrei mai fatta".