UNIVERSITÀ POLITECNICA DELLE MARCHE
FACOLTÀ DI INGEGNERIA
Dipartimento di Ingegneria dell'Informazione
Corso di Laurea in Ingegneria Informatica e dell'Automazione

TESI DI LAUREA

Realizzazione di case study nel contesto dell'Intelligenza Artificiale: riconoscimento di immagini, traduzione, estrazione di testi

Realization of case studies in the Artificial Intelligence context: image recognition, translation and text extraction

Relatore

Prof. Domenico Ursino

Candidato

Amir Othmani

ANNO ACCADEMICO 2022-2023

Se ci si aspetta che la macchina sia infallibile, allora essa non può essere anche intelligente.

Alan Turing

Sommario

L'Intelligenza Artificiale (IA) è una tecnologia rivoluzionaria che sta rapidamente trasformando il nostro mondo. La sua capacità di apprendimento continuo consente ad essa di adattarsi e migliorare nel tempo, aprendo la strada a nuove applicazioni e innovazioni. In questa tesi viene data una panoramica sull'IA a partire dalle sue radici storiche, per poi discutere come si è evoluta nel tempo e diventare così influente attualmente. Successivamente, vengono esaminati alcuni case study di applicazione dell'IA e vengono discussi i risultati che essa ha prodotto. L'obiettivo della tesi è mostrare le potenzialità di questa tecnologia e i vastissimi campi di applicazione, evidenziandone anche i principali limiti, in modo da poter comprendere come e quanto può ancora migliorare.

Keyword: Intelligenza Artificiale, Reti neurali, Machine Learning, Deep Learning, Amazon Web Services, Servizi cloud, Riconoscimento di immagini, Traduzione automatica, Big Data, Estrazione di testi

Indice

Elenco delle figure

Introduzione

Negli ultimi anni l'Intelligenza Artificiale (IA) è diventata sempre più pervasiva, trovando continuamente nuovi campi di applicazione e assistendoci a svolgere molteplici compiti. L'IA è un campo di studi per sua natura interdisciplinare e trae concetti da numerose discipline, quali informatica, neuroscienze, logica formale, statistica, filosofia, lingustica, economia e ricerca operativa. L'obiettivo di questi studi è quello di fornire una rappresentazione di come funziona il ragionamento umano e trovare il modo di applicarlo alle macchine.

Nella storia furono tentati molteplici approcci, che si riconducono principalmente a due filoni: uno è quello basato sui sistemi esperti, che, attraverso una serie di principi ben definiti, svolgono il loro lavoro effettuando inferenze, seguendo una logica simbolica; l'altro filone è basato sulle reti neurali, che, invece, svolgono il loro lavoro apprendendo in maniera empirica, dopo lunghi periodi di addestramento.

Benchè l'interesse sull'IA sia recente, in realtà, gli studi in merito cominciarono dagli anni '50, quando Alan Turing propose una prima definizione di Intelligenza Artificiale attraverso il "Test di Turing"; dopodiché, tra gli anni '50 e gli anni '60, si assistette a un periodo di grande entusiasmo in cui si credette che si potesse costruire una macchina per risolvere problemi di carattere generale; tuttavia, tra gli anni '70 e '80, questo entusiasmo si fermò e in questo periodo molti dei finanziamenti alle ricerce in ambito IA vennero ritirati. Dagli anni '90 in poi, soprattutto grazie alle novità introdotte dalle reti neurali, l'IA ha ricominciato ad attirare interesse, grazie ai risultati che è stata in grado di produrre. Attualmente, l'IA è capace di adattarsi a compiti molto variegati tra loro, come l'analisi dei dati, la robotica, l'automazione industriale, la chirurgia a distanza, l'assistenza virtuale, e molti altri ancora.

L'obiettivo di questa tesi consiste nel fornire una spiegazione sintetica, ma quanto più efficace possibile, della lunga storia che c'è stata dietro alla diffusione dell'IA, fissando, di volta in volta, tutti i concetti fondamentali alla base del suo funzionamento, in modo tale da fornire un quadro quanto più completo possibile. Inoltre, si cercherà di spiegare come i servizi offerti da Amazon Web Services (AWS) permettano di usufruire dell'Intelligenza Artificiale utilizzando le risorse fornite da AWS stessa, tramite le infrastrutture cloud che mette a disposizione. Infine, con l'analisi dei case study, si forniranno alcuni esempi di applicazione dell'IA, valutando gli ambiti in cui essa può essere utilizzata e i vantaggi che offrirebbero, e osservando i risultati che essa produce, considerando sia i successi che gli insuccessi, in modo da poter trarre qualche spunto di miglioramento per applicazioni future.

La presente tesi è composta da sette capitoli strutturati come di seguito specificato:

- Nel Capitolo 1 sarà introdotto il concetto di Intelligenza Artificiale e verrà descritta la sua evoluzione nella storia. In seguito, verrano descritte l'IA simbolica, con la logica che sta alla base di essa e le tecniche con cui si riproducono i ragionamenti, e l'IA basata su

reti neurali, approfondendo la loro struttura a strati e gli algoritmi di Machine Learning e Deep Learning che stanno alla base.

- Nel Capitolo 2 verranno descritte le caratteristiche di AWS e la vasta gamma di servizi offerti e verranno illustrati i vantaggi che il cloud offre, quali scalabilità, affidabilità, sicurezza e flessibilità.
- Nel Capitolo 3 verrà fornita una panoramica sul riconoscimento di immagini e si effettueranno dei test tramite Amazon Rekognition.
- Nel Capitolo 4 verrà spiegata la traduzione automatica, con un cenno sulla sua evoluzione, per poi effettuare test con Amazon Translate e proporre un confronto con DeepL.
- Nel Capitolo 5 verranno descritti il tema della digitalizzazione dei documenti e quello dei Big Data, dopodiché si effettueranno alcuni test con Amazon Textract.
- Nel Capitolo 6 verranno discussi i risultati ottenuti.
- Nel Capitolo 7 verranno tratte le conclusioni e verranno delineati alcuni possibili sviluppi futuri.

CAPITOLO 1

Introduzione all'Intelligenza Artificiale

In questo capitolo verrà trattato il concetto di Intelligenza Artificiale (IA), ovvero ci si occuperà di come si è originato questo campo di studi e come si è sviluppato nel tempo fino a diventare quello che conosciamo oggi. In particolare si vedranno quali sono i fondamenti teorici che hanno permesso all'IA di diventare rilevante per l'epoca moderna, quali sono le caratteristiche chiave alla base di questa tecnologia e quali effetti ha sul nostro modo di approcciare all'informatica.

1.1 Nascita del concetto di IA

Gli studi sull'Intelligenza Artificiale nascono fondamentalmente perché ci si è chiesto se fosse possibile costruire delle macchine che potessero "pensare" e comportarsi come gli esseri umani. L'obiettivo era di creare sistemi in grado di risolvere problemi e compiere attività che, normalmente, richiedono l'intervento umano.

Le potenzialità dell'IA sono vaste, visto che è in grado di occuparsi di molte attività, a partire da quelle generali (come l'apprendimento, il ragionamento, la percezione, etc.) fino a quelle specifiche, come giocare a scacchi, dimostrare teoremi matematici, scrivere poesie, guidare un'automobile o diagnosticare delle malattie.

In effetti, l'IA è stata a sua volta influenzata da tanti campi di studio, oltre ovviamente all'informatica, la filosofia, la matematica, l'economia, le neuroscienze, la psicologia, la cibernetica, le scienze cognitive e la linguistica.

1.1.1 Macchine di Turing

Uno dei primi contributi in tal senso fu portato da Alan Turing (1912 - 1954), considerato uno dei padri dell'informatica.

Il suo più grande contributo consiste nell'invenzione della "Macchina di Turing" (MdT), che sarebbe una macchina ideale (ma che si può anche implementare nella realtà) in grado di fare calcoli potenzialmente all'infinito ed è, quindi, in grado di darci una misura della complessità di un algoritmo. Per essere più precisi, la nozione di algoritmo è riconducibile proprio alle macchine di Turing, in quanto, secondo la tesi di Church-Turing[1], qualsiasi algoritmo è modellabile tramite una macchina di Turing.

[1] È chiamata "tesi" solo perché nella storia è stata ampiamente accettata e considerata come valida, ma in realtà è una congettura.

Una macchina di Turing è costituita da tre componenti principali (Figura 1.1):

1. Un nastro infinito diviso in caselle, ognuna delle quali può contenere un simbolo.
2. Una testina di lettura/scrittura che può leggere e scrivere simboli sul nastro, e può muoversi a sinistra o a destra di una casella alla volta.
3. Un insieme di istruzioni o un programma che controlla il comportamento della macchina in base al simbolo corrente letto dalla testina.

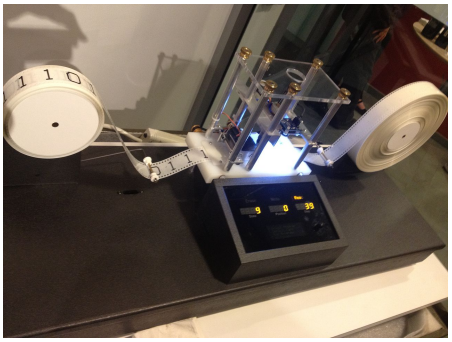

Figura 1.1: Un esempio di Macchina di Turing

Il funzionamento di una macchina di Turing è piuttosto semplice. Inizia con il nastro vuoto o con alcuni simboli su di esso. La macchina legge il simbolo sotto la testina e, in base al suo stato corrente e al simbolo letto, esegue un'azione che può includere la scrittura di un nuovo simbolo, lo spostamento della testina a sinistra o a destra, e la transizione a un nuovo stato.

L'importanza di questo concetto risiede nel fatto che qualsiasi macchina in grado di fare calcoli, indipendentemente dalla sua complessità (e, quindi, qualsiasi computer o calcolatore elettronico in generale), può essere ricondotta al modello di MdT.

Dunque, questo modello risulta essere estremamente versatile, poiché può essere impiegata una MdT per risolvere un qualsiasi problema computazionale o anche per verificare la correttezza di un teorema logico-matematico. Infatti, riprendendo la tesi di Church-Turing, "Se un problema è umanamente calcolabile, allora esisterà una macchina di Turing in grado di risolverlo (cioè di calcolarlo)".

Pertanto ci si è chiesti se è possibile in qualche modo simulare i processi del ragionamento umano tramite delle macchine di Turing. Turing stesso si chiese se una macchina possa in qualche modo "pensare" e propose come criterio di verifica il "Test di Turing", nel quale una macchina può essere considerata "intelligente" se le risposte che fornisce ai quesiti posti da un operatore sono indistinguibili da quelle che fornisce un essere umano.

1.1.2 Nascita effettiva della disciplina

Uno dei primissimi programmi di IA fu il "Logic Theorist" (LT), creato nel 1956 da Allen Newell, Herbert A. Simon, e Cliff Shaw. Il Logic Theorist è un programma di ragionamento automatico, che manipola le espressioni logiche, capace di dimostrare teoremi della logica del primo ordine.

Il suo più importante risultato fu quello di essere riuscito a dimostrare 38 dei 52 teoremi di logica matematica presenti nel Capitolo 2 del Principia Mathematica[2].

[2]Il Principia Mathtematica fu un'opera scritta da Bertrand Russel e Alfred N. Whitehead nel 1910, che si proponeva di definire rigorosomaente le basi della matematica (in particolare l'aritmetica) a partire da alcuni assiomi e regole logiche. Questo tentativo non fu mai riuscito del tutto, perché quegli stessi assiomi portavano ad alcuni paradossi (i celebri "Paradossi di Russel").

Questo risultato diede un po' di entusiasmo alla ricerca, in quanto fu il primo programma in grado di "pensare" in termini non strettamente numerici.

In quello stesso periodo fu organizzato dallo studioso informatico John McCarthy (1927 - 2011) un convegno al quale parteciparono undici studiosi tra scienziati e matematici, per discutere della possibilità di creare una macchina che simuli "ogni aspetto dell'apprendimento o una qualsiasi altra caratteristica dell'intelligenza umana" (Proposta di Dartmouth, p.1), che va sotto il nome di "Conferenza di Dartmouth".

Durante quella conferenza vennero discussi vari temi, tra cui le reti neurali, la teoria della computabilità, la creatività e l'elaborazione del linguaggio naturale. Tale convegno, piuttosto che essere una conferenza ben strutturata, era un dibattito aperto caratterizzato da varie sessioni brainstorming collettivo, da cui iniziò ufficialmente un nuovo filone di studi coniato col termine "Intelligenza Artificiale".

Ispirati dalle idee discusse durante la conferenza di Dartmouth, Newell, Simon e Shaw procedettero con i loro studi e, nel 1957, crearono il "General Problem Solver" (GPS), che, similmente al Logict Theorist, si proponeva di risolvere problemi tramite ragionamento automatico, ma invece di soffermarsi solo ai teoremi di matematici, cercava di occuparsi di qualsiasi problema che un essere umano può porsi, purché sia formalizzabile attraverso la sintassi della logica matematica.

Il GPS è considerato come il primo programma in grado di simulare in maniera soddisfacente il pensiero umano. Ciononostante, il GPS si rivelò subito un approccio estremamente limitato, in quanto non dimostrò una performance buona quando dovette gestire problemi più complessi.

Nel prossimo paragrafo si vedranno meglio i dettagli di questo fallimento e soprattutto quali sono precisamente i limiti di un IA basata unicamente sulla programmazione logica.

1.2 IA basata su programmazione logica (IA simbolica)

L'approccio adottato da Newell, Simon e Shaw diede vita a un intero filone metodologico che, all'interno degli studi sull'IA, fu dominante dalla metà degli anni '50 fino alla fine degli anni '80. Questo approccio si basa sul presupposto che molti aspetti dell'intelligenza possano essere raggiunti mediante la manipolazione dei simboli, da cui, infatti, viene la definizione di "IA simbolica".

1.2.1 Fondamenti teorici

L'IA simbolica necessita,innanzitutto, di un modo di rappresentare le conoscenze di partenza, a partire dalle quali il programma potrà risolvere il problema che è stato posto ad essa e trovare la soluzione migliore possibile.

La rappresentazione della conoscenza in questo tipo di IA si definisce tramite un determinato simbolismo con cui, attraverso regole precise, si possono costruire espressioni complesse.

In particolare, si deve avere una serie di costrutti per definire la sintassi del dominio di interesse (le regole sulle quali costruire delle asserzioni accettabili), una serie di operatori (quantificatori, operatori modali, etc.) che permettano di dare un significato e un valore di verità alle asserzioni rispetto al modello di riferimento.

Attraverso il linguaggio scelto si effettueranno svariate asserzioni sul mondo, che costituiranno una base di conoscenza (KB, Knowledge Base). È, inoltre, importante che il linguaggio scelto per fare le asserzioni sia anche in grado di operare sulla KB per estrarre e aggiungere nuova conoscenza.

Le regole sintattiche con cui costruire tutte le espressioni sono definite tramite una logica formale, in particolare la logica proposizionale e la logica del primo ordine.

Per rappresentare la conoscenza si possono usare anche alberi di decisione. Un albero di decisione consiste in un grafo strutturato gerarchicamente a forma d'albero che permette di analizzare tutte le decisioni che si possono prendere in merito a un problema, con tutte le conseguenze previste (inclusi i relativi costi, risorse e rischi). Un esempio di albero di decisione viene riportato in Figura 1.2.

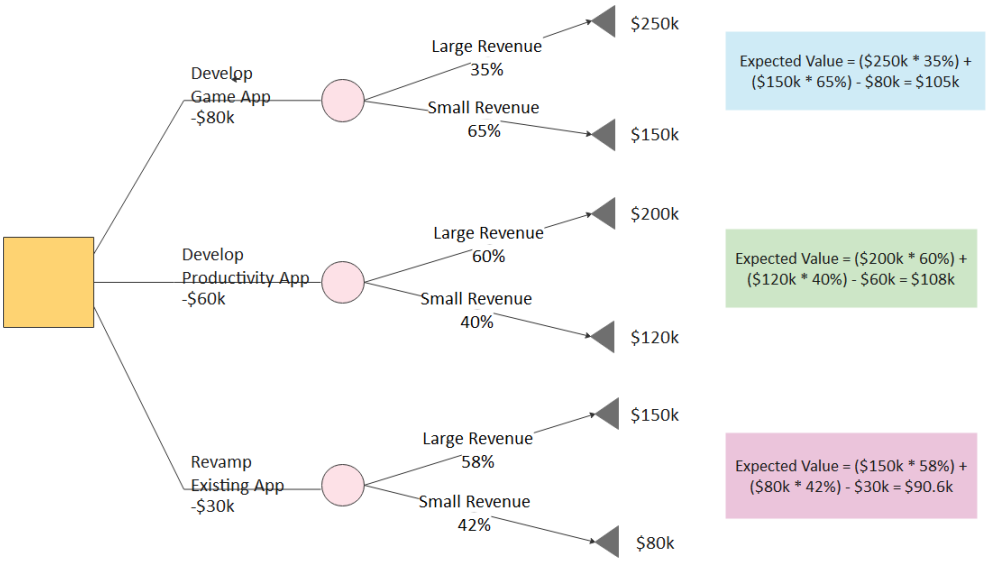

Figura 1.2: Esempio di albero di decisione

Una volta definite sia la conoscenza di base che le regole di sintassi, è necessario stabilire delle regole di inferenza.

Innanzitutto, l'inferenza è un processso della logica attraverso il quale l'IA può ricavare una proposizione a partire da altre che fungono da premesse.

I meccanismi di inferenza sono principalmente tre:

- *induzione*, che consiste nel trarre una conclusione generale a partire dall'esame di molti casi specifici;
- *deduzione*, che è l'inverso dell'induzione e consiste nel trarre una conclusione su un caso specifico a partire dalla conoscenza di una legge generale;
- *abduzione*, che consiste nel trarre una conclusione a partire dalla conoscenza di un insieme di regole e fatti; le ipotesi di partenza possono essere sia fatti specifici che fatti generali e, per questo motivo, è necessario distinguere l'abduzione dall'induzione e dalla deduzione.

Si noti che l'unica inferenza tra le tre appena elencate rigorosamente valida è la deduzione, mentre i risultati prodotti dall'induzione e dall'abudzione non possono essere considerati come validi a priori, perché è comunque necessaria una conferma dalla realtà dei fatti a sostegno.

Questi, così come altri tipi inferenza, devono essere definiti attraverso le regole stabilite dalla logica che adottiamo. Nel caso più comune, la logica adottata è la logica del primo ordine.

1.2.2 Tecnologie ed implementazioni

Per poter mettere in pratica i principi appena descritti sono necessari alcuni strumenti. Uno fondamentale è un linguaggio di programmazione, che permette di implementare la logica su cui l'IA deve basarsi.

A differenza della programmazione tradizionale, sviluppata da linguaggi ad alto livello, la programmazione logica richiede, e allo stesso tempo consente, al programmatore di descrivere la struttura logica del problema piuttosto che il modo di risolverlo. Da un punto di vista concettuale, il programmatore si può così concentrare sugli aspetti logici del problema e sul modo migliore per rappresentarli, senza essere focalizzato sulla necessità di determinare in dettaglio il modo di pervenire ai risultati.

Uno dei linguaggi di programmazione logica più usati è il Prolog (contrazione dal francese PROgrammation LOGique), che è un linguaggio che implementa la logica del primo ordine.

La caratteristica fondamentale di Prolog è che si basa sul calcolo dei predicati, in particolare sulle clausole di Horn. Queste ultime sono delle clausole composte da disgiunzioni di letterali di cui solo uno deve essere positivo e, grazie alle opportune equivalenze[3], è possibile convertirle in una qualsiasi proposizione logica complessa.

Un altro componente fondamentale per una IA simbolica è il motore inferenziale. Quest'ultimo può trarre delle conclusioni di tipo deduttivo oppure di tipo induttivo.

Un motore inferenziale è costituito dai seguenti elementi:

- *Interprete*: decide la regola da applicare.
- *Schedulatore*: decide l'ordine di esecuzione delle regole.
- *Memoria di lavoro*: in essa viene memorizzato un elenco delle operazioni svolte e da svolgere.
- *Rafforzatore di consistenza*: ha il compito di testare la veridicità delle ipotesi fatte.

1.2.3 Insuccessi e limiti

Questo tipo di IA si rivelò nel tempo parecchio limitata, in particolare perché la rappresentazione del mondo in modo simbolico si è rivelata estremamente complessa e risultava estremamente difficile tradurre tutto ciò che l'essere umano sa in una serie di simboli e regole logiche.

Un altro problema particolarmente rilevante fu la difficoltà a gestire l'incertezza; l'IA simbolica era spesso inefficiente nel gestire enormi quantità di dati, in quanto i sistemi basati su simboli richiedevano una grande quantità di risorse computazionali.

Inoltre l'approccio simbolico risultava anche molto rigido e questo comportò grandi difficoltà nell'ambito dell'apprendimento automatico, visto che era molto difficile applicare lo stesso programma di IA a nuovi problemi o a situazioni che non sono state previste in anticipo.

Un celebre esempio di queste difficoltà fu il tentativo negli anni '50-'60 di traduzione automatica di documenti russi in inglese, di particolare interesse del governo americano, dato che era il periodo della Guerra Fredda.

Quello che successe in tal caso era che i ricercatori sottovalutarono l'ambiguità del linguaggio naturale. Per esempio, si era tentato di tradurre una frase che originariamente significava "lo spirito è forte, ma la carne è debole" che, tentando di tradurlo dal russo in inglese, diventava "la vodka è buona, ma la carne è marcia".

[3] In logica proposizionale è possibile convertire una qualsiasi proposizione in una proposizione composta soltanto dagli operatori logici $\neg, \wedge, \vee$. Per esempio la proposizione $x \Rightarrow y$ può essere convertita in $\neg x \vee y$.

Questo episodio, insieme a molti altri simili, portò al taglio dei fondi alle ricerche in Intelligenza Artificiale in quel periodo, evento ricordato tutt'ora come "AI winter".

1.3 IA basata sulle reti neurali

Nella storia è esistito, e tuttora esiste, un approccio alternativo all'IA simbolica, ovvero uno basato sulle reti neurali artificiali.

Questo approccio è fortemente ispirato dal funzionamento del cervello umano; in questo caso, invece di cercare di rappresentare la conoscenza tramite una logica formale, si cerca invece di ricreare gli stessi meccanismi con cui il cervello umano apprende le informazioni e applica le conoscenze acquisite per risolvere i problemi.

Le prime reti neurali furono inventate nel 1943 da Warren Sturgis McCulloch e Walter Pitts.

I due scienziati provarono a modellare un primissimo neurone artificiale schematizzando quello che venne identificato come un "combinatore lineare a soglia", sistema dove lo strato di input prevedeva dati binari multipli in entrata mentre per l'output era previsto un singolo dato binario in uscita. Questo neurone artificiale era in grado di calcolare semplici funzioni booleane.

Tuttavia la prima rete neurale vera e propria fu introdotta nel 1958 da Frank Rosenblatt; essa era chiamata "Perceptron" (in italiano "percettrone"). La novità sta nell'introduzione di due strati, uno di ingresso e uno di uscita, e un algoritmo di apprendimento basato sulla minimizzazione degli errori chiamato "back-propagation".

Le reti neurali non riscontrarono molto successo fino agli anni '80, quando nel 1986 David Rumelhart propose di aggiungere un terzo strato (detto "nascosto") al modello preesistente di percettrone.

1.3.1 Struttura di una rete neurale

Come detto prima, una rete neurale artificiale è una struttura fatta di tanti nodi interconnessi e che si ispira al cervello umano. La rete neurale è un grafo e, in quanto tale, è una struttura composta da archi e nodi. (Figura 1.3)

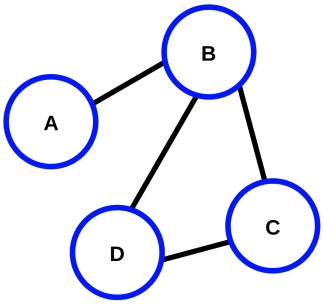

Figura 1.3: Esempio di un grafo

Il nodo è, dunque, l'unità fondamentale della rete neurale ed è, a sua volta, ispirato dal neurone biologico. La funzione principale di un nodo è quella di ricevere input, elaborarli e produrre un output che viene poi inviato agli altri nodi della rete.

Il nodo è in grado di trasformare l'input di partenza e produrre un output applicando una propria funzione caratteristica (detta funzione di attivazione). Nella Figura 1.4 viene mostrato un semplice esempio di output prodotto da un nodo.

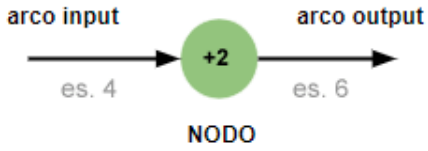

Figura 1.4: Esempio semplice di output prodotto da un nodo

Oltre alla funzione di attivazione, può essere applicata all'input anche un peso, detto "sinaptico", che è caratterizzato dall'arco in cui si trova l'input e che ne modifica il valore prima di essere passato alla funzione. (Figura 1.5)

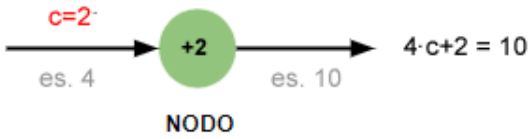

Figura 1.5: Esempio di output prodotto a partire da un input pesato

L'output di un nodo può quindi essere passato ad altri nodi nella rete o può essere l'output finale della rete, a seconda della posizione del nodo rispetto alla rete.

Tutti i nodi di una rete neurale sono organizzati in alcuni strati, che possono essere di tre tipi:

1. *Strato di input*: questo strato riceve i dati in input, che possono rappresentare le caratteristiche di un problema.
2. *Strati nascosti*: questi strati contengono nodi che elaborano i dati in modo da identificare modelli complessi. ogni nodo in uno strato nascosto è collegato a tutti i nodi negli strati di input e di output.
3. *Strato di output*: Questo strato produce l'output della rete sulla base delle elaborazioni svolte negli strati nascosti.

Nella Figura 1.6 viene riportato un esempio semplificato di rete neurale.

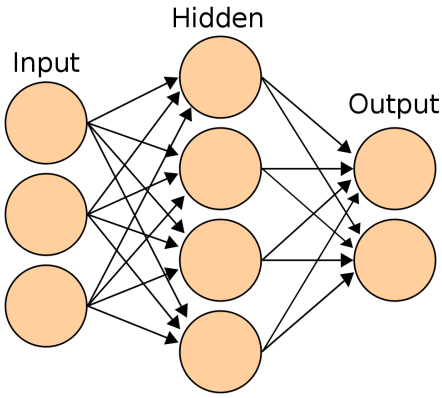

Figura 1.6: Esempio semplificato di rete neurale

A seconda della complessità della rete neurale si possono avere uno o più strati nascosti.

1.3.2 Apprendimento delle reti neurali

Dunque, in una rete neurale, le informazioni viaggiano dal nodo di uno strato a un nodo dello strato successivo. Questo processo avviene in due fasi: la propagazione in avanti (forward propagation) e la propagazione all'indietro (back-propagation).

La propagazione in avanti inizia con gli input della rete. Questi ultimi vengono moltiplicati per i pesi associati ai collegamenti tra i neuroni di ingresso e i neuroni nascosti. I risultati di queste moltiplicazioni vengono, quindi, sommati e sottoposti a una funzione di attivazione. La funzione di attivazione determina l'output del neurone nascosto.

Il processo viene ripetuto per ogni strato nascosto della rete. Alla fine, i segnali raggiungono lo strato di output della rete. I neuroni di output generano una previsione del risultato desiderato.

La propagazione all'indietro è il processo attraverso il quale la rete viene aggiornata per migliorare la sua precisione. Questo processo inizia con il calcolo dell'errore tra la previsione della rete e il risultato desiderato. L'errore viene quindi propagato all'indietro attraverso la rete, a partire dagli strati di output.

A ogni livello, l'errore viene utilizzato per aggiornare i pesi dei collegamenti tra i neuroni. Gli aggiornamenti dei pesi determinano come la rete risponderà agli input futuri.

Il processo di propagazione all'indietro viene ripetuto fino a quando l'errore non è sufficientemente piccolo.

Questa proprietà delle reti neurali permette all'IA di portare a termine in maniera accurata esempi o compiti nuovi, che non ha mai affrontato, dopo aver fatto esperienza su un insieme di dati di apprendimento. Questo processo è noto come "apprendimento automatico" o, in inglese, "Machine Learning[4]" (ML).

Esistono molteplici algoritmi di Machine Learning; questi, generalmente, possono essere classificati in una delle seguenti categorie:

- *Apprendimento supervisionato*, in cui al modello vengono forniti degli esempi nella forma di possibili input e dei rispettivi output desiderati. L'obiettivo è quello di produrre un'ipotesi induttiva, ossia una funzione in grado di "apprendere" dai risultati forniti durante la fase di esempio e in grado di avvicinarsi a dei risultati desiderati per tutti gli esempi non forniti.

- *Apprendimento non supervisionato*, in cui al modello vengono forniti degli input che non vengono etichettati in nessun modo. A differenza dell'apprendimento supervisionato, il modello ha lo scopo di trovare autonomamente le relazioni che possono esistere tra i dati analizzati. Il suo utilizzo è adatto a cercare modelli nascosti nei dati che sfuggono all'osservazione sia perché oscurati da altre informazioni sia perché la quantità di dati è talmente grande da non poter essere osservata facilmente senza un ausilio computazionale.

- *Apprendimento per rinforzo* (chiamato anche *reinforcement learning*), in cui il modello interagisce con un ambiente dinamico nel quale cerca di raggiungere un obiettivo (per esempio guidare un veicolo), avendo un "insegnante" che dice ad esso solo se ha raggiunto l'obiettivo. Questo paradigma si occupa di problemi di decisioni sequenziali, in cui l'azione da compiere dipende dallo stato attuale del sistema e ne determina quello futuro. La qualità di un'azione è data da un valore numerico di "ricompensa", ispirata al concetto di rinforzo, che ha lo scopo di incoraggiare comportamenti corretti

[4]Nel 1983 Tom M. Mitchell scrisse un libro intitolato con questo termine; in questo libro egli scrisse che un programma apprende se c'è un miglioramento delle prestazioni dopo un compito svolto. Questa definizione è emblematica del cambio di approccio rispetto all'IA simbolica, in quanto defnisce l'apprendimento in maniera operativa, piuttosto che in termini cognitivi.

dell'agente. Questo tipo di apprendimento è solitamente modellizzato tramite i processi decisionali di Markov[5].

1.3.3 Modelli di apprendimento avanzati

Esistono modelli di ML un po' più avanzati rispetto a quelli descritti precedentemente. Uno di questi è il Deep Learning (DL), o anche in italiano "apprendimento profondo", che utilizza reti neurali più complesse rispetto a quelle degli algoritmi precedenti.

Ciò che caratterizza le reti neurali utilizzate nel Deep Learning è la presenza di molteplici strati nascosti tra input e output.

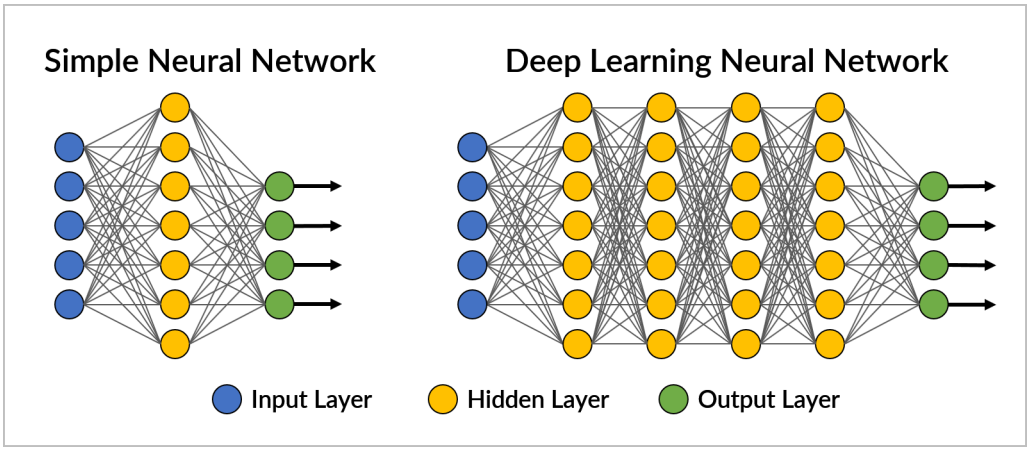

Figura 1.7: Confronto tra reti neurali utilizzate normalmente nel ML e quelle utilizzate nel DL

Gli algoritmi di DL sono in grado di cogliere delle strutture nascoste che ci sono nei dati elaborati e possono essere utilizzati per riconoscere modelli nei grandi volumi di dati, oltre che per implementare previsioni che potrebbero non essere immediatamente evidenti.

Le reti neurali possono essere applicate anche ai processi decisionali in cui i dati sono complessi o in cui la logica non è immediatamente evidente. Il loro uso può aiutare ad analizzare i dati, riconoscere le tendenze e prevedere risultati.

Le reti neurali possono anche essere utilizzate per creare sistemi di controllo che reagiscono automaticamente a determinate situazioni, come la guida autonoma di veicoli o il controllo del traffico aereo.

Anche se la richiesta di immense capacità computazionali rappresenta un ostacolo, la scalabilità del Deep Learning, grazie all'aumento dei dati disponibili e degli algoritmi, è ciò che lo distingue dal Machine Learning.

I sistemi di Deep Learning, infatti, migliorano le proprie prestazioni mano a mano che i dati aumentano, mentre le applicazioni di Machine Learning, noti anche come sistemi di apprendimento superficiale, una volta raggiunto un determinato livello di performance, non sono più scalabili nemmeno aggiungendo esempi e dati di training alla rete neurale.

Tuttavia, va precisato che il calcolo computazionale richiesto per il funzionamento dei sistemi di Deep Learning è molto più impattante rispetto ai modelli di ML di base, sia dal punto di vista computazionale che dal punto di vista economico.

[5]Un processo decisionale di Markov è un modello matematico che permette a un agente intelligente di prendere decisioni in un ambiente dinamico. Esso definisce le decisioni che l'agente può intraprendere e tutti gli stati possibili in cui si può trovare. In questo contesto, l'obiettivo dell'agente è di prendere le decisioni che permettano ad esso di massimizzare la ricompensa attesa.

Esistono svariati modelli di Deep Learning; di seguito ne verranno elencati alcuni:

- *Reti Neurali Convoluzionali (CNN)*: utilizzate principalmente per l'elaborazione di immagini e riconoscimento di pattern in dati bidimensionali.
- *Reti Neurali Ricorrenti (RNN)*: adatte per dati sequenziali, come il linguaggio naturale, e in grado di gestire informazioni temporali.
- *Autoencoder*: utilizzato per apprendere rappresentazioni efficienti e comprimere dati, spesso usato per la riduzione della dimensionalità.
- *Transformer*: introdotto originariamente per il trattamento del linguaggio naturale, ha dimostrato di essere altamente efficace anche in altri contesti grazie alla sua struttura attenzionale. Questo modello, in particolare, è quello su cui si basa GPT.

1.3.4 Limiti delle reti neurali

Per quanto l'apprendimento automatico sia stato molto impattante in ambito IA, i programmi di ML non sempre sono in grado di fornire i risultati attesi.

Ci possono essere molteplici ragioni per tale incapacità, tra cui:

- *Bias*: similmente a quello che può succedere a un essere umano, la rete neurale potrebbe giungere a una conclusione errata perché ha trovato una correlazione tra alcuni dati che, in realtà, non sussiste.
- *Scatola nera*: la rete neurale spesso si comporta come una "scatola nera", in quanto non sempre è possibile fornire una spiegazione chiara dietro alle decisioni che essa prende.
- *Overfitting*: le reti neurali possono essere suscettibili di overfitting, il che significa che possono adattarsi troppo ai dati di addestramento e avere difficoltà a generalizzare su nuovi dati. Diverse tecniche, come la regolarizzazione e l'aumento dei dati, sono utilizzate per mitigare questo problema.

1.4 Impatto dell'IA sull'informatica

1.4.1 Evoluzione di alcuni campi dell'informatica già esistenti

L'introduzione dell'IA nell'informatica ha fatto in modo che alcuni ambiti si evolvessero o, perlomeno, integrassero l'IA utilizzandola come supporto.

Uno di questi è la programmazione e, in generale, lo sviluppo di qualsiasi software. La fase di programmazione può essere resa più efficiente grazie alla possibilità di generare automaticamente il codice di base, di creare le viste e/o di sviluppare alcuni algoritmi specifici. Similmente è possibile automatizzare parzialmente lo sviluppo di test. Inoltre, l'IA può essere utilizzata per aiutare a risolvere i problemi nel software e quindi ridurre il tempo necessario per identificare e risolvere gli errori. Pertanto, automatizzando alcune attività che normalmente richiederebbero l'intervento umano, l'IA è in grado di ridurre i costi di sviluppo del software e di portare ad un aumento della velocità di lancio di nuovi prodotti.

L'IA offre alcune possibilità anche in ambito di cybersecurity, in quanto può essere utilizzata per analizzare grandi quantità di dati in tempo reale, alla ricerca di anomalie o comportamenti sospetti che potrebbero indicare un attacco informatico. Questo può aiutare a rilevare e rispondere agli attacchi più rapidamente e in modo più efficace. Allo stesso modo, l'IA può essere utilizzata per identificare e mitigare le vulnerabilità che potrebbero essere sfruttate dagli aggressori.

L'IA può essere utilizzata anche nell'ambito della data science. Grazie alla sua capacità di gestire grandi quantità di dati, essa può essere utilizzata per automatizzare i processi di raccolta dei dati, come la scansione di documenti, l'analisi di immagini o l'analisi dei dati in tempo reale. Grazie ai meccanismi di apprendimento automatico, l'IA è capace di trovare delle correlazioni tra i dati o dei modi di classificarli che potrebbero facilmente sfuggire a un osservatore umano.

1.4.2 Nascita di alcuni nuovi campi dell'informatica

Oltre a permettere il miglioramento di alcuni campi già esistenti, l'IA è stata in grado di dar vita a nuovi rami di applicazione dell'informatica, permettendo di adottare degli approcci del tutto nuovi ad alcuni problemi che sono stati posti nel corso del tempo.

Uno di questi è la visione artificiale (in inglese, detta, anche, "computer vision"), che è una branca che si occupa di riprodurre artificialmente la visione umana. Per la precisione, si occupa di creare un modello approssimato del mondo reale (3D) a partire da immagini bidimensionali (2D).

Tuttavia, l'obiettivo di questi studi non è soltanto di fare in modo che la macchina possa "vedere", ma è soprattuto quello di permettere alla macchina di ricavare tutte le informazioni che ha raccolto tramite la "vista", affinché possa prendere delle decisioni di conseguenza.

Un sistema di visione artificiale è costituito dall'integrazione di componenti ottiche, elettroniche e meccaniche che permettono di acquisire, registrare ed elaborare immagini sia nello spettro della luce visibile che al di fuori di essa (infrarosso, ultravioletto, raggi X, etc.). Il risultato dell'elaborazione è il riconoscimento di determinate caratteristiche dell'immagine per varie finalità di controllo, classificazione, selezione, etc.

Un problema classico nella visione artificiale è quello di determinare se l'immagine contiene o meno determinati oggetti (object recognition) o attività. Il problema può essere risolto efficacemente e senza difficoltà per oggetti specifici in situazioni specifiche, per esempio il riconoscimento di specifici oggetti geometrici come poliedri, il riconoscimento di volti o di caratteri scritti a mano. Le cose si complicano nel caso di oggetti arbitrari in situazioni arbitrarie.

Un altro contesto in cui l'IA trova un'applicazione innovativa è l'elaborazione del linguaggio naturale (in inglese "Natural Language Processing", o NLP). Si tratta di una branca di studi che mette insieme la linguistica e l'informatica e che si occupa di creare macchine che siano in grado di comprendere e generare messaggi in linguaggio naturale. In particolare, lo scopo è rendere la tecnologia in grado di "comprendere" il contenuto dei documenti e le loro sfumature contestuali, in modo tale che possa estrarre con precisione informazioni e idee contenute nei documenti, nonché classificare e categorizzare i documenti stessi.

Questo processo è reso particolarmente difficile e complesso a causa delle caratteristiche intrinseche di ambiguità del linguaggio umano. Per questo motivo il processo di elaborazione viene suddiviso in fasi diverse e, tuttavia, simili a quelle che si possono incontrare nel processo di elaborazione di un linguaggio di programmazione. Tali fasi sono:

- *analisi lessicale*: scomposizione di un'espressione linguistica in token (in questo caso, le parole);
- *analisi grammaticale*: associazione delle parti del discorso a ciascuna parola nel testo;
- *analisi sintattica*: arrangiamento dei token in una struttura sintattica (ad albero: parse tree);
- *analisi semantica*: assegnazione di un significato alla struttura sintattica e, di conseguenza, all'espressione linguistica, disambiguando dove è necessario.

Infine, esiste anche l'IA generativa, che è un tipo di IA particolare in grado di generare automaticamente del testo o vari tipi di media a seconda della richiesta che viene posta ad essa. Tra i più noti attualmente vi sono ChatGPT, un chatbot creato da OpenAI utilizzando i modelli linguistici GPT-3 e GPT-4, Bard, sviluppato da Google, o Bedrock, sviluppato da Amazon. L'Intelligenza Artificiale generativa ha potenziali applicazioni in una vasta gamma di settori, tra cui lo sviluppo software, il marketing e la moda, l'editoria, la predizione della struttura proteica e la scoperta di farmaci (a partire da catene di aminoacidi o rappresentazioni di molecole, come la codifica SMILES, che rappresenta DNA o proteine).

CAPITOLO 2

Amazon Web Services

In questo capitolo verrà introdotto Amazon Web Services (AWS), che consiste un insieme di servizi di cloud computing, tra cui quelli che hanno permesso la realizzazione dei case study di questo documento. In particolare, si parlerà di com'è strutturato AWS e delle sue potenzialità a livello globale; successivamente verranno trattati la struttura del cloud a sostegno del servizio e i modelli utilizzati, per poi, infine, analizzare gli strumenti di machine learning che hanno permesso la realizzazione dei case study, che verranno trattati nei capitoli successivi.

2.1 Caratteristiche di AWS

AWS è una piattaforma di cloud computing che, attualmente, offre oltre 200 servizi, tra cui calcolo, archiviazione, networking, deployment, system management, servizi applicativi, machine learning, tool per sviluppo software e per l'IoT, etc.

Le origini di AWS risalgono al 2002, quando Amazon iniziò a sviluppare la propria infrastruttura di cloud computing per supportare il proprio sito web in rapida crescita. Nel 2006, Amazon ha deciso di rendere disponibile questa infrastruttura a terzi. Furono, dunque, fondati Amazon S3, come servizio di archiviazione su cloud, e Amazon EC2, come servizio di cloud computing che permetteva di affitare macchine virtuali su cui eseguire le applicazioni.

L'intento era di agevolare gli sviluppatori nello sviluppo dei propri software, permettendo loro di non preoccuparsi di dove vengano memorizzati i dati, assicurando loro che verranno salvati, oltre che mantenuti protetti e disponibili per ogni volta che ne avranno bisogno; tutto questo assicurando, comunque, la disponibilità del server in maniera continua.

AWS ha avuto un successo immediato e, nel giro di pochi anni, è diventata la piattaforma di cloud computing più utilizzata al mondo. Questo successo è dovuto a una serie di fattori, tra cui:

- *la scalabilità e la disponibilità della piattaforma*, che consentono ai clienti di scalare le proprie risorse in base alle proprie esigenze;
- *la sicurezza e la conformità della piattaforma*, che soddisfano i requisiti dei clienti più esigenti;
- *il prezzo competitivo della piattaforma*, che la rende accessibile a un'ampia gamma di clienti.

2.1.1 Gamma di servizi offerti da AWS

Tra i vari servizi offerti da AWS vi sono servizi di calcolo, fatti per soddisfare le esigenze di applicazioni e carichi di lavoro diversi. Uno dei servizi più popolari è Elastic Compute Cloud (EC2), che ha una capacità di calcolo scalabile, che consente di lanciare rapidamente macchine virtuali con vari sistemi operativi. Esso, inoltre, permette di configurare delle macchine virtuali e di scalare facilmente verso l'alto o verso il basso, in base alle proprie esigenze.

Un altro servizio particolarmente noto è Amazon Simple Storage Service (S3), che consiste in un servizio di archiviazione a oggetti che consente di archiviare, recuperare e gestire grandi quantità di dati, come documenti, immagini e video. S3 è fortemente scalabile e fornisce una soluzione altamente durevole e disponibile per l'archiviazione dei dati.

Inoltre, c'è anche AWS Lambda, un servizio di serverless computing, che consente di eseguire codice senza necessità di fornire o gestire server. Con Lambda è possibile eseguire il codice in risposta a eventi, come le modifiche ai dati in un bucket S3 o una nuova richiesta API. Questo servizio è ideale per la creazione di microservizi e applicazioni basate su eventi.

AWS fornisce, anche, svariati servizi di database, tra cui DynamoDB, che è un servizio di database NoSQL. Esso supporta modelli di dati sia documentali che di tipo valore-chiave e offre una scalabilità illimitata, ideale per applicazioni cloud-native.

Un altro servizio di database è Amazon Relational Database Service (RDS), che consiste in un servizio di database relazionale gestito, che semplifica l'impostazione, la gestione e la scalabilità di un database relazionale nel cloud. RDS supporta diversi motori di database, tra cui Amazon Aurora, Microsoft SQL Server, Oracle, PostgreSQL e MySQL. Esso fornisce backup automatici, patch del software e rilevamento e ripristino automatico dei guasti.

AWS offre, inoltre, diversi servizi di rete e di distribuzione dei contenuti, tra cui, in particolare, Virtual Private Cloud (VPC) e Route 53. VPC consente di lanciare le risorse AWS in una sezione logicamente isolata del cloud AWS, dove è possibile archiviare i dati in modo sicuro e accedervi attraverso la propria rete. VPC controlla l'ambiente di rete virtuale, compresi l'intervallo di indirizzi IP, le sottoreti, le tabelle di routing e i gateway di rete. Invece, Amazon Route 53 è un servizio DNS (Domain Name System) che permette di instradare il traffico verso le applicazioni o i siti web in base alla latenza o alla geolocalizzazione. Route 53 fornisce anche servizi di registrazione dei domini, consentendo di gestire facilmente i nomi di dominio e i record DNS in un unico posto.

AWS offre, anche, vari strumenti di sviluppo e distribuzione del software, come CodeStar e CodeBuild. AWS CodeStar è un servizio completamente gestito che semplifica lo sviluppo, la creazione e la distribuzione di applicazioni su AWS. Esso fornisce ambienti di sviluppo preconfigurati per vari linguaggi di programmazione e framework, tra cui Java, .NET, Node.js, Python e Ruby. AWS CodeBuild, invece, consente l'integrazione di codice sorgente con repository esterni (come GitHub) e fornisce ambienti di compilazione preconfigurati per vari linguaggi di programmazione e framework. CodeBuild fornisce, inoltre, un'infrastruttura di compilazione scalabile e altamente disponibile, ideale per carichi di lavoro di compilazione e test su larga scala.

2.1.2 Dimensione e portata globale di AWS

AWS offre una copertura globale attraverso la sua infrastruttura, che comprende diverse regioni e zone di disponibilità. Ogni regione AWS è costituita da un minimo di tre zone di disponibilità isolate e fisicamente separate all'interno di un'area geografica. Attualmente vi sono 31 regioni, come si può osservare nella Figura 2.1.

Ogni regione è composta da diverse zone di disponibilità, ciascuna delle quali ha la propria capacità di alimentazione, raffreddamento e sicurezza fisica.

Figura 2.1: Zone di disponibilità di Amazon; le zone indicate in verde sono le regioni attualmente esistenti, quelle indicate in rosso sono le regioni che Amazon sta progettando e realizzerà in futuro.

Alcune delle zone di disponibilità di AWS hanno la copertura AWS Wavelength. Queste ultime sono distribuzioni di infrastrutture AWS che integrano i servizi di calcolo e storage di AWS nelle reti 5G dei fornitori di servizi di telecomunicazione, permettendo al traffico delle applicazioni da dispositivi 5G di raggiungere i server delle applicazioni in esecuzione nelle zone Wavelength senza uscire dalla rete delle telecomunicazioni.

Ciò permette di evitare che la latenza che risulterebbe dal traffico delle applicazioni attraversi più hop in Internet per raggiungere la destinazione, consentendo ai clienti di sfruttare totalmente i vantaggi della latenza e della larghezza di banda offerti dalle reti 5G moderne.

2.1.3 Strumenti e servizi di sicurezza forniti da AWS

AWS offre una vasta gamma di strumenti e servizi di sicurezza per aiutare gli sviluppatori a costruire, eseguire e scalare le loro applicazioni su un'infrastruttura cloud sicura.

Di seguito verranno elencati i principali strumenti e servizi di sicurezza offerti da AWS:

- *Gestione dell'identità e dell'accesso*: i servizi di identità AWS aiutano a gestire in modo sicuro le proprie identità, le proprie risorse e i propri permessi.
- *Rilevamento e risposta*: i servizi di rilevamento e risposta di AWS aiutano a migliorare il livelllo di sicurezza e a semplificare le operazioni di sicurezza in tutto il proprio ambiente AWS.
- *Protezione della rete e delle applicazioni*: i servizi di protezione della rete e delle applicazioni aiutano a far rispettare una politica di sicurezza dettagliata nei punti di controllo della rete in tutta la propria organizzazione, permettendo una scalabilità sicura.
- *Protezione dei dati*: AWS permette di controllare i propri dati utilizzando potenti servizi e strumenti che consentono di scegliere il luogo in cui sono archiviati i propri dati, come sono protetti e chi può accedervi; servizi quali AWS Identity and Access Management

(IAM) consentono di gestire in sicurezza l'accesso ai servizi e alle risorse AWS. AWS CloudTrail e Amazon Macie consentono conformità, rilevamento e verifica, mentre altri servizi come AWS CloudHSM e AWS Key Management Service (KMS) permettono di generare e gestire in modo sicuro le chiavi di crittografia.

- *Conformità*: AWS offre una visione completa dello stato di conformità e monitora continuamente l'ambiente dei clienti, utilizzando controlli di conformità automatizzati basati sulle migliori pratiche di AWS e sugli standard di settore che le proprie organizzazioni seguono.

2.2 Struttura e gestione del cloud a disposizione

2.2.1 Concetti base del cloud computing

Il cloud computing nacque[1] dall'esigenza di hardware sempre più potente per poter lavorare in modo agevole; per questo motivo, si cominciarono a erogare servizi agli utenti tramite la rete Internet (servizi quali l'archiviazione e l'elaborazione dei dati), a partire da un insieme di risorse preesistenti, configurabili e disponibili in remoto sotto forma di architettura distribuita.

Tipicamente, le risorse non vengono pienamente configurate e messe in opera dal fornitore appositamente per l'utente, ma gli sono assegnate convenientemente grazie a procedure automatizzate, a partire da un insieme di risorse condivise con altri utenti, lasciando all'utente parte dell'onere della configurazione.

Utilizzare un cloud può presentare diversi vantaggi, come la riduzione dei costi, eliminando la necessità di investire in hardware, la scalabilità rapida delle risorse in base alle esigenze, l'accessibilità globale e la possibilità di backup automatici. Inoltre, offre agilità operativa, con aggiornamenti software e gestione dell'infrastruttura forniti dal fornitore.

Tuttavia, ci sono anche svantaggi, inclusi problemi di sicurezza e sulla privacy dei dati e dipendenza da una connessione Internet stabile. Per questo motivo, i servizi di cloud adottano diverse strategie per ovviare a questi problemi. Una di queste è la ridondanza sulla rete, così da permetterne il funzionamento anche in caso di guasto di uno o più nodi al suo interno; un altro consiste nell'adottare misure di sicurezza a livello di rete, utilizzando firewall e crittografando i dati in transito, in modo da garntirne l'integrità; inoltre, offrono backup automatici e soluzioni di ripristino dati per garantire la continuità operativa. Tutto questo viene fatto con un monitoraggio continuo del sistema, in modo da poter agire tempestivamente in caso di qualche malfunzionamento, prima che questi ultimi possano causare l'interruzione del servizio.

[1] Gli studi sul cloud computing iniziarono durante gli anni '60, ma la pratica cominciò a consolidarsi a partire dagli anni '90-2000.

2.2.2 Principali modelli di cloud offerti da AWS

I principali modelli di servizi cloud offerti da AWS sono:

- *Infrastructure as a Service (IaaS)*: vengono offerte risorse hardware, quali server, capacità di rete, sistemi di memoria e archivio; inoltre, il cliente può scegliere cosa eseguire nelle macchine (sistemi operativi compresi).
- *Platform as a Service (PaaS)*: viene eseguita in remoto una piattaforma software che può essere costituita da diversi servizi, programmi, librerie, etc.; tutto questo consentendo al cliente di sviluppare le applicazioni aziendali senza i costi e la complessità associati alla gestione dell'hardware.
- *Software as a Service (SaaS)*: viene messo a disposizione del cliente un software applicativo che può essere eseguito da remoto su una macchina cloud, a cui il cliente può accedervi facilmente tramite browser; questo piano è utile, in quanto vengono garantite continuamente la sincronizzazione e un'elevata versatilità, senza che il cliente debba configurarsi da solo un sistema di backup.

Inoltre, AWS supporta diversi modelli di distribuzione del cloud, ovvero:

- *Completamente su cloud*: in questo modello tutta l'implementazione, esecuzione compresa, viene fatta direttamente sul cloud; questo modello permette di sfruttare al massimo tutti i vantaggi offerti dal cloud discussi finora. Le applicazioni possono, inoltre, essere costruite su componenti di infrastruttura di basso livello o possono utilizzare servizi di livello superiore che forniscono astrazioni dai requisiti di gestione, architettura e scalabilità dell'infrastruttura principale.
- *Ibrido*: questo modello permette di connettere le risorse del cloud con quelle non presenti in esso; per esempio, questa tecnica può essere utilizzata per sviluppare le applicazioni utilizzando le risorse locali e, all'occorrenza, estenderle utilizzando le risorse offerte dal cloud.
- *Locale*: detto anche "cloud privato", permette la distribuzione di risorse in locale, utilizzando strumenti di virtualizzazione e gestione delle risorse. L'implementazione locale non offre molti dei vantaggi del cloud computing, ma, a volte, viene richiesta per la sua capacità di fornire risorse dedicate.

2.3 Servizi di Machine Learning

2.3.1 Reti neurali utilizzate da AWS

AWS fa un ampio uso di reti neurali ricorrenti (in inglese, dette "Recurrent Neural Networks", RNN). Esse sono addestrate per elaborare e convertire un input di dati sequenziale in un output di dati sequenziale specifico. I dati sequenziali sono dati, come parole, frasi o dati di serie temporali, in cui i componenti sequenziali sono correlati in base a regole semantiche e sintattiche complesse.

Nella Figura 2.2 viene mostrata la struttura di una RNN.

Le RNN funzionano trasmettendo i dati sequenziali che ricevono ai livelli nascosti un passaggio alla volta. Tuttavia, hanno anche un flusso di lavoro ricorrente o a ciclo automatico: il livello nascosto può ricordare e utilizzare gli input precedenti per le previsioni future in una componente di memoria a breve termine. Grazie a questo meccanismo, queste reti neurali utilizzano l'input corrente e la memoria archiviata per prevedere la sequenza successiva.

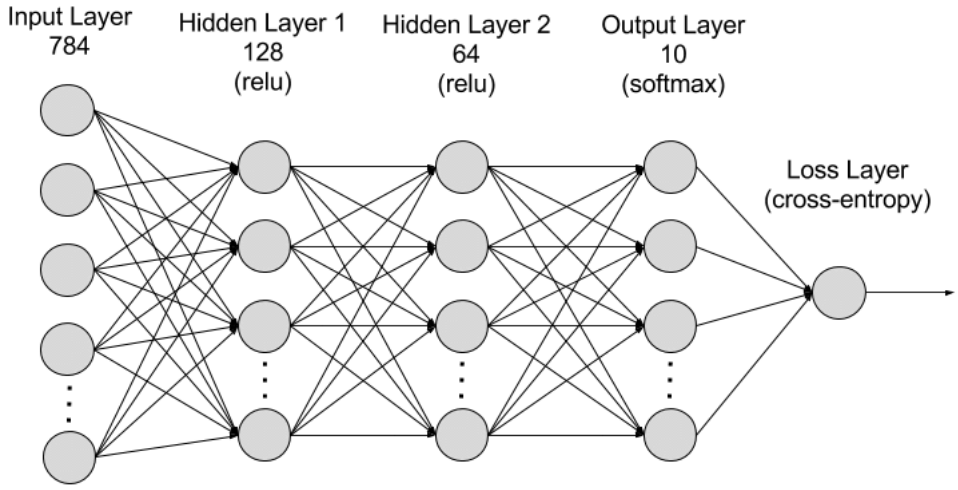

Figura 2.2: Struttura di una RNN

Le RNN sono spesso caratterizzate da architetture di tipo "uno a uno", dove a un input è associato un output. Tuttavia, sono possibili altri modelli di RNN:

- *Uno a molti*: a partire da un singolo input vengono prodotti molteplici output; questo modello è tipico delle applicazioni che richiedono di generare un'intera frase a partire da una singola parola chiave.
- *Molti a molti*: si utilizzano più input per prevedere più output; questo modello è tipico dei software di traduzione, che richiedono un'intera frase come input per poi restituire, come output, una frase in un'altra lingua.
- *Molti a uno*: diversi input sono mappati su un output; ciò è utile in applicazioni come l'analisi del sentiment, in cui il modello prevede i sentiment dei clienti come positivi, negativi e neutri in base alle testimonianze in entrata.

Le RNN presentano anche alcuni limiti importanti:

- *Gradiente esplosivo*: una RNN può prevedere erroneamente l'output dell'addestramento iniziale, per cui è necessario effettuare diverse iterazioni per regolare i parametri del modello e ridurre il tasso di errore; la sensibilità del tasso di errore può essere vista come un gradiente, pertanto averne uno molto alto significa avere una rete neurale molto veloce ad apprendere; tuttavia, se questo gradiente è troppo alto, la rete neurale diventa instabile e si verificano problemi di prestazioni, come l'overfitting (fenomeno in cui la rete neurale è in grado di fare previsioni accurate con i dati di addestramento, ma non lo è con i dati del mondo reale).
- *Scomparsa del gradiente*: è il problema inverso rispetto a quello del gradiente esplosivo e si verifica quando il gradiente è così basso da diventare quasi nullo; quando il gradiente scompare in questo modo, la rete neurale non riesce più ad apprendere dai dati di addestramento e, pertanto, non è in grado di effettuare alcuna previsione (questo fenomeno è noto come underfitting).
- *Tempo di addestramento lento*: una RNN elabora i dati in modo sequenziale, il che limita la sua capacità di elaborare un gran numero di testi in modo efficiente. Ad esempio, un modello RNN può analizzare il sentiment di un acquirente a partire da un paio di

frasi. Tuttavia, richiede un'enorme potenza di calcolo, spazio di memoria e tempo per riassumere una pagina di un saggio.

Alcuni di questi limiti possono essere superati utilizzando dei trasformatori, che utilizzano delle unità di auto-attenzione. Esse permettono di elaborare i dati in parallelo, senza che sia necessario utilizzare degli strati nascosti, il che consente di addestrare ed elaborare sequenze più lunghe in minor tempo rispetto a una RNN. Inoltre, siccome i trasformatori sono capaci di elaborare i dati in parallelo, essi non sono soggetti a restrizioni di retropropagazione, perché i gradienti possono fluire liberamente verso tutti i pesi. Questo meccanismo di parallelismo permette ai trasformatori di lavorare con modelli più grandi rispetto a una RNN, il che può essere utile, ad esempio, per gestire attività complesse di NLP.

2.3.2 Principali algoritmi di apprendimento

AWS dispone di molti algoritmi per il Machine Learning. Uno di questi si chiama proprio Amazon Machine Learning, che utilizza la discesa stocastica del gradiente (Stochastic Gradient Descent - SGD) come tecnica di ottimizzazione. Questo algoritmo effettua passate sequenziali sui dati di addestramento e, durante ogni passata, aggiorna i pesi delle feature, un campione alla volta, con l'obiettivo di raggiungere i pesi ottimali, in grado di ridurre al minimo la perdita. Amazon ML usa i seguenti algoritmi di apprendimento:

- Per la classificazione binaria, impiega la regressione logistica (funzione di perdita logistica + SGD).
- Per la classificazione multiclasse, impiega la regressione logistica multinomiale (perdita logistica multinomiale + SGD).

AWS utilizza, inoltre, molti algoritmi e framework per il deep learning, tra cui TensorFlow, che consiste in un framework open-source che si può utilizzare per applicazioni di ML e DL, oppure Apache MXNet, che è un framework di DL scalabile supportato da Apache, oppure SageMaker, che offre direttamente servizi di creazione, addestramento e distribuzione dei modelli di apprendimento, etc.

2.3.3 Servizi scelti per l'analisi dei case study

Per l'analisi dei case study che verranno discussi nei prossimi capitoli, sono stati scelti i seguenti servizi:

- *Amazon Rekognition*: è un servizio che effettua l'analisi di immagini e video, capace di riconoscere volti, oggetti, scene o attività.
- *Amazon Translate*: è un servizio che è in grado di effettuare la traduzione, sia in tempo reale, sia di interi documenti.
- *Amazon Textract*: è un servizio che è in grado di estrarre automaticamente testi, tabelle e dati strutturati da vari documenti, come fatture, moduli, questionari, etc.

CAPITOLO 3

Riconoscimento di immagini

In questo capitolo verranno analizzati gli strumenti forniti da Amazon Rekognition per il riconoscimento di immagini e video. In particolare, ci si concentrerà nel riconoscimento dei volti celebri, nel riconoscimento dei volti tramite confonto facciale e nel riconoscimento dei contesti o eventi raffigurati nei video. Inoltre, si vedrà come l'IA svolge queste attività con alcuni esempi e, in seguito, si osserveranno i limiti di questa tecnologia.

3.1 Introduzione al riconoscimento di immagini

Prima di prendere in esame i case study in particolare, si vedranno come funziona il riconoscimento di immagini e le potenzialità offerte da questo strumento; in seguito, si vedrà come viene utilizzato il Machine Learning per questo scopo.

3.1.1 Proprietà del riconoscimento di immagini

Il riconoscimento di immagini è un campo dell'informatica che si occupa di sviluppare algoritmi e modelli per consentire ai computer di identificare e comprendere il contenuto di un'immagine. Questo processo è una componente essenziale dell'Intelligenza Artificiale e della visione artificiale. In generale, il riconoscimento di immagini si basa su tecniche di Machine Learning e Deep Learning per insegnare ai computer a riconoscere pattern e caratteristiche visive nelle immagini.

Il riconoscimento di immagini, in generale, viene svolto in alcune fasi:

- *Acquisizione dell'immagine*: l'immagine viene acquisita da una fotocamera o da un altro dispositivo di imaging.
- *Preprocessing*: l'immagine viene preprocessata per migliorare la qualità e ridurre il rumore; ciò può includere operazioni come il bilanciamento del colore, la normalizzazione e la riduzione del rumore.
- *Addestramento del modello*: utilizzando un set di dati annotato, il modello di rete neurale viene addestrato per riconoscere pattern specifici oppure oggetti nelle immagini; durante l'addestramento, il modello impara a regolare i suoi parametri in modo che le sue previsioni si avvicinino sempre di più alle etichette di training.
- *Estrazione delle caratteristiche*: vengono estratte le caratteristiche più significative dell'immagine; questo, in genere, lo si fa tramite delle reti neurali convoluzionali (CNN)[1], che

apprendono automaticamente quali caratteristiche devono estrarre durante la fase di addestramento.

- *Testing e valutazione*: il modello viene testato su nuovi dati per valutare le sue prestazioni; questa fase serve ad assicurarsi che esso sia in grado di generalizzare e, quindi, riconoscere immagini che non ha mai visto.

- *Utilizzo in tempo reale*: una volta addestrato e valutato, il modello può essere utilizzato per riconoscere automaticamente oggetti o pattern nelle nuove immagini in tempo reale.

In generale, il riconoscimento di immagini può trovare varie applicazioni in molteplici ambiti diversi. Per esempio, può essere utilizzato per riconoscere difetti di produzione in ambito industriale, oppure può essere impiegato per la guida autonoma, permettendo di rilevare automaticamente pedoni, cartelli stradali, altri veicoli e ostacoli di varia natura, oppure, ancora, può essere applicato alla diagnostica medica, in modo tale da identificare immediatamente possibili malattie e patologie a partire dalle immagini fornite dalle microscopie, dai raggi X, dalle risonanze magnetice, etc.

In particolare, in questa in tesi, si vedrà il riconoscimento di immagini applicato nell'ambito del riconoscimento facciale, con particolare focus sull'identificazione delle persone, e in quello della classificazione di oggetti, eventi e contesti raffigurati nei video, attraverso l'assegnazione di etichette.

Un altro aspetto da considerare sono i benefici che si possono trarre da questa tecnologia; tra questi citiamo:

- *Costanza, affidabilità, oggettività dei controlli*: al contrario degli operatori umani, un sistema di visione è in grado di garantire un controllo secondo criteri di valutazione sempre omogenei e di operare senza cali significativi di prestazione, anche per lunghi periodi di tempo.

- *Operabilità in ambienti ostili*: in condizioni ambientali limite, come ambienti molto rumorosi, esposti ad agenti chimici, temperature molto elevate o molto fredde, un sistema di visione può operare in tranquillità senza mettere in pericolo i lavoratori.

- *Piccole dimensioni degli oggetti da controllare*: la dimensione degli oggetti può costituire un limite per la verifica e il controllo umano; i sistemi di visione consentono di analizzare particolari non visibili, o difficilmente identificabili dall'uomo, grazie a ottiche e software specifici.

- *Elevata precisione del controllo*: in genere i sistemi di visione consentono di raggiungere una precisione ed un'accuratezza del controllo di gran lunga superiore a quella umana.

- *Elevata velocità di controllo*: oltre ad essere più precisi, in genere i sistemi di visione sono anche molto più veloci ad effettuare il controllo rispetto agli operatori umani.

[1] Le reti neurali convoluzionali sono delle reti neurali di tipo "feed-forward", ovvero caratterizzati dall'assenza di strutture cicliche tra i nodi, che sono progettate con dati strutturati in griglia; per questo motivo, si prestano molto bene nel riconoscimento di immagini e video.

3.1.2 Amazon Rekognition nel riconoscimento di immagini

I servizi offerti da Amazon Rekognition si basano su due API, Rekognition Image e Rekognition Video.

Rekognition Image è un servizio di riconoscimento delle immagini che rileva oggetti, scene, attività, punti di riferimento, volti, colori dominanti e qualità dell'immagine. Rekognition Image estrae anche il testo, riconosce le celebrità e identifica i contenuti inappropriati nelle immagini. Esso consente anche di ricercare e confrontare i volti.

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.

Per quanto riguarda le immagini, Amazon Rekognition supporta i formait .jpg e .png; in alternativa, le immagini possono essere caricate come oggetti di S3; la dimensione massima che l'immagine può avere è di 15 MB, se caricata come oggetto di S3, mentre è di 5 MB negli altri casi; inoltre, per ottenere dei risultati sufficentemente affidabili, è consigliato utilizzare immagini che abbiano una risoluzione di 640×480, o superiore.

Per quanto rigarda i video invece, Amazon Rekognition Video supporta file di tipo .mp4, .avi e .mov, la cui dimensione massima deve essere di 10 GB; se, invece, il video viene caricato come oggetto di S3, esso può essere lungo fino a 6 ore.

3.2 Analisi dei case study

Come già anticipato precedentemente, i case study presi in considerazione in questo capitolo rientreranno nell'ambito dell'identificazione tramite riconoscimento facciale e dell'identificazione di oggetti, contesti ed eventi all'interno dei video.

I servizi di Amazon Rekognition scelti per questo studio sono i seguenti:

- riconoscimento di volti celebri;
- confronto facciale;
- analisi dei video archiviati.

Ciascuno di questi servizi riceverà come input un'immagine o un video e produrrà come output un file JSON che contiene una serie di parametri con i loro rispettivi valori, da cui, in seguito, ricaverà il risultato finale dell'analisi compiuta.

3.2.1 Riconoscimento di volti celebri

Questo servizio di Amazon analizza un'immagine caricata in input e, come risultato, produce una risposta in cui specifica quale personaggio famoso è riuscito ad identificare e con quanta confidenza (espressa in percentuale) fornisce questa risposta.

Per esempio, come mostrato nella Figura 3.1, viene fornita un'immagine di Ken Thompson (uno dei principali inventori del sistema operativo UNIX) e il servizio lo riconosce con un 75% di confidenza.

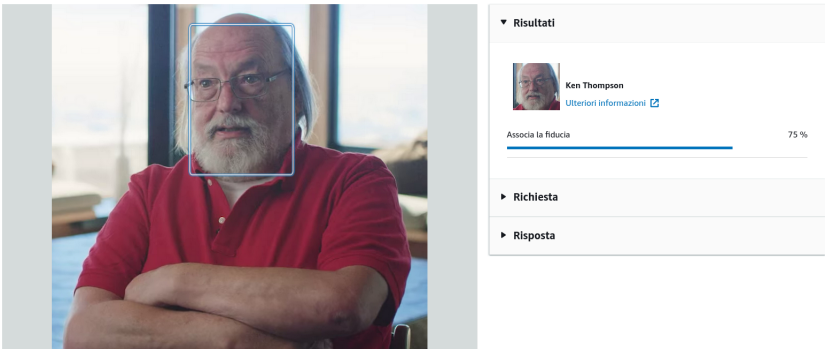

Figura 3.1: Risultato dell'analisi di un'immagine raffigurante Ken Thompson.

Un altro esempio, mostrato nella Figura 3.2, è l'analisi di un'immagine che raffigura Charlie Chaplin, attore famoso per numerosi film del cinema muto del periodo del Novecento, che viene riconosciuto con un grado di affidabilità del 97,6%.

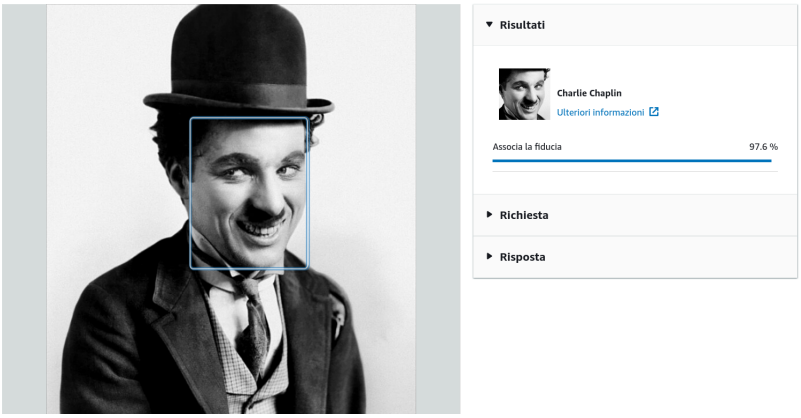

Figura 3.2: Risultato dell'analisi di un'immagine raffigurante Charlie Chaplin

Tuttavia, è molto interessante osservare come non sempre queste analisi vadano a buon fine; infatti, se si tenta di analizzare un'immagine di Charlie Chaplin quando era più giovane (e non aveva i suoi iconici baffi), il servizio non riesce a riconoscerlo (Figura 3.3).

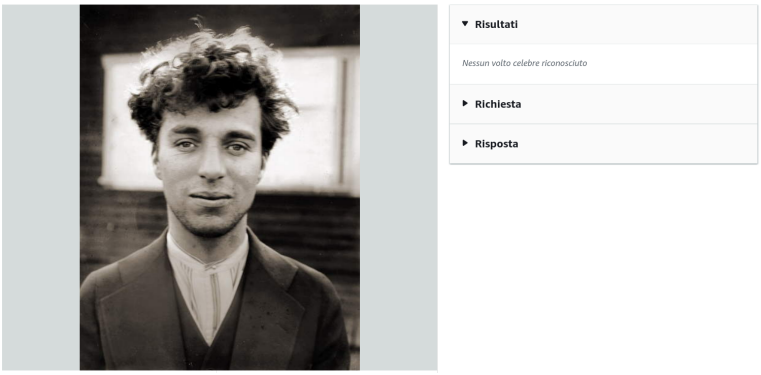

Figura 3.3: Analisi di un'altra immagine di Charlie Chaplin che, però, non viene riconosciuta dal sistema.

3.2.2 Confronto facciale

Questo servizio riceve come input due immagini e verifica in ognuna di esse se c'è un volto, in maniera simile a quanto fatto nel paragrafo precedente. Dopodiché, esso effettua un confronto tra le due immagini per vedere se ci sono corrispondenze tra i volti, indipendentemente dall'espressione, dalla presenza di barba e capelli o dall'età, sempre fornendo una percentuale che rappresenta il grado di affidabilità.

Per esempio, si può osservare che, prendendo due immagini del noto attore Keanu Reeves, una proveniente da una scena del film "John Wick" (2014), l'altra proveniente dal film "Matrix" (1999), Rekognition li identifica come la stessa persona, con un grado di confidenza del 99,7% (Figura 3.4).

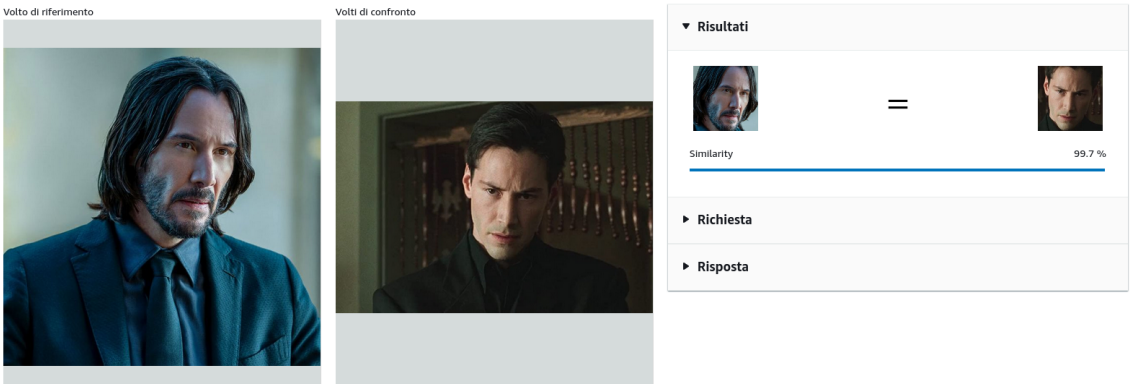

Figura 3.4: Risultato del confronto facciale di due immagini di Keanu Reeves

Un altro esempio interessante lo si può trovare riutilizzando le immagini, già analizzate nel precedente paragrafo, di Charlie Chaplin; si può osservare come i due volti vengano riconosciuti come la stessa persona con un grado di confidenza del 90,6% (Figura 3.5).

Dunque, è interessante osservare come questo strumento, che analizza soltanto due immagini, facendo direttamente il confronto dei lineamenti del volto, sia in grado di effettuare un riconoscimento più accurato rispetto a quello precedente, che, invece, analizzava il volto di una singola immagine e ne faceva il confronto tra i volti archiviati nel proprio database.

Tuttavia, neppure questo strumento produce sempre il risultato corretto. Nell'esempio riportato nella Figura 3.6, viene effettuato il confronto tra due immagini del noto cantante Chester Bennington, di cui una scattata nel 2014, quando faceva parte del gruppo musicale Linkin Park (quello con cui è diventato effettivamente famoso), mentre l'altra scattata nel

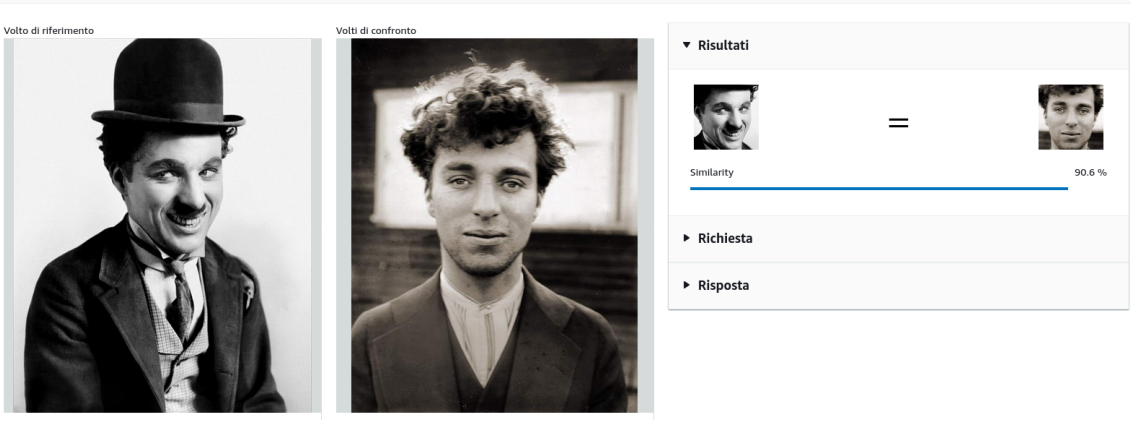

Figura 3.5: Risultato del confronto facciale di due immagini di Charlie Chaplin

1996, quando ancora faceva parte del gruppo Grey Daze. Seppur queste due foto ritraggono la stessa persona, Rekognition non è stato in grado di riconoscerlo.

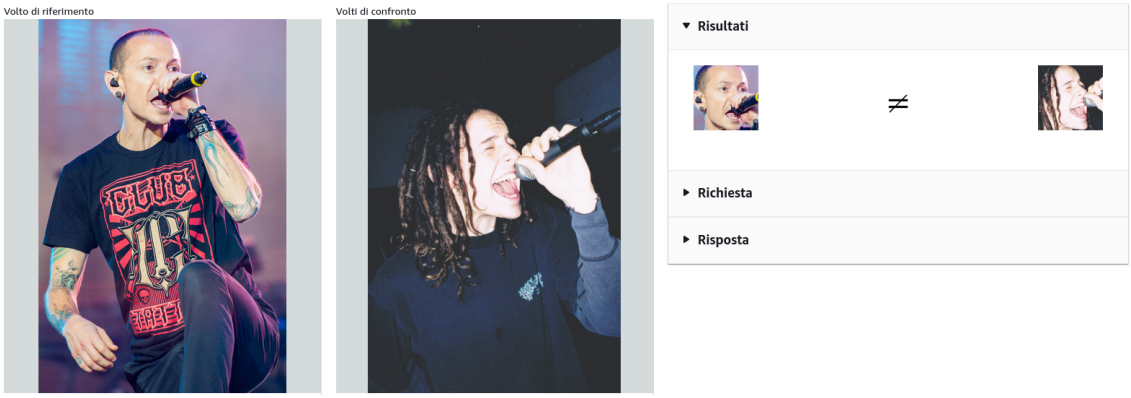

Figura 3.6: Risultato del confronto facciale di due immagini di Chester Bennington

3.2.3 Analisi dei video archiviati

Questo servizio di Amazon Rekognition è capace di analizzare un video per poi rilevare la presenza di etichette, volti, persone, celebrità, oggetti o attvità all'interno di esso. Per ogni elemento individuato all'interno del video, verrà, inoltre, specificato in quali intervalli è effettivamente presente nel video.

Per mostrarne il funzionamento, si può prendere come esempio uno spezzone (di 19 secondi) di una partita di scacchi e vedere quali elementi vengono individuati (Figura 3.7).

Per far vedere come vengono evidenziati gli intervalli in cui ogni elemento individuato è presente nel video, si può prendere come esempio uno degli spettatori della partita, che viene mostrato in secondo piano nel video (Figura 3.8).

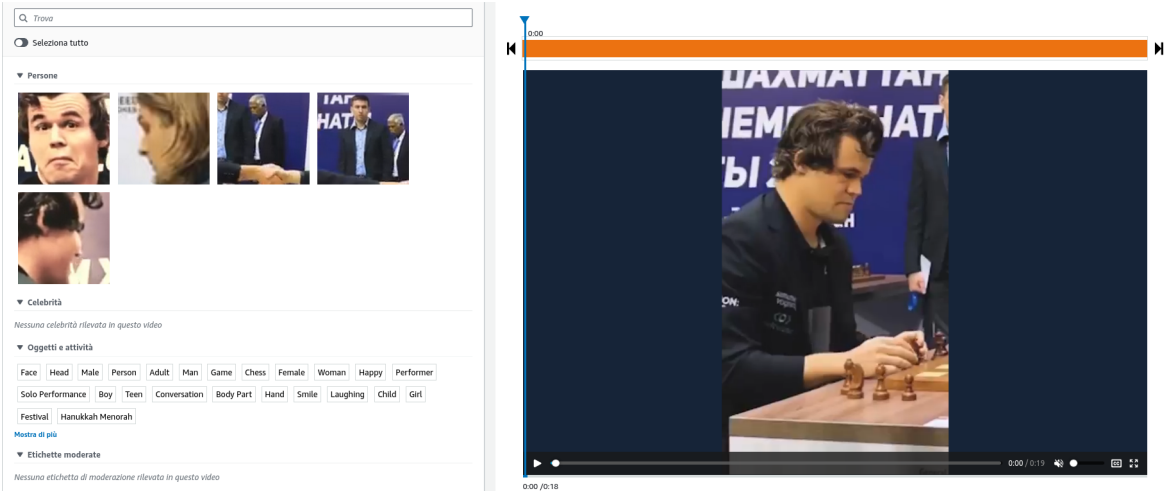

Figura 3.7: Analisi di un video raffigurante una partita di scacchi

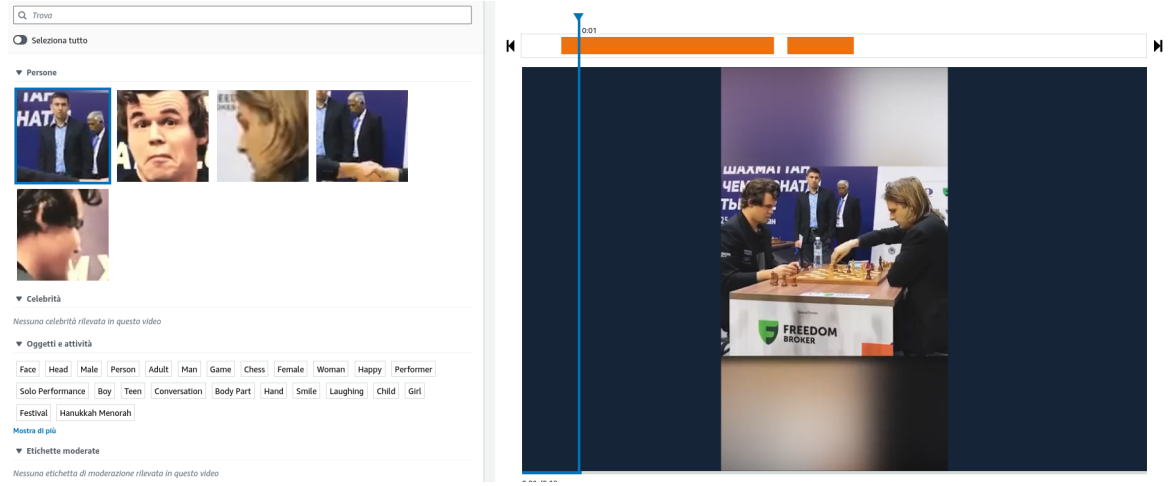

Figura 3.8: In questa figura vengono mostrati gli intervalli in cui la persona selezionata si vede nel video

Questo tool è decisamente più complesso dei due precedenti e, in quanto tale, è capace di fornire molte più informazioni a partire da un singolo file; tuttavia, tende abbastanza facilmente a individuare alcuni elementi che, in realtà, non sono presenti nel video; infatti, nella Figura 3.7 e nella Figura 3.8 si possono notare che, nella sezione "Oggetti e attività", sono presenti anche etichette come "Woman" e "Child", anche se non erano presenti donne e bambini in tutto il video, oppure come "Hannukah Menorah", che è un evento che nulla ha a che vedere con la partita di scacchi.

CAPITOLO 4

Traduzione

Questo capitolo tratterà il tema della traduzione automatica, fornendo una panoramica sulle sue caratteristiche essenziali, sulla sua evoluzione tramite Machine Learning e Natural Language Processing (NLP) e sulla sua applicazione tramite il servizio Amazon Translate. Successivamente, verranno analizzati alcuni case study di traduzione in tempo reale effettuati con Amazon Translate, i cui risultati verranno confrontati con quelli ottenuti tramite DeepL.

4.1 Traduzione automatica

La traduzione automatica (abbreviata in MT, dall'inglese Machine Translation) è un campo della linguistica computazionale e della scienza della traduzione che si occupa di tradurre testi da una lingua naturale a un'altra mediante programmi informatici.

Principalmente esistono tre approcci alla traduzione automatica:

- *Traduzione basata su regole*: è un approccio che, in genere, utilizza un processo traduttivo suddiviso in tre fasi; nella prima fase, nota come fase di analisi, il sistema esegue il parsing[1]delle frasi del testo di partenza e le trasforma in diagrammi ad albero (morfologici, sintattici e/o semantici); nella seconda fase, detta fase di trasferimento, gli alberi appena creati vengono trasformati in alberi con la struttura sintattica della lingua d'arrivo; nella terza fase, chiamata fase di generazione o sintesi, le parole della lingua di partenza vengono tradotte nella lingua d'arrivo e inserite nell'albero d'arrivo, seguendo le regole sintattiche di tale lingua, in modo da ottenere frasi di senso compiuto. Esiste, anche, una variante di questa tecnica, che consiste nel tradurre il testo di partenza in una lingua intermedia, la cui struttura è indipendente da quella della lingua originale e da quella della lingua finale, da cui, in seguito, si ottiene il testo nella lingua d'arrivo. La principale debolezza di questa tecnica è che effettuare una traduzione basata su regole è un approccio molto rigido e, al fine di garantire sempre traduzioni di qualità, costringerebbe gli autori del testo originale ad adeguare il loro stile di scrittura, il che non è applicabile naturalmente.

- *Traduzione automatica statistica*: viene abbreviata anche in SMT (dall'inglese, Statistical Machine Translation); questa tecnica necessita di due banche dati molto voluminosi, una di testi nella lingua di partenza, con le relative traduzioni nella lingua d'arrivo, e un'altra di testi solo nella lingua d'arrivo. Quando si vuole tradurre un nuovo testo, il sistema genera possibili traduzioni delle sequenze di parole che trova nel testo stesso sulla base delle corrispondenze che riscontra nella prima banca dati; tra le varie

proposte di traduzione, seleziona, poi, la migliore sulla base della seconda banca dati. Il vantaggio della traduzione automatica statistica è che, una volta impostato il sistema secondo le specifiche richieste dal cliente, questi ha a disposizione uno strumento in grado di fornire una discreta qualità traduttiva di testi simili tra loro; il lato negativo è che, affinché il sistema fornisca risultati di un certo livello, occorre utilizzare delle banche dati molto corpose.

- *Traduzione automatica neurale*: viene abbreviata anche in NMT (dall'inglese, Neural Machine Translation); si basa sull'uso di reti neurali artificiali per imparare a tradurre il testo da una lingua all'altra. Le reti neurali sono in grado di apprendere relazioni complesse tra le parole e le frasi, e sono, quindi, capaci di fornire traduzioni più accurate e naturali. Quest'ultimo approccio verrà approfondito successivamente.

4.1.1 Natural Language Processing

Prima di approfondire l'NMT è utile soffermarsi sul processo con cui un programma di Intelligenza Artificiale interagisce col linguaggio naturale in generale.

Questo processo va sotto il nome di elaborazione del linguaggio naturale o, dall'inglese, Natural Language Processing (anche abbreviato in NLP). Il suo scopo è quello di far "comprendere" il significato di un contenuto testuale, insieme a tutte le sfumature di contesto che la lingua comporta. Questa tecnologia risulta molto utile per estrarre informazioni importanti dai documenti ed organizzarli in categorie.

Il processo di elaborazione del testo è composto da numerose fasi, le quali aiutano il sistema a superare le ambiguità dovute al linguaggio naturale. Tali fasi comprendono:

- *Tokenization*: il testo viene scomposto in token, che possono corrispondere a spazi, parole, punteggiatura o frasi; durante questa fase, è importante saper distinguere il significato dei caratteri a seconda del contesto; per esempio, il punto, di norma, indica la fine della frase, ma in altri casi può essere parte di un'abbreviazione, di una data o di un link.

- *Analisi morfologica e lessicale*: è il processo di determinazione della parte del discorso di una particolare parola, o pezzo di testo, in base al relativo uso e al contesto; per esempio, la parola "riso" può essere inteso come verbo nella frase "ho riso molto" oppure come sostantivo nella frase "ho mangiato riso".

- *Analisi sintattica e generazione di parse tree*: l'analisi sintattica consente di individuare le varie parti del discorso e dedurre quali funzioni svolgono all'interno della frase (soggetto, predicato o complemento), verificando, dunque, che l'ordine delle parole sia corretto. Anche questa fase è soggetta ad ambiguità, per esempio, nella frase "Giorgio vide Giulio con un telescopio" non si può capire con certezza se il telescopio appartenga a Giorgio o a Giulio.

- *Named Entity Recognition*: consiste nell'individuazione, tramite regole o approcci statistici di Machine Learning, di parole, o gruppi di parole, che rappresentano entità del mondo reale, come nomi di persone, luoghi, dati, compagnie.

- *Analisi semantica*: in questa fase si cerca di dedurre il significato di un'intera frase partendo dal significato di ciascun termine che la compone e dalle relazioni esistenti tra gli stessi.

[1] Il parsing è il processo con cui un apposito programma analizza un flusso di dati in modo da determinare la correttezza della sua struttura grazie ad una data grammatica formale.

- *Analisi del discorso*: in questa fase vengono messe a confronto le varie frasi presenti nel testo e vengono individuate le parole che si riferiscono alle stesse entità, in modo da garantire una comprensione delle parole coerente con l'intero testo.

4.1.2 Neural Machine Translation

Questa tecnica di traduzione cominciò a diffondersi a partire dal 2010, superando alcune difficoltà che venivano incontrate dalla SMT, fino a diventare l'approccio attualmente dominante.

Come la SMT, la NMT viene "allenata" attraverso la raccolta di un vastissimo corpus di traduzioni parallele nei linguaggi sorgente e destinazione; tuttavia, a differenza della SMT, questo nuovo tipo di architettura opera attraverso l'uso di reti neurali, in particolare utilizza le Recurrent Neural Networks (RNN).

Il processo di traduzione consiste in tre fasi, ovvero:

- *Training*: in questa fase viene fatto analizzare alla rete neurale il corpus contenente le traduzioni bilingue; per ogni parola individuata, la rete memorizzerà una serie di informazioni sia di tipo semantico sia di tipo grammaticale, tramite la tecnica del word embedding[2]; questo viene fatto separatamente per entrambe le lingue, dopodiché ogni termine della prima lingua verrà associato a uno o più termini della seconda lingua che contengono informazioni simili.

- *Encoder*: in questa fase la rete neurale analizza il testo di partenza e rieffettua un'operazione di word embedding, cercando tutte le connessioni tra le varie parole presenti nel testo.

- *Decoder*: una volta ottenuta l'immagine vettoriale del testo di partenza, il sistema di decodifica della NMT è in grado di leggere i valori delle singole unità che compongono la frase nella lingua di partenza e trasformarli in un output nella lingua di arrivo.

4.1.3 Vantaggi e limiti della traduzione automatica

Alla luce di quanto discusso nei paragrafi precedenti, la traduzione automatica risulta essere uno strumento molto potente, in quanto permette di tradurre grandi volumi di testo molto più velocemente di quanto possano fare gli operatori umani, diminuendo notevolmente i costi. Per di più, generalmente, un servizio di traduzione automatica è sempre disponibile; pertanto, può aiutare a rendere le informazioni più accessibili, superando i problemi dovuti alle barriere linguistiche.

Tuttavia, è una tecnologia molto limitata e, in ogni caso, necessita della supervisione umana, al fine di garantire traduzioni di sufficiente qualità. I principali limiti della traduzione automatica sono i seguenti:

- *Comprensione del contesto*: attualmente le macchine hanno ancora grandi difficoltà a comprendere il contesto linguistico in cui le parole vengono utilizzate.

- *Sfumature linguistiche*: le lingue sono piene di sfumature, come il tono, l'ironia e l'umorismo, che possono essere difficili da catturare in una traduzione automatica.

[2] Il word embedding è una tecnica di NLP che associa ad ogni parola un vettore che ne rappresenta il significato semantico e sintattico; in questo spazio vettoriale, vettori che sono più vicini corrispondono a parole simili nel significato.

- *Errori grammaticali*: nonostante i progressi di questa tecnologia, possono verificarsi ancora svariati errori in seguito alla traduzione, soprattutto con le lingue più complesse e meno diffuse.
- *Traduzione di nomi propri e termini tecnici*: i nomi propri e i termini tecnici possono essere difficili da tradurre correttamente, poiché spesso non hanno un equivalente diretto in un'altra lingua.
- *Cultura e costumi locali*: le macchine non hanno una comprensione profonda della cultura e dei costumi locali, il che può influenzare la qualità della traduzione.

Proprio a causa di questi difetti, in alcuni contesti delicati, come quello giuridico, la traduzione automatica deve essere sottoposta alla revisione umana, al fine di evitare errori (con conseguenze potenzialmente pesanti) dovuti alla mancata comprensione di alcune sfumature. In alcuni tribunali è, addirittura, completamente vietata la traduzione automatica durante qualsiasi procedimento legale.

4.2 Analisi dei case study

Per mettere alla prova l'IA nell'ambito della traduzione verranno presi tre porzioni di testo provenienti da contesti differenti e verrà analizzata la qualità della traduzione.

Tali traduzioni verranno effettuate dal servizio di AWS Amazon Translate e i corrispondenti risultati verranno confrontati con quelli ottenuti tramite DeepL, da molti considerato il miglior traduttore automatico attualmente esistente.

Gli esempi che verranno considerati sono i seguenti:

- la definizione fornita da Treccani di Intelligenza Artificiale;
- un estratto di "La patente", uno dei testi che compongono le "Novelle per un anno" di Luigi Pirandello;
- alcuni versi tratti da "L'infinito", poesia di Giacomo Leopardi.

Per motivi di chiarezza e semplicità, ognuno di questi tre esempi verrà tradotto in inglese e, in seguito, ritradotto in italiano per valutare la qualità della traduzione.

4.2.1 Definizione di IA di Treccani

Come primo esempio di studio per comprendere come viene svolta la traduzione e quanto risulta fedele al senso originale del testo, verrà selezionata una voce dell'enciclopedia Treccani. In particolare, si è scelta la definizione che l'enciclopedia fornisce di Intelligenza Artificiale:

> *Disciplina che studia se e in che modo si possano riprodurre i processi mentali più complessi mediante l'uso di un computer. Tale ricerca si sviluppa secondo due percorsi complementari: da un lato l'i. artificiale cerca di avvicinare il funzionamento dei computer alle capacità dell'intelligenza umana, dall'altro usa le simulazioni informatiche per fare ipotesi sui meccanismi utilizzati dalla mente umana.*

Se si fa traddure questo paragrafo ad Amazon Translate, il risultato che si ottiene è questo:

> *Discipline that studies whether and how the most complex mental processes can be reproduced through the use of a computer. This research is developed according to two complementary paths: on the one hand, artificial intelligence seeks to bring the functioning of computers closer to the capabilities of human intelligence, on the other hand, it uses computer simulations to make assumptions about the mechanisms used by the human mind.*

Si può osservare che la traduzione è molto accurata in questo caso; addirittura viene colta automaticamente l'abbreviazione "i. artificial" e tradotta correttamente, dimostrando una certa comprensione del contesto trattato nel paragrafo.

A maggior riprova del fatto che la traduzione è corretta, viene mostrato di seguito il testo tradotto nuovamente in italiano:

> *Disciplina che studia se e come i processi mentali più complessi possono essere riprodotti attraverso l'uso di un computer. Questa ricerca si sviluppa secondo due percorsi complementari: da un lato, l'intelligenza artificiale cerca di avvicinare il funzionamento dei computer alle capacità dell'intelligenza umana, dall'altro utilizza simulazioni al computer per formulare ipotesi sui meccanismi utilizzati dalla mente umana.*

Come si può notare, quest'ultimo testo risulta essere quasi identico a quello originale. Per questo specifico case study non verrà riportata la traduzione effettuata con DeepL, dal momento che anch'essa è molto accurata e ha prodotto risultati molto simili.

4.2.2 Testo tratto da "La patente" di Luigi Pirandello

Per l'analisi di questo case study è stata scelta la novella "La patente", la cui storia consiste nella vicenda di un uomo, Rosario Chiarchiaro, che ha perso il lavoro e vive in condizioni di miseria, a causa della fama di iettatore che, suo malgrado, gli è stata attribuita. Nello specifico, il testo che verrà sottoposto alla traduzione corrisponde alla scena in cui Chiarchiaro, si rivolge al giudice D'Andrea, per farsi dare una patente ufficiale che attesti legalmente la sua capacità di portare iella[3].

Il testo da tradurre è il seguente:

> *- Lei, padrone mio, per esercitare codesta professione di giudice, anche così male come la esercita, mi dica un po', non ha dovuto prender la laurea?*
> *- La laurea, sì.*
> *- Ebbene, voglio anch'io la mia patente, signor giudice! La patente di jettatore. Col bollo. Con tanto di bollo legale! Jettatore patentato dal regio tribunale.*

Traduzione con Amazon Translate

Il testo viene tradotto da Amazon Translate in questo modo:

> *- You, my master, in order to practice this profession of judge, even as badly as you practice it, didn't you have to graduate?*
> *- The degree, yes.*
> *- Well, I want my driver's license too, judge! The jettler's license. With the stamp. Complete with a legal stamp! Jet jet patented by the Royal Court.*

[3] L'intenzione di Chiarchiaro sarebbe quella di essere legalmente autorizzato a chiedere denaro per stare lontano dalla gente, dal momento che non può trovare nessun lavoro.

Quest'ultima traduzione non ha alterato il senso del dialogo in maniera eccessiva, ma comunque presenta alcuni problemi. Innanzitutto, la risposta del giudice, seppur tradotta correttamente, risulta poco naturale all'interno del dialogo, invece di rispondere "The degree, yes." sarebbe stato più logico se avesse detto semplicemente "Yes, I did.".

Un altro problema è la traduzione di "patente", che in questo caso, è stato tradotto in maniera troppo letterale ed è diventata "driver's license", riferendosi alla patente per poter guidare i veicoli.

Inoltre, la parola "jettatore" non è stata per niente riconosciuta dal sistema (a causa del fatto che in genere si scrive "iettatore") e viene tradotta in maniera improvvisata, una volta con "jettler", che è una parola che non esiste in inglese, una con "jet jet".

Se si traduce questo testo nuovamente in italiano, si otterrà il seguente risultato:

> *- Tu, mio maestro, per esercitare questa professione di giudice, anche se male la eserciti, non dovevi laurearti?*
> *- La laurea, sì.*
> *- Beh, anch'io voglio la mia patente, giudice! La patente del jettler. Con il francobollo. Completo di timbro legale! Jet jet brevettato dalla Corte Reale.*

Quest'ultima traduzione si porta con sé i problemi della precedente traduzione, ma a questi se ne aggiunge un altro. Il termine "master" è stato ora tradotto in "maestro"; questo è dovuto al fatto che, in inglese, il termine "master" ha molte sfumature e si può usare sia per riferirsi a un insegnante, sia per riferirsi a qualcuno che ha una posizione di comando (anche se nel testo originale è in senso figurato). Inoltre, in questa seconda traduzione, si è persa un po' di formalità, e quello che all'inizio era un "lei" è diventato un "tu".

Traduzione con DeepL

Riprendendo il testo originale, verrà ora mostrata la traduzione effettuata con DeepL:

> *- You, my master, in order to practice this coded profession of a judge, even as badly as you practice it, tell me a little bit, didn't you have to take a degree?*
> *- The degree, yes.*
> *- Well, I want my license too, Mr. Judge! Jettor's license. With a stamp. With a legal stamp! Jettatore licensed by the royal court.*

Questa traduzione riscontra problemi simili a quelli della traduzione fatta con Amazon Translate, ma, in questo caso, la parola "patente" viene tradotta in maniera più corretta utilizzando "license", che è un termine più generale per intendere una patente o una licenza. Tuttavia, DeepL non è stato in grado di riconoscere la parola "codesta", facendola diventare "this coded", che, oltre a non avere senso in quel contesto, significa una cosa completamente differente.

Infatti, se si ri-traduce il testo in italiano, si ottiene:

> *- Lei, mio maestro, per esercitare questa professione codificata di giudice, anche se male come la esercita lei, mi dica un po', non ha dovuto prendere una laurea?*
> *- La laurea, sì.*
> *- Ebbene, voglio anch'io la mia licenza, signor giudice! Patente di jettatore. Con un timbro. Con un timbro legale! Jettatore con licenza della corte reale.*

4.2.3 Testo tratto da "L'infinito" di Giacomo Leopardi

L'ultimo case study scelto nell'ambito della traduzione è un estratto dalla poesia "L'infinito" di Giacomo Leopardi, che recita[4]:

> *Sempre caro mi fu quest'ermo colle, e questa siepe, che da tanta parte dell'ultimo orizzonte il guardo esclude.*
> *Ma sedendo e mirando, interminati spazi di là da quella, e sovrumani silenzi, e profondissima quiete io nel pensier mi fingo, ove per poco il cor non si spaura.*

Traduzione con Amazon Translate

La traduzione fornita da Amazon Translate è la seguente:

> *Always dear to me was this empty hill, and this hedgerow, which the eye excludes from so much of the last horizon.*
> *But sitting and looking, endless spaces beyond that, and superhuman silences, and very deep stillness, I pretend in my thoughts, where the heart almost doesn't get scared.*

La maniera con cui sono state formulate le frasi in questa poesia mette molto in difficoltà l'IA, che, di conseguenza, restituirà delle frasi che risultano essere poco accettabili in inglese; per di più, la traduzione ha del tutto perso il significato originale in alcuni passaggi. Per esempio, alla fine del testo la traduzione di "ove per poco il cor non si spaura" è diventata "where the heart almost doesn't get scared", che, detto in questo modo, sembra significare il contrario di quello che c'è scritto nella poesia; se l'autore, con la prima frase, intendeva che per poco non si era spaventato, la traduzione fa sembrare che in realtà è rimasto quasi impassibile.

Le difficoltà legate alla traduzione risultano ancora più evidenti quando viene fatto ri-tradurre il testo in italiano:

> *Mi è sempre stata cara questa collina vuota e questa siepe, che l'occhio esclude da gran parte dell'ultimo orizzonte.*
> *Ma seduto a guardare, tra spazi infiniti e silenzi sovrumani e un'immobilità molto profonda, fingo nei miei pensieri, dove il cuore quasi non si spaventa.*

Si noti come la traduzione di "ermo colle" è diventata "collina vuota", che, a grandi linee, è corretta come traduzione, ma non rende del tutto l'idea di un luogo desolato. Il passaggio successivo è ancora più problematico "e questa siepe, che l'occhio esclude da gran parte dell'ultimo orizzonte", da queste parole sembra che l'osservatore veda una siepe, ma la ignora per guardare qualcos'altro, quando, in realtà, il poeta intendeva che la siepe gli impediva di vedere quello che c'è oltre, il che gli ha permesso di vagare con l'immaginazione e fare riflessioni più profonde. Invece, nelle ultime due righe è stata, più o meno, rispettata la struttura originale del testo, tranne per quella "quiete" che è diventata "immobilità" che ha fatto perdere un po' di significato.

[4] Ai fini della traduzione è risultato necessario rompere lo schema originale della poesia e mostrare il testo come se fosse scritto in prosa.

Traduzione con DeepL

La traduzione fornita da DeepL è la seguente:

> *Always dear to me was this ermo hill, and this hedge, which from so much of the last horizon the look excludes.*
> *But sitting and gazing, interminable spaces beyond it, and superhuman silences, and profoundest quietness I in thought finish myself, where for a little the heart is not frightened.*

In questo caso, si riscontrano ulteriori problemi, come il fatto che la parola "ermo" non viene tradotta per niente oppure il fatto che "per poco" viene tradotto in "for a little", che, in questo caso, non è corretto perché significherebbe "per poco tempo" (sarebbe stato più opportuno tradurlo in "almost"). Un altro problema abbastanza importante è un passaggio alla fine del testo, dove "mi fingo", in qualche modo, è diventato "I... finish", che ha un significato totalmente differente.

Dunque, se si ri-traduce il testo in italiano, si ottiene il seguente risultato:

> *Sempre caro mi fu questo colle ermo, e questa siepe, che da tanta parte dell'ultimo orizzonte lo sguardo esclude.*
> *Ma sedendo e guardando, interminabili spazi al di là di esso, e sovrumani silenzi, e profondissima quiete io nel pensiero mi finisco, dove per un poco il cuore non si spaventa.*

Quest'ultimo testo risulta essere molto simile al testo originale; tuttavia, è molto importante sottolineare che questa somiglianza è, in realtà, una coincidenza, dato che, nella fase di traduzione in inglese, il software ha commesso alcuni errori con cui poi, però, è stato coerente nella fase di ri-traduzione in italiano.

In conclusione, la traduzione è, come appena visto in questo capitolo, un'attività molto complessa, al momento ancora difficile da automatizzare in maniera efficace.

CAPITOLO 5

Estrazione di testi

Questo capitolo è dedicato al tema dell'estrazione di testi. Inizialmente verrà dato uno sguardo più ampio sul tema della digitalizzazione dei documenti e dei Big Data per comprendere, da un lato, che scopo ha l'attività di estrazione di testi, e dall'altro quali implicazioni ha questa attività su larga scala. Dopodiché, verranno analizzati 3 casi di studio, di cui ognuno rappresenterà uno scenario concreto, che mostreranno le potenzialità di questa tecnologia.

5.1 Introduzione all'estrazione di testi

L'estrazione di testi è un'attività nell'ambito dell'NLP che fa ampio uso delle tecnologie fornite dal Machine Learning. Essa rientra in temi più ampi, come la digitalizzazione dei documenti e i Big Data.

5.1.1 Digitalizzazione dei documenti

La digitalizzazione dei documenti è il processo di conversione dei documenti cartacei in formato digitale, rendendoli accessibili e facilmente gestibili attraverso dispositivi elettronici.

Gestire i documenti in maniera digitale presenta numerosi vantaggi, tra cui:

- *Maggiore velocità*: un documento digitale è molto più veloce da creare, condividere e archiviare rispetto ai documenti cartacei; inoltre, a parità di precisione, la ricerca dei dati al suo interno è molto più rapida.
- *Minori costi*: la digitalizzazione dei documenti consente di ridurre i costi associati alla gestione dei documenti cartacei, come la stampa, la spedizione e l'archiviazione.
- *Maggiore accessibilità*: i documenti digitali sono accessibili da qualsiasi parte del mondo, permettendo di risparmiare i costi dovuti alla spedizione.
- *Maggiore sicurezza*: i documenti digitalizzati sono più sicuri in quanto possono essere crittografati e archiviati su server protetti; inoltre, permettono alle organizzazioni di esercitare maggiore controllo su chi ha accesso ai vari documenti.
- *Migliore permanenza*: i documenti fisici possono rovinarsi, strapparsi o rompersi nel tempo, rischiando di far perdere alcune informazioni; utilizzare documenti digitali permette di evitare questo problema, purché si effettui regolarmente un processo di backup.

Tipicamente, i documenti che sono più adatti ad essere digitalizzati sono documenti amministrativi (come documenti fiscali e tributari o i documenti di identità), contratti (di lavoro o di vendita), curriculum lavorativi e documenti legali (come le sentenze).

5.1.2 Big Data

La digitalizzazione dei documenti comporta come conseguenza il salvataggio dei loro dati su qualche dispositivo di archiviazione; questo implica che bisognerà salvare in memoria una quantità sempre maggiore di dati, il che, a sua volta, comporterà una serie di problemi a cascata, come il mantenimento, la sicurezza e la ricerca dei dati.

Quando i volumi di dati diventano troppo grandi sono necessarie delle tecnologie e dei metodi analitici molto specifici, che permettano di estrarre velocemente i dati nonostante l'enorme mole; in tal caso si parla di Big Data.

In realtà, non esiste una separazione netta tra Big Data e altri dati, né esiste una dimensione che determini in modo univoco quando un volume di dati diventa Big Data. In genere si parla di Big Data quando l'insieme di dati è talmente grande e complesso che richiede la definizione di nuovi strumenti e metodologie per estrapolare, gestire e processare informazioni entro un tempo ragionevole.

Le caratteristiche dei Big Data sono studiate dal cosiddetto modello delle "3V" (Figura 5.1), che corrispondono alle seguenti proprietà:

- *Volume*: corrisponde alla quantità di dati generati e memorizzati; in genere, si parla di Big Data quando la dimensione è nell'ordine dei petabyte (circa 10^{15} byte) o superiore.
- *Varietà*: i dati possono essere strutturati, non strutturati o semi-strutturati e possono provenire da fonti eterogenee; i dati semi-strutturati e non strutturati richiedono un'ulteriore elaborazione preliminare per poter ricavare informazioni utili e recuperabili.
- *Velocità*: ci sono dati che possono essere elaborati e gestiti offline; tuttavia, ci sono anche altri dati che devono essere gestiti quasi in tempo reale, se non proprio in tempo reale; si pensi, ad esempio, ai dati provenienti dai sensori o ai file di log dei server elaborati nell'ambito della sicurezza informatica.

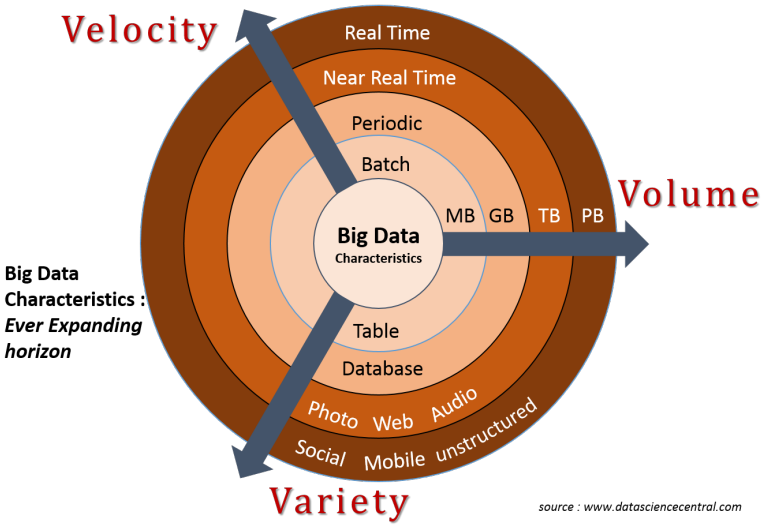

Figura 5.1: Rappresentazione grafica del modello delle 3V

Con il tempo, sono state introdotte altre 2 "V" nel modello; esse corrispondono a:

- *Veridicità*: considerando la varietà dei dati e la velocità alla quale possono variare, è molto probabile che non si riesca sempre a garantire la stessa qualità dei dati; pertanto, è importante assegnare un indice di veridicità agli stessi, in modo da avere una misura dell'affidabilità.
- *Valore*: esso rappresenta il valore nelle informazioni che possono essere raggiunte con l'elaborazione e l'analisi di grandi set di dati; questo parametro è importante per capire quanto conviene investire per elaborare i dati desiderati.

Lo studio di tecniche per gestire Big Data è giustificato dal grande interesse che ha suscitato nelle aziende (tanto sono stati investiti oltre 15 miliardi di dollari negli ultimi anni) e dalla crescita esponenziale dei dati (mostrata nella Figura 5.2, misurata in Exabyte[1]), che renderà sempre più necessario ricorrere a questa tecnologia.

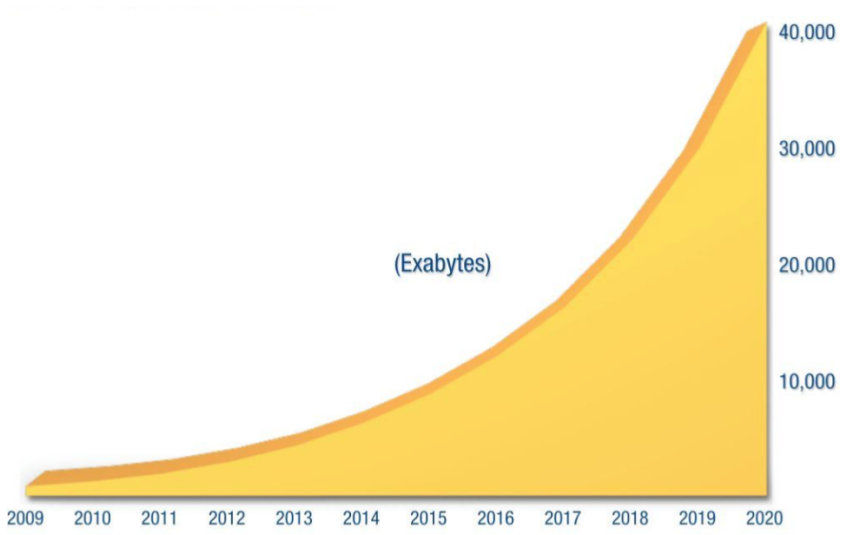

Figura 5.2: Andamento esponenziale della crescita dei dati negli anni

L'estrazione di conoscenza dai Big Data e l'impiego della stessa per il miglioramento delle attività decisionali sono subordinati alla definizione di alcuni processi. Ciascun processo modifica lo stato ed il contenuto di quelli precedenti, contribuendo a convertire moli di dati ancora grezzi in valore, e dunque ad arricchire il modello analitico dei dati. Tali processi sono raggrupati in alcune fasi che compongono il ciclo di vita dei Big Data (Figura 5.3). Le fasi sono le seguenti:

- *Generazione*: le fonti di dati sono eterogenee e possono essere suddivise in tre categorie, ovvero, dati generati dall'uomo, dati generati da macchine e dati generati da attività commerciali.
- *Acquisizione*: include la raccolta, la trasmissione e l'elaborazione dei dati; i dati raccolti possono essere ridondanti o incoerenti, e ciò può dar luogo ad analisi non accurate; pertanto, i dati devono subire processi di elaborazione.
- *Immagazzinamento*: i dati vengono memorizzati per un uso successivo; questa fase richiede anche tecniche specifiche di integrazione dei dati, visto che i dati possono essere molto eterogenei.

[1] 1 Exabyte corrisponde a circa 10^{18} byte.

- *Analisi*: si utilizzano metodi statistici e informatici per estrapolare e mettere in relazione i dati, al fine di scoprire i legami tra fenomeni diversi e prevedere quelli futuri.
- *Intepretazione*: i risultati dell'analisi vengono interpretati e utilizzati per prendere decisioni.

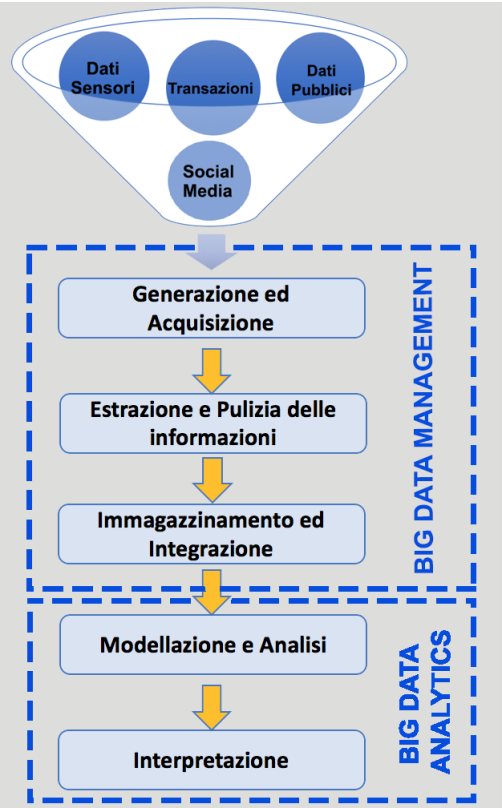

Figura 5.3: Rappresentazione grafica delle fasi che caratterizzano il ciclo di vita dei Big Data

5.1.3 Amazon Textract

Amazon Textract è un servizio di Machine Learning che estrae automaticamente testo, scrittura a mano, elementi di layout e dati da documenti scansionati. Questo servizio va oltre il semplice riconoscimento ottico dei caratteri (OCR) per identificare, capire ed estrarre dati da moduli e tabelle.

Attualmente, molte aziende devono estrarre dati manualmente da documenti scansionati in vari formati, come PDF, immagini, tabelle e moduli, oppure utilizzando software OCR semplici che richiedono una configurazione manuale, la quale, spesso, deve essere aggiornata ad ogni modifica del modulo. Per superare questi processi dispendiosi e manuali, Textract utilizza il Machine Learning per leggere ed elaborare accuratamente qualsiasi tipo di documento, estrarre testo, scrittura a mano, tabelle e altri dati senza richiedere intervento manuale.

Le principali funzionalità fornite da Textract sono le seguenti:

- *Rilevamento di testo*: questa funzione consente di rilevare il testo in documenti di input a pagina singola o multipagina.
- *Analisi di documenti*: questa funzione consente di analizzare le relazioni tra i vari elementi di testo presenti in un documento.

- *Analisi di fatture e ricevute*: questa funzione consente di estrarre i dati rilevanti, come le informazioni di contatto, gli articoli acquistati e il nome del fornitore, da quasi tutte le fatture o ricevute senza la necessità di modelli o configurazione.
- *Analisi di documenti di identità*: questa funzione consente di estrarre informazioni pertinenti da documenti di identità, come passaporti o patenti di guida.

5.2 Analisi dei case study

I case study scelti per mettere alla prova le funzionalità di Amazon Textract sono i seguenti:

- *Case study n°1*: verrà analizzato un documento di autodichiarazione generico compilato a mano.
- *Case study n°2*: verrà analizzato un fac-simile di una bolletta.
- *Case study n°3*: verrà effettuata la ricerca dei dati presenti nel documento analizzato nel case study precedente tramite query apposite.

5.2.1 Estrazione di testo da modulo di autodichiarazione

Il primo case study preso in esame è un modulo di autodichiarazione generico compilato e firmato a mano, di cui tutte le varie informazioni personali (quali nome, cognome, codice fiscale, ecc.), sono di fantasia. Il contenuto del modulo viene mostrato di seguito, nella Figura 5.4 (naturalmente, anche il nome dell'azienda è inventato).

Dichiarazione sostitutiva di certificazione

(art. 46 D.P.R.28 dicembre 2000 n. 445)

Il/la Sottoscritto/ MARIO ROSSI c.f. MRSS41548484880 1P

nato a MILANO (MI) il 19/07/91

residente a MILANO (MI) in VIA ROMA n° 1

consapevole che chiunque rilascia dichiarazioni mendaci è punito ai sensi del codice penale e delle leggi speciali in materia, ai sensi e per gli effetti dell'art. 46 D.P.R. n. 445/2000

DICHIARA

DI AVER CONSEGUITO LA LAUREA MAGISTRALE IN INGEGNERIA INFORMATICA E DELL'AUTOMAZIONE PRESSO L'UNIVERSITÀ POLITECNICA DELLE MARCHE CON VOTO 110 E LODE NELL'ANNO ACCADEMICO 2015-2016 E DI AVER AVUTO ESPERIENZE LAVORATIVE NEGLI ANNI TRA IL 2017 E IL 2022 PRESSO L'AZIENDA PIATTAFORMA CYBERSECURITY S.R.L. NEI RUOLI DI SECURITY ANALIST E MACHINE LEARNING SPECIALIST

Luogo, 22/03/2023

M. R.

Firma del dichiarante
(per esteso e leggibile)

Ai sensi dell'art. 10 della legge 675/1996 e successive modificazioni, le informazioni indicate nella presente dchiarazione verranno utilizzate unicamente per le finalità per le quali sono state acquisite.

Figura 5.4: Autodichiarazione presa come riferimento per l'analisi del case study

Se si fa analizzare questa immagine a Textract, il risultato che si ottiene è il seguente (Figura 5.5).

Figura 5.5: Risultati dell'analisi effettuata da Textract

A giudicare da questi risultati, possiamo osservare che Textract è in grado di riconoscere buona parte del testo presente nel documento; l'unico difetto è che, in alcuni casi, le lettere non vengono riconosciute come tali e, a volte, vengono confuse con caratteri numerici o speciali; per esempio, la "I" di "VIA ROMA" è stata confusa con il carattere "/" e la lettera "D" di "DI AVER CONSEGUITO" è stata scambiata per la cifra "0".

Questa analisi può essere svolta anche per testi scritti in corsivo, di cui viene mostrato un esempio nella Figura 5.6 (il modulo è lo stesso della Figura 5.4, ma si riporta solo una parte del testo).

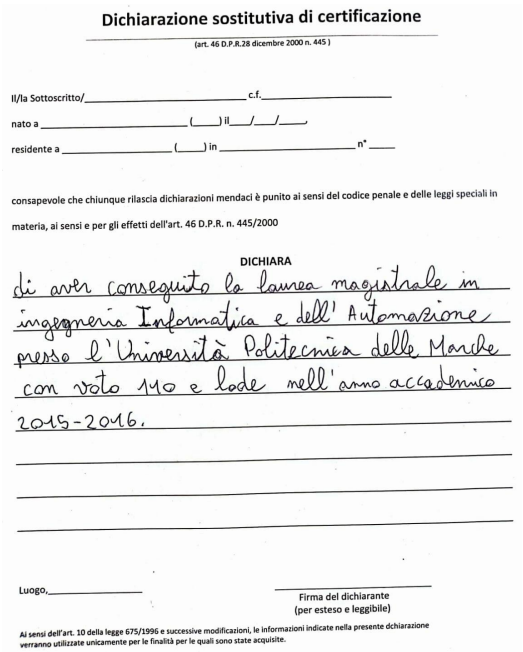

Figura 5.6: Esempio di testo scritto in corsivo

I risultati dell'analisi di questo testo scritto in corsivo sono riportati nella Figura 5.7.

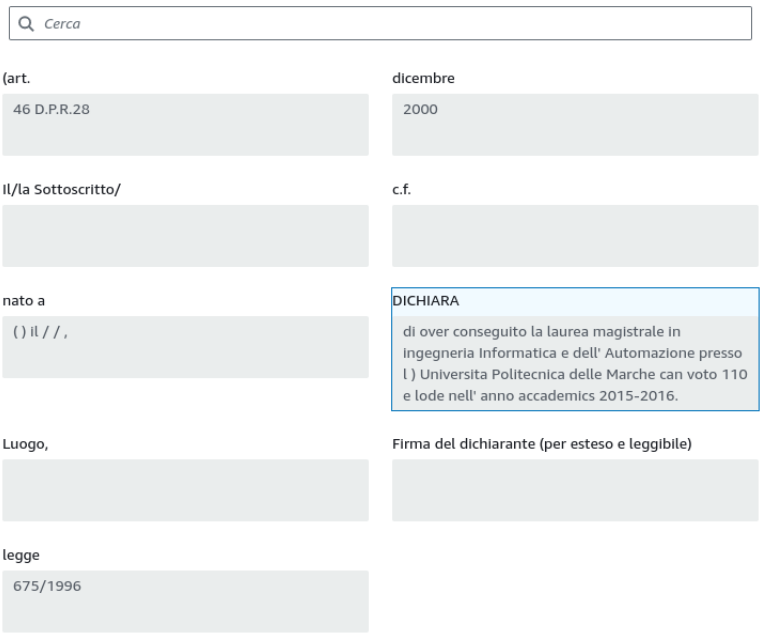

Figura 5.7: Risultati dell'analisi del testo scritto in corsivo.

I risultati ottenuti da quest'ultima analisi dimostrano che Textract è in grado di riconoscere testi scritti in corsivo con una buona accuratezza (paragonabili a quella dell'analisi del testo in stampatello); tuttavia, è importante ricordare che questo non vale per tutti i testi scritti in corsivo e l'accuratezza può cambiare notevolmente a seconda della calligrafia.

In conclusione, uno strumento di questo tipo potrebbe essere utile per acquisire e trasferire velocemente moduli, come autodichiarazioni, permessi di qualunque tipo, report e altro ancora, permettendo di semplificare parte della burocrazia (sia in termini di velocità che in termini di costi).

5.2.2 Estrazioni di dati dal fac-simile di una bolletta

In questo case study verrà condotta un'analisi simile a quella fatta per il case study precedente; in questo caso, però, tutti i testi sono scritti al computer. Nello specifico, verrà analizzato il fac-simile di una bolletta, sempre con nomi e informazioni di fantasia, al fine di valutarne l'accuratezza dei testi estratti e osservare come i dati raccolti vengono organizzati.

Il fac-simile preso in esame verrà mostrato nella Figura 5.8.

I risultati dell'analisi di questo fac-simile sono riportati nella Figura 5.9 e sono organizzati in vari campi, di cui ognuno è identificato tramite una chiave e contiene al proprio interno un'informazione che funge da valore.

Naturalmente, non vengono riportati tutti dati per una mera questione di spazio. Ad ogni modo, si può notare che l'estrazione di testo è estremamente accurata e non riporta problemi particolarmente significativi. Per quanto riguarda l'estrazione e l'organizzazione dei dati, si può osservare che una buona parte di essi viene estratta correttamente, ma, in alcuni campi, capita che il valore non viene individuato (come nel campo "COMUNICAZIONE RECLAMI"), anche se nel documento originale è presente; di conseguenza, tali campi rimangono vuoti. Pertanto, questo servizio si rivela molto potente per digitalizzare i documenti, ma, allo stato attuale, necessita ancora un'attenta supervisione da parte di un operatore umano, altrimenti si rischia di perdere alcuni dati.

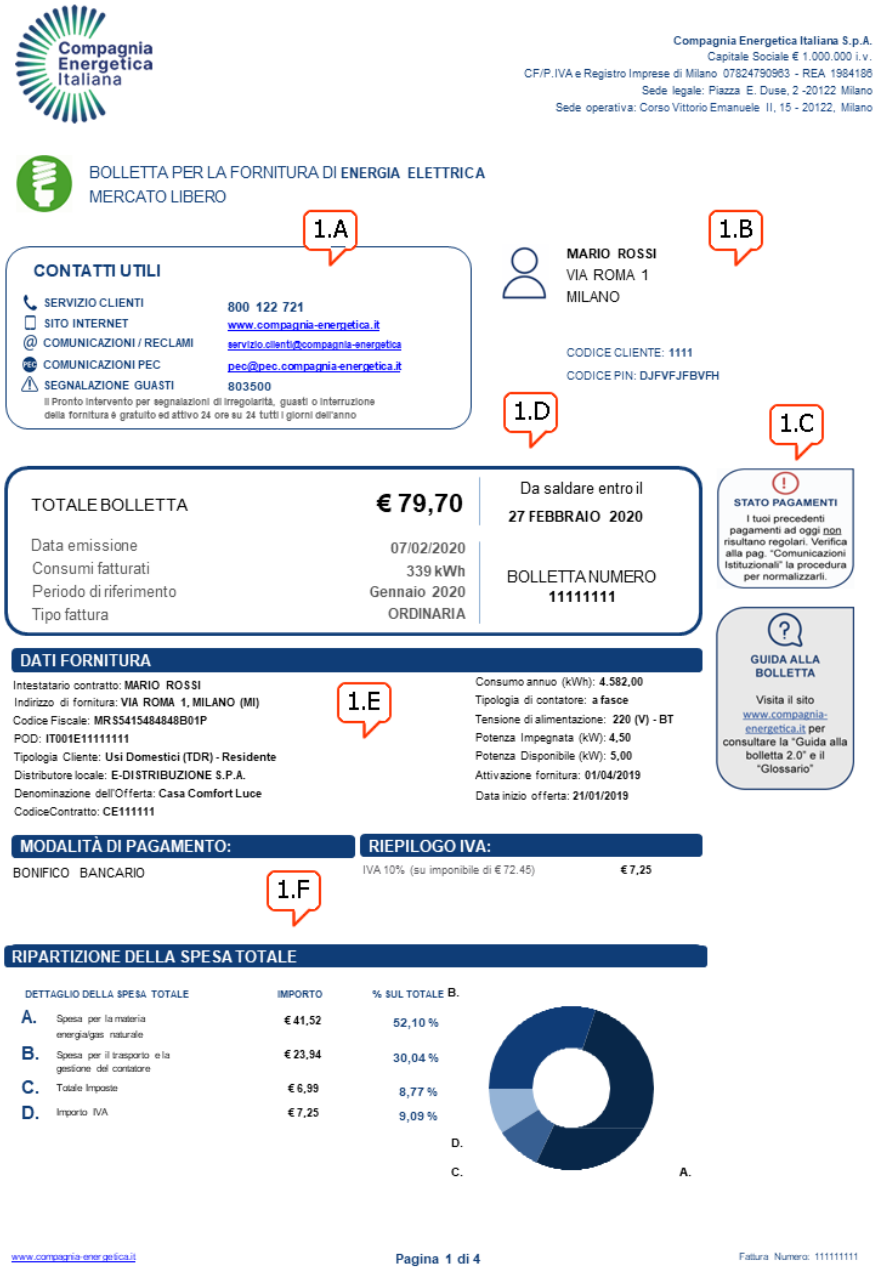

Figura 5.8: Fac-simile di una bolletta riempita con dati di fantasia

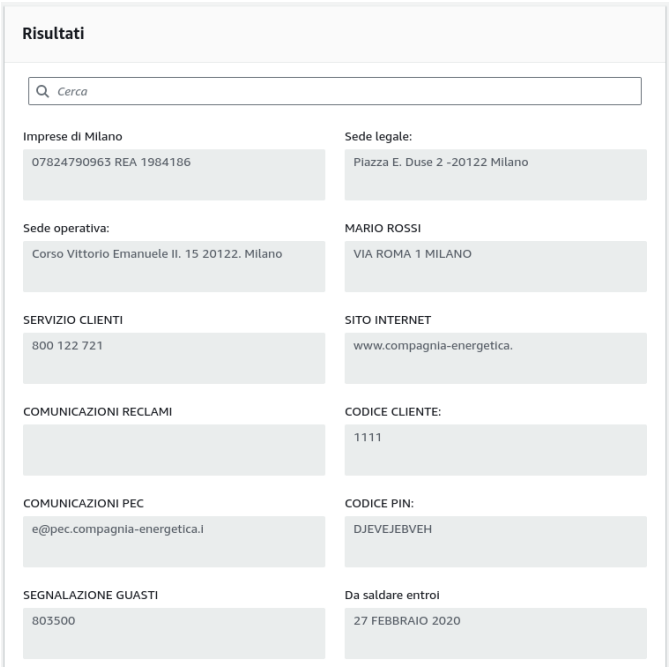

Figura 5.9: Risultati dell'analisi del fac-simile

Un'altra funzionalità interessante offerta da Textract è la rilevazione automatica di dati organizzati in una tabella. Nella Figura 5.10 ne viene mostrato un esempio, in cui vengono rappresentati i dati della ripartizione della spesa totale.

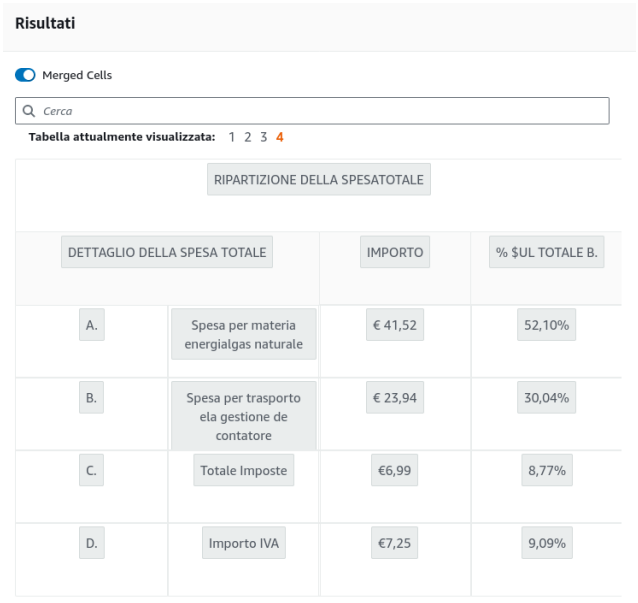

Figura 5.10: Esempio in cui vengono rilevati alcuni dati organizzati in una tabella.

5.2.3 Ricerda dei dati tramite query

In questo case study verrà ripreso il fac-simile di quello precedente e si mostrerà come i dati contenuti possano essere ricavati tramite delle query scritte in linguaggio naturale.

Nella Figura 5.11 vengono mostrati alcuni esempi di query con le risposte che hanno generato.

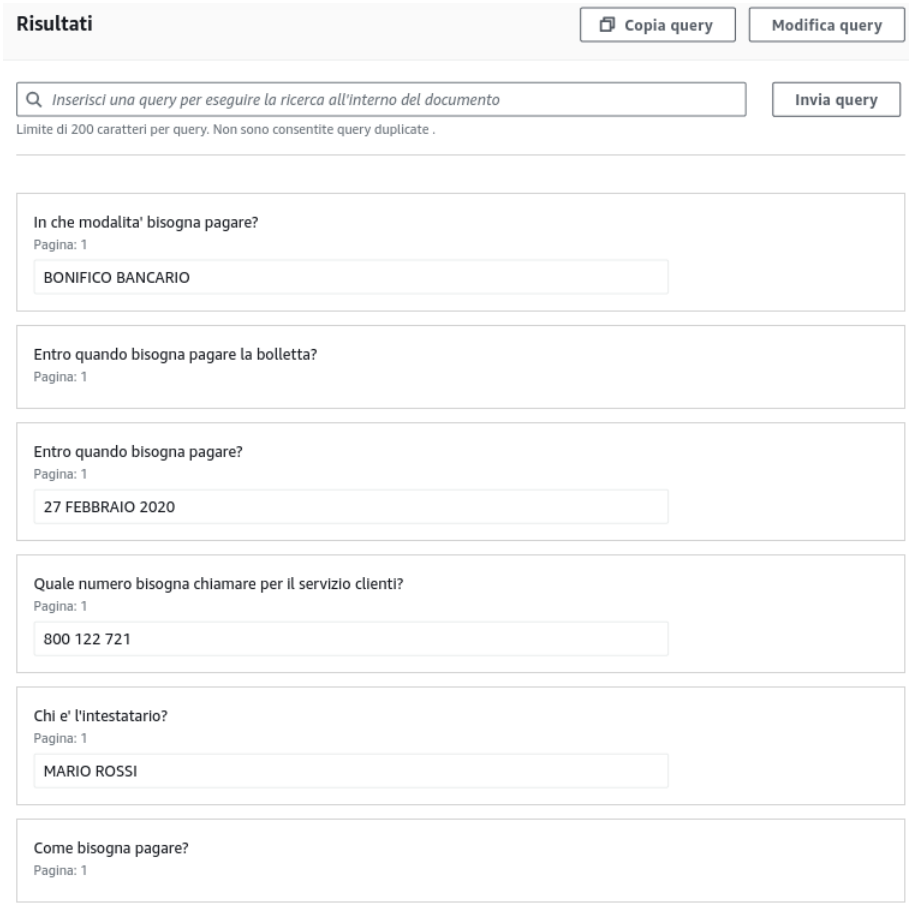

Figura 5.11: Alcune query eseguite per mostrare come i dati possano essere ricavati su richiesta

La cosa interessante di questo servizio è che le query possono essere scritte come delle domande vere e proprie; quindi si possono cercare le informazioni di interesse chiedendole in maniera naturale; in altre parole, non si è costretti a scrivere "Quanto vale campo x?". Tuttavia, questo tool è abbastanza rigido, in quanto non è in grado di elaborare e interpretare i campi; quindi, se si scrivono le query in maniera elaborata, non sempre la richiesta viene compresa, ed è necessario che ci siano alcune parole chiave all'interno di esse, altrimenti il servizio non è in grado di comprendere le richieste.

Per esempio, quando è stato chiesto quale fosse la modalità di pagamento, si è riscontrato che se la parola "modalità" non è presente nella richiesta, quest'ultima non viene riconosciuta e, quindi, non viene fornita nessuna risposta; una cosa analoga succede anche quando è stato chiesto quale fosse la data di scadenza di pagamento, dove è stato sufficiente accorciare la frase per ottenere il dato desiderato.

In conclusione, se si parte dal presupposto di avere già raccolto e organizzato tutti i dati presenti nel documento, questo servizio può essere molto utile per semplificare e automatizzare il processo di ricerca, ma, attualmente, necessita ancora di tanto miglioramento per poter elaborare correttamente le richieste e fornire i dati che si desiderano conoscere.

CAPITOLO 6

Discussione sulle esperienze condotte

In questo capitolo verranno esaminati i risultati dell'analisi dei case study condotta tramite l'IA fornita da Amazon Web Services. In particolare, si trarranno conclusioni in merito al riconoscimento di immagini, traduzione automatica ed estrazione di testi, analizzando i principali punti di forza e di debolezza.

6.1 Discussione sul riconoscimento di immagini

Il riconoscimento di immagini è una tecnologia ancora in evoluzione, ma sta già dimostrando di essere molto potente e di fare passi da gigante. Esso può essere utile in diversi ambiti; alcuni esempi riguardano l'automazione industriale, dove l'IA può essere impiegata per l'ispezione automatica dei prodotti, la gestione della qualità e la manutenzione predittiva delle macchine; oppure, riguardano la salute, dove l'IA potrebbe essere utilizzata per l'analisi delle immagini, consentendo una diagnosi più rapida e accurata; oppure, ancora, possono riguardare la sicurezza stradale, attraverso la rilevazione e la prevenzione degli incidenti. In particolare, il riconoscimento facciale può essere utilizzato per il controllo degli accessi a luoghi sensibili, come edifici governativi, aziende o istituzioni finanziarie, sostituendo o integrando i tradizionali sistemi di accesso basati su badge o password. Ancora, può essere utilizzato per la sicurezza pubblica, supportando le indagini e l'individuazione dei criminali ricercati; infine, può essere utilizzato come strumento di identificazione per l'autenticazione su dispositivi mobili e non, andando a sostituire o complementare l'uso delle password. Tutto questo sta diventando sempre più alla portata di tutti, grazie a servizi come quelli offerti da Amazon Web Services, che permettono di utilizzare software di Machine Learning senza dover necessariamente possedere le conoscenze e la potenza di calcolo necessarie per creare e addestrare un modello adeguato. Alla luce degli studi condotti in questa tesi, il servizio di riconoscimento di immagini di AWS, ovvero Amazon Rekognition, si è dimostrato molto potente nei casi di riconoscimento di volti celebri, di confronto facciale e di analisi dei video archiviati; per quanto non sempre le analisi condotte sono perfette, le difficoltà di riconoscimento incontrate nel terzo capitolo (come il mancato riconoscimento di Charlie Chaplin o di Chester Bennington), possono essere superate facendo più prove e incrociando i risultati, anche di tool diversi; si pensi, all'esperimento di Charlie Chaplin, che può essere riconosciuto anche da giovane se si utilizza anche il confronto facciale. Inoltre, questi servizi possono migliorare anche attraverso un maggiore addestramento, dando la possibilità alle reti neurali di avere più informazioni a disposizione per poter effettuare analisi quanto più accurate possibile.

6.2 Discussione sulla traduzione automatica

La traduzione automatica è un'attività che viene studiata e applicata da molto tempo, fin dagli anni'50. Prima dell'avvento dell'IA esistevano alcune tecniche; la prima, quella basata su regole, consisteva nello scomporre il testo in parole (o gruppi di poche parole) e poi convertirlo in un'altra lingua; questo approccio risultò molto rigido, in quanto non era in grado di adattarsi alle varie sfumature linguistiche, generando errori dovuti alla mancanza di comprensione del contesto. Dunque, questa tecnica fu superata da un'altra basata sulla statistica; quest'ultima, invece di utilizzare le regole, fa affidamento alla probabilità, analizzando testi già esistenti e, sulla base delle sequenze di parole più frequentemente trovate, traduce di conseguenza. Il Machine Learning, in generale, è capace di apprendere molto velocemente da grandi set di dati e, pertanto, si adatta molto bene a questo scopo. Dunque, anche l'approccio statistico viene superato, dando spazio alla Neural Machine Translation, che è in grado di migliorare sempre di più attraverso i dovuti addestramenti. Gli studi condotti tramite Amazon Translate e DeepL, hanno dimostrato quanto l'IA possa semplificare notevolmente il processo di traduzione, mantenendo in gran parte la coerenza a livello sintattico e semantico con il testo originale. Tuttavia, necessita ancora di un'attenta supervisione, in quanto gli errori che commette, anche se pochi, possono essere cruciali per comprendere correttamente il significato di un discorso; oltretutto, le prestazioni dell'IA tendono a peggiorare notevolmente se si analizzano testi con una struttura non convenzionale (come le poesie) oppure che contengono dei modi di dire tipici dei dialoghi. In conclusione, serve un'ulteriore evoluzione delle tecniche di NLP che attualmente conosciamo, in modo che le traduzioni che ci verranno fornite, anche in presenza di errori, ci possano comunque far comprendere pienamente il contesto in cui ci troviamo, senza essere costretti a sapere in anticipo quali sono le ambiguità della lingua originale che hanno generato tali errori.

6.3 Discussione sull'estrazione di testi

L'estrazione di testi è uno strumento molto utile per ricavare informazioni importanti da documenti cartacei non ancora registrati su nessun database; i benefici di questa tecnologia risiedono nella possibilità di poter immaganizzare un'enorme quantità di dati e avere la certezza di ritrovarli in un dispositivo o in un cloud. Se a questa tecnologia affianchiamo anche le tecniche di gestione dei Big Data, sarà possibile la ricerca di tali dati, organizzare dati che inizialmente non erano strutturati e fare confronti tra molti di essi, in modo tale da poter trovare corrispondenze ed effettuare previsioni. Tutto questo può essere fatto in maniera estremamente veloce, nonostante le dimensioni del dataset. Le esperienze fatte tramite l'utilizzo di Amazon Textract ci suggeriscono che l'IA è in grado di riconoscere testi scritti a mano, confondendo pochissime volte i caratteri (e anche quando succede, in genere, risulta abbastanza facile capire qual è il carattero giusto). Questo è senz'altro vero se si scrive in stampatello, ma lo stesso non si può dire se si scrive in corsivo, poiché, in genere, non si segue un font standardizzato, il che rende notevolmente più difficile il compito di estrazione dei testi, generando, nel migliore dei casi, risultati con più errori e, nel peggiore, risultati quasi del tutto incomprensibili. Detto ciò, una volta estratti i testi, Textract è in grado di riconoscere automaticamente la struttura del documento e riorganizza i dati raccolti di conseguenza. In questa operazione, esso si rivela molto efficace, in quanto è capace di evidenziare se sono presenti titoli, paragrafi, coppie chiave-valore, tabelle, indici, etc. Tuttavia, a volte, capita che Textract non riesca a leggere alcuni dati e, perciò, lascia i campi corrispondenti vuoti. Inoltre, esso è in grado di fornire i dati raccolti su richiesta, permettendo di ottenere i dati che servono senza doverli cercare manualmente all'interno del documento. Quest'ultimo tool necessita, comunque, di una certa attenzione, in quanto è necessario porre le domande in

maniera molto mirata, altrimenti potrebbe o non trovare il dato richiesto oppure trovarne uno errato. Se su piccola scala potrebbe sembrare una tecnologia limitata, se ci spostiamo su larga scala possiamo intuire come questi servizi permettano di relazionarci molto meglio con i dati che abbiamo a disposizione. Essi si potrebbero usare per verificare se le informazioni contenute in un certo documento sono veritiere, facendo il confronto con altri documenti già in possesso, oppure per verificare l'autenticità dei documenti stessi, o, ancora, per semplificare le operazioni di fact-checking, permettendo di migliorare l'accesso alle informazioni.

Conclusioni

In questa tesi si è analizzato, innanzitutto, che cosa vuol dire "Intelligenza Artificiale" e quali implicazioni comportava tale definizione, e si è imparato come, a partire dalla concezione di "intelligenza" che si adottava, ne risultava una tecnologia con caratteristiche e performance differenti. Se si cercava di fare in modo che la macchina, invece di ragionare in maniera simile agli esseri umani, ragionasse, piuttosto, seguendo una serie di principi e una logica formale entrambi universalmente validi, quello che si otteneva era la cosiddetta IA simbolica. Invece, se si cercava di fare in modo che la macchina apprendesse in maniera simile a come fanno gli esseri umani, quello che si otteneva era l'IA basata sulle reti neurali, le quali apprendono in maniera empirica, analizzando gli errori e cercando di minimizzarli nei tentativi successivi. Per ognuno di questi due tipi di IA esistono alcuni pro e alcuni contro, l'IA simbolica produce i suoi risultati secondo delle procedure già note a priori, ma è applicabile soltanto a contesti limitati ed è, quindi, poco flessibile. D'altra parte, le reti neurali hanno la capacità di adattarsi a situazioni nuove e, pertanto, di arricchire sempre di più le proprie conoscenze semplicemente continuando ad essere addestrate, ma le procedure con cui esse producono i risultati non possono essere conosciute.

Successivamente, ci siamo soffermati sulla piattaforma di cloud computing Amazon Web Services (AWS), sulla sua divisione in regioni e in zone di disponibilità e sui vantaggi che è in grado di offrire. Inoltre, abbiamo fornito una panoramica di tutti i servizi che offre, con particolare focus alle misure di sicurezza adottate per garantire la qualità di essi. Si è, poi, spiegato come AWS faccia ampio uso di reti neurali ricorrenti (RNN), specificando i vari modelli che si possono adottare, e, subito dopo, si sono descritti gli algoritmi di ML e DL impiegati.

Proprio grazie ad AWS, è stato possibile effettuare i case study trattati in questa tesi, permettendoci di analizzare le prestazioni dell'IA in ambito di riconoscimento di immagini, traduzione ed estrazione di testi, attraverso i servizi offerti, rispettivamente, da Rekognition, Translate, Textract. Le esperienze fatte durante questi studi ci suggeriscono nuove idee e prospettive in ambito di sicurezza, raccolta dati, identificazione e autenticazione, proiezioni statistiche e molto altro ancora. Infine, sono stati discussi i risultati ottenuti da queste esperienze, cercando di analizzare in maniera critica il comportamento dell'IA, tenendo conto sia dei punti di forza che di quelli di debolezza.

Dunque, l'IA è un campo di studi ancora in costante evoluzione, che necessita ancora di tanto addestramento per svolgere in maniera soddisfacente i nostri scopi; nonostante ciò, gli usi che se ne possono fare sono potenzialmente sconfinati e il potere che ha di automatizzare compiti che, altrimenti, sarebbero molto ripetitivi, noiosi e/o pericolosi, cambierà di sicuro la nostra società molto presto. In tale prospettiva, è importante rimanere consapevoli delle sfide

e delle questioni etiche che potrebbero emergere in futuro.

Bibliografia

BERTRAND RUSSELL, A. N. W. (1910), *Principia Mathematica*.

BISHOP, C. M. (2007), *Pattern Recognition and Machine Learning (Information Science and Statistics)*.

DALL'AGATA, M. (2021), «Neural Machine Translation e traduzione letteraria», .

IAN GOODFELLOW, A. C., YOSHUA BENGIO (2008), *Deep Learning*.

MARMO, R. (2021), *Algoritmi per l'intelligenza artificiale. Progettazione dell'algoritmo, dati e machine learning, neural network, deep learning*.

MCCARTHY, J. (1960), «Programs with common sense», Rap. tecn.

REZZANI, A. (2013), *Big data. Architettura, tecnologie e metodi per l'utilizzo di grandi basi di dati*, Svezia.

STUART RUSSELL, P. N. (2021), *Aritificial Intelligence: A Modern Approach*.

TURING, A. (1950), «Computing Machinery and Intelligence», .

Siti web consultati

- Amazon Web Services – `https://docs.aws.amazon.com`
- Ai4Business – `https://www.ai4business.it`
- Andrea Minini – `https://www.andreaminini.com`
- BNOVA – `https://www.bnova.it/`
- Enciclopedia Treccani – `https://www.treccani.it`
- RedHat – `http://www.redhat.com`
- DeepL – `https://www.deepl.com`

- Wikipedia – `www.wikipedia.org`
- IBM – `www.ibm.com`

Ringraziamenti

Innanzitutto, ringrazio la mia famiglia, che mi ha sostenuto durante tutto questo percorso, appoggiando le scelte che ho deciso di intraprendere, compresa quella di cambiare facoltà, quando ho avuto quella breve e, purtroppo, infruttuosa esperienza a Medicina. In particolare, ci terrei a ringraziare mia madre Habiba, che ha sempre dato tutta se stessa per farmi vedere il lato positivo delle cose, anche quando divento talmente pessimista da non vederci nulla, e mio padre Mohamed, che ha sempre lavorato sodo per assicurarmi il futuro che sto cercando piano piano di conquistare. Ringrazio, inoltre, mia sorella Sabrine, che mi ha dato preziosi consigli durante le mie prime esperienze universitarie e, in generale, rappresenta una fonte di ispirazione, e mio fratello Anis, con cui ho condiviso tanti bei momenti, che, a volte, mi hanno aiutato a smaltire lo stress che ho accumulato negli anni.

Vorrei ringraziare il Prof. Ursino, che mi ha aiutato e seguito tanto durante lo svolgimento di questa tesi e che è stato, in generale, un professore formidabile, che ha saputo trasmettermi parte del suo entusiasmo in questi studi ed è sempre stato disponibilissimo, sia per fornire chiarimenti che anche semplicemente per dare consigli.

Un altro importante ringraziamento va a due ragazzi che ritengo straordinari e che hanno sempre avuto la mia ammirazione, Francesco e Alessandro. Voi mi avete accompagnato in tutti i progetti e siete stati la migliore squadra che io abbia mai avuto, ho sempre apprezzato molto l'impegno che avete messo in tutto quello che fate, ma mantenendo comunque una certa leggerezza, che molte volte mi ha rassicurato. Vi auguro il meglio con il percorso che avete intrapreso con la magistrale, siete dei ragazzi molto in gamba e so che farete passi da gigante.

Ringrazio anche tutti i ragazzi con cui ho studiato in questi anni: Gabriel, Tosca, Federico, Luca, Lorenzo, Leonardo, Omar, Nicola, Michele, Giorgia, Giulia, Riccardo. Ad essere sincero, a me mancano le parole per poter esprimere quanto mi ritengo fortunato ad avervi conosciuti; non solo mi avete sempre aiutato con gli studi, ma insieme abbiamo condiviso tante di quelle esperienze (anche non legate all'università), che ritengo estremamente preziose e che non dimenticherò mai. Ci tengo a specificare che grazie a voi, quello che all'inizio credevo fosse un percorso che mi avrebbe portato un sacco di frustrazione, è diventato invece qualcosa di molto più bello, che mi ha permesso anche di crescere molto come persona. Inoltre, per quanto riguarda Gabriel e Tosca, sono veramente molto entusiasta di laurearmi con voi e sono molto orgoglioso del traguardo che avete appena raggiunto. Anche a voi, vi auguro un buon lavoro con il percorso che intraprenderete con la magistrale, sappiate che ho sempre creduto in voi e sempre ci crederò.

Vorrei infine dedicare i miei utlimi due ringraziamenti a quelli che sono da molti anni miei amici e su cui ho sempre potuto contare. Uno è dedicato a Matteo, che ringrazio per avermi

ascoltato in tutti questi anni e avermi fatto sentire molto spesso capito; l'altro è dedicato a Fernando, che ringrazio per avermi sempre aiutato a ridimensionare i problemi con cui, di volta in volta, avevo a che fare, che altrimenti avrei creduto molto spesso che sarebbero stati troppo grandi per poterli affrontare.

Ringrazio anche tutte le persone che mi hanno fatto compagnia durante il percorso universitario, anche se solo per poco tempo o per brevi momenti, vi ringrazio tutti.