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



TESI DI LAUREA

Progettazione e implementazione di case study nel contesto dell'Intelligenza Artificiale Discriminativa

Design and implementation of case studies in the context of Discriminative Artificial Intelligence

Relatore

Prof. Domenico Ursino

Candidata

Naomi De Michele

ANNO ACCADEMICO 2025-2026

The only way of discovering the limits of the possible
is to venture a little way past them into the impossible.
Arthur C. Clarke, *Profiles of the Future*

Sommario

Nella presente tesi si ha l'obiettivo di fornire una visione concreta di alcuni strumenti di Intelligenza Artificiale discriminativa, partendo da un breve inquadramento teorico dell'AI, della sua evoluzione e dei suoi principali paradigmi di apprendimento. Il lavoro si concentra sull'attività svolta durante il periodo di tirocinio, nel quale sono stati selezionati e sperimentati tre servizi di Intelligenza Artificiale offerti da Amazon Web Services (AWS): Content Moderation, Amazon Rekognition e Amazon SageMaker. La Content Moderation è stata prima sperimentata tramite AWS e successivamente analizzata in relazione al controllo delle immagini pubblicate sui Social Media. Attraverso Amazon Rekognition è stato sperimentato il servizio di Facial Analysis, successivamente collegato al suo impiego nell'ambito delle indagini di polizia. Infine, Amazon SageMaker è stato utilizzato per analizzare testi poetici attraverso Sentiment Analysis ed Emotion Classification. Per ciascun servizio, l'analisi sperimentale svolta attraverso AWS è stata messa in relazione con applicazioni e casi reali. Il confronto ha permesso di evidenziare le differenze tra quanto osservato durante le sperimentazioni e le problematiche che caratterizzano l'impiego di queste tecnologie in contesti reali.

Keyword: Intelligenza Artificiale (AI), Machine Learning (ML), Amazon Web Services (AWS), Content Moderation, Social Media, Facial Analysis, Sentiment Analysis, Emotion Classification

Introduzione

Negli ultimi anni, l'Intelligenza Artificiale ha assunto un ruolo sempre più rilevante nella vita di tutti i giorni, grazie alla creazione di sistemi in grado di analizzare grandi quantità di dati, riconoscere immagini o testi e supportare attività che in passato richiedevano l'intervento umano. La crescente diffusione di queste tecnologie ha fatto sì che esse venissero utilizzate in diversi contesti, dai Social Media alla sicurezza, fino all'analisi del linguaggio e dei contenuti testuali. Le possibilità offerte dall'Intelligenza Artificiale hanno portato alla nascita di un forte interesse nei confronti di queste tecnologie e, allo stesso tempo, alla necessità di comprenderne concretamente il funzionamento e le problematiche che possono emergere quando tali sistemi vengono utilizzati in contesti reali. Da questa considerazione prende forma il presente lavoro, che si propone di affiancare a un iniziale inquadramento teorico dell'Intelligenza Artificiale un'analisi maggiormente applicativa.

Il lavoro si concentra principalmente sulle attività svolte durante il periodo di tirocinio, nel corso del quale sono stati selezionati e sperimentati alcuni strumenti di Intelligenza Artificiale messi a disposizione da Amazon Web Services (AWS). In particolare, è stata prestata attenzione alla Content Moderation, ad Amazon Rekognition e ad Amazon SageMaker, scelti per analizzare differenti applicazioni dell'Intelligenza Artificiale. Le sperimentazioni svolte tramite AWS sono state successivamente messe in relazione con applicazioni e casi reali, con l'obiettivo di confrontare quanto osservato durante le prove con le problematiche che possono caratterizzare l'uso concreto di queste tecnologie.

L'analisi riguardante la moderazione dei contenuti è stata svolta, in un primo momento, su immagini generiche e, successivamente, su immagini studiate seguendo le linee guida di Social Media reali, prendendo in considerazione TikTok e Instagram. Lo studio relativo al riconoscimento facciale è stato articolato in più fasi: inizialmente è stata utilizzata la Facial Comparison, confrontando due immagini di quella che si suppone essere la stessa persona; successivamente è stata utilizzata la Facial Analysis sul singolo volto, applicandola a un caso più complesso. Quanto osservato è stato, poi, messo in relazione con l'impiego di questa tecnologia nell'ambito delle indagini di polizia. Infine, è stata svolta un'analisi di testi poetici, inizialmente attraverso una classificazione binaria, con il riconoscimento dell'orientamento del testo come POSITIVE o NEGATIVE, e successivamente attraverso l'Emotion Classification, in grado di individuare emozioni più specifiche.

Lo scopo del lavoro è quello di fornire una visione dell'Intelligenza Artificiale che non si limiti alla descrizione teorica degli strumenti applicati, ma che consideri anche il loro funzionamento attraverso la sperimentazione e il confronto con situazioni reali.

Questa tesi è strutturata come di seguito specificato:

- Nel primo capitolo viene presentato il concetto di Intelligenza Artificiale, ripercorren-

done l'evoluzione storica e i principali fondamenti teorici. Vengono descritti il Machine Learning, il Deep Learning, i principali paradigmi di apprendimento e il ruolo dei Big Data. Viene, poi, approfondita la distinzione tra Intelligenza Artificiale discriminativa e generativa.

- Nel secondo capitolo viene presentata la piattaforma Amazon Web Services (AWS), introducendo i principi del cloud computing e i principali modelli di servizio e distribuzione. Successivamente vengono descritte l'infrastruttura e l'architettura di AWS, le modalità di accesso e i servizi principali.

- Nel terzo capitolo viene presentato un case study relativo alla moderazione automatica dei contenuti attraverso l'utilizzo di Amazon Rekognition, analizzando immagini caratterizzate da differenti livelli di rischio. Vengono, poi, approfondite le politiche di Content Moderation di Meta e TikTok.

- Nel quarto capitolo viene presentato un case study relativo al riconoscimento facciale mediante l'utilizzo di Amazon Rekognition. La sperimentazione viene applicata a un contesto investigativo, con l'obiettivo di valutare le capacità del sistema e le eventuali criticità.

- Nel quinto capitolo viene presentato un case study relativo all'analisi dei sentimenti mediante Amazon SageMaker, applicato a testi letterari in lingua inglese. La sperimentazione confronta un modello di Sentiment Analysis binaria con un modello di Emotion Classification multi-classe.

- Nel sesto capitolo viene proposta una discussione dei case study precedentemente analizzati, approfondendo le possibili conseguenze degli errori dei sistemi di Intelligenza Artificiale in contesti reali.

Indice

Introduzione **ii**

1 Introduzione all'Intelligenza Artificiale **1**
1.1 Definizione di Intelligenza Artificiale e principi fondamentali 1
1.1.1 Il test di Turing: criterio di valutazione dell'Intelligenza Artificiale . . 2
1.1.2 Il modello cognitivo umano applicato all'Intelligenza Artificiale . . . 3
1.1.3 Approccio razionale allo studio dell'Intelligenza Artificiale 4
1.1.4 Approccio basato sull'agente razionale nell' Intelligenza Artificiale . . 4
1.2 Evoluzione storica e sviluppo dell'Intelligenza Artificiale 5
1.2.1 Dalle origini al "primo inverno" dell'Intelligenza Artificiale (1943–1979) 5
1.2.2 L'affermazione dei sistemi esperti (1980–1987) 7
1.2.3 Il ritorno delle reti neurali (1987–2000) 8
1.2.4 L'affermazione del Machine Learning (anni '90-oggi) 9
1.2.5 I Big Data (2001–oggi) 9
1.3 Machine Learning come fondamento dell'AI moderna 9
1.3.1 Definizione di Machine Learning 10
1.3.2 Paradigmi di apprendimento 10
1.3.3 L'evoluzione del Deep Learning 12
1.4 Modelli discriminativi e generativi 12
1.4.1 Modelli discriminativi 12
1.4.2 Modelli generativi 13
1.4.3 Confronto fra Intelligenza Artificiale Discriminativa e Generativa . . . 13
1.5 Applicazioni dell'Intelligenza Artificiale discriminativa oggetto dei case study 14
1.5.1 Moderazione dei contenuti 15
1.5.2 Riconoscimento visivo 16
1.5.3 Analisi dei sentimenti 16

2 Amazon Web Services **18**
2.1 Introduzione ad Amazon Web Services 18
2.1.1 Fondamenti del cloud computing: IaaS, PaaS e SaaS 19
2.1.2 AWS nel panorama degli hyperscaler 21
2.1.3 Storia ed evoluzione di AWS 21
2.2 Architettura e infrastruttura di AWS 22
2.2.1 Infrastruttura globale 23
2.2.2 Architettura tecnica dei servizi 24

2.2.3 Accesso e interfaccia utente . . . 26
2.3 Principali servizi e tool di AWS . . . 26
2.3.1 Servizi di storage . . . 26
2.3.2 Servizi di calcolo . . . 28
2.3.3 Servizi di Intelligenza Artificiale e Machine Learning . . . 29

3 Un case study relativo alla moderazione dei contenuti 33
3.1 Obiettivi e progettazione del case study . . . 33
3.2 Sperimentazione con Amazon Rekognition . . . 34
3.2.1 Analisi di un contenuto Safe . . . 34
3.2.2 Analisi di un contenuto Unsafe . . . 36
3.2.3 Analisi di un contenuto Borderline . . . 38
3.3 Content Moderation nelle piattaforme social . . . 39
3.3.1 Politiche di moderazione di Meta . . . 40
3.3.2 Politiche di moderazione di TikTok . . . 40
3.4 Confronto sperimentale tra Amazon Rekognition, Meta e TikTok . . . 42
3.4.1 Contenuti caratterizzati da rischio implicito . . . 42
3.4.2 Contenuti sensibili e contestualizzati . . . 46
3.4.3 Analisi dei risultati . . . 47

4 Un case study relativo al riconoscimento visivo 50
4.1 Facial Analysis . . . 50
4.1.1 La Facial Analysis in ambito investigativo . . . 51
4.1.2 Amazon Rekognition e Facial Comparison . . . 51
4.2 Obiettivo e progettazione del case study . . . 52
4.3 Sperimentazione con Amazon Rekognition . . . 53
4.3.1 Aging Case . . . 53
4.3.2 Basic Case . . . 54
4.3.3 Twins Case . . . 57
4.3.4 Facial Analysis dei soggetti del Twins Case . . . 58
4.3.5 Analisi dei risultati . . . 64

5 Un case study relativo all'analisi dei sentimenti 67
5.1 Amazon SageMaker e Natural Language Processing . . . 67
5.2 Sentiment Analysis ed Emotion Classification . . . 67
5.3 Obiettivo e progettazione del case study . . . 68
5.4 Implementazione e sperimentazione . . . 68
5.4.1 Emily Dickinson: The Battle-Field . . . 69
5.4.2 Luis Sepúlveda: The Story of a Seagull and the Cat Who Taught Her to Fly . . . 71
5.4.3 Edgar Allan Poe: Annabel Lee . . . 74
5.4.4 William Wordsworth: My Heart Leaps Up . . . 76
5.4.5 Analisi dei risultati . . . 78

6 Discussione 80
6.1 Discussione sul Case Study 1: i limiti nella Content Moderation nei Social Media 80
6.1.1 Il caso Meta . . . 81
6.1.2 Il caso TikTok . . . 82
6.1.3 Conclusioni . . . 83
6.2 Discussione sul Case Study 2: riconoscimento facciale nelle indagini . . . 83
6.2.1 Problematiche nella tutela della privacy . . . 85

6.2.2 Conclusioni 86
6.3 Discussione sul Case Study 3: sentiment analysis nei testi letterari 86
6.3.1 I limiti della Sentiment Analysis 87
6.3.2 Riconoscere, comprendere o provare emozioni? 88
6.3.3 Conclusioni 89

Conclusioni **90**

Bibliografia **92**

Sitografia **94**

Ringraziamenti **97**

Elenco delle figure

1.1 Rappresentazione del Test di Turing (fonte: ResearchGate). 3
1.2 Storia dell'Intelligenza Artificiale (fonte: Noovle). 5
1.3 ELIZA (fonte: NeonTools). 6
1.4 Reti neurali (fonte: ZeroUno). 8
1.5 Le gerarchie dell'Intelligenza Artificiale (fonte: IONOS Digital Guide). 10
1.6 Paradigmi di apprendimento (fonte: Deep Learning Italia su Medium). 11
1.7 Applicazioni dell'Intelligenza Artificiale Generativa (fonte: Cittadinanza Digitale). 14
1.8 Confronto tra modello discriminativo e modello generativo (fonte: Wikipedia). 14
1.9 Moderazione dei contenuti (fonte: Copyleaks). 15
1.10 Riconoscimento facciale (fonte: AI4Business). 16

2.1 Schema AWS (fonte: Maceforce). 19
2.2 Schema differenziazione hyperscaler e data center tradizionali (fonte: Solarwinds). 21
2.3 Schema crescita AWS (fonte: Kinsta). 22
2.4 Infrastruttura globale (fonte: Amazon). 23
2.5 Fibre ottiche (fonte: Linkedin). 24
2.6 Microservizi (fonte: Amazon). 25
2.7 Alcuni tool e servizi AWS (fonte: Zerounoweb). 27
2.8 Servizi di storage di AWS (fonte: Amazon). 28
2.9 Servizi di calcolo AWS (fonte: Amazon). 30
2.10 Servizi di AI AWS (fonte: Itmagic). 30

3.1 Tramonto (fonte: Canva). 35
3.2 Risultato della prima sperimentazione (Svolta su: Amazon Web Services) . . 35
3.3 Militari (fonte: Rid). 36
3.4 Prima parte del risultato della seconda sperimentazione (Svolta su: Amazon Web Services) 36
3.5 Seconda parte del risultato della seconda sperimentazione (Svolta su: Amazon Web Services) 37
3.6 Squalo (fonte: Veratour). 38
3.7 Risultato della terza sperimentazione (Svolta su: Amazon Web Services) . . . 39
3.8 Parkour (fonte: Kaique Rocha). 43

3.9 Risultato della prima sperimentazione di confronto (Svolta su: Amazon Web Services) . . . 43
3.10 Urbex (fonte: Cottonbro Studio). . . . 44
3.11 Risultato della seconda sperimentazione di confronto (Svolta su: Amazon Web Services) . . . 45
3.12 Immagine di Softair rielaborata da (fonte: Pixabay). . . . 46
3.13 Prima parte del risultato della sperimentazione contestualizzata (Svolta su: Amazon Web Services) . . . 47
3.14 Seconda parte del risultato della contestualizzata (Svolta su: Amazon Web Services) . . . 48

4.1 Rappresentazione della pipeline di Fiacial Analysis (fonte: Semanticscholar). 52
4.2 Aging Case: confronto tra immagini dello stesso soggetto in età differenti mediante Amazon Rekognition (immagini fornite da Amazon Rekognition). . 53
4.3 Aging Case: prima parte del risultato. . . . 54
4.4 Aging Case: seconda parte del risultato. . . . 55
4.5 Aging Case: terza parte del risultato. . . . 55
4.6 Aging Case: quarta parte del risultato. . . . 56
4.7 Basic Case (fonte: Pexels . . . 56
4.8 Basic Case: prima parte del risultato. . . . 56
4.9 Basic Case: seconda parte del risultato. . . . 57
4.10 Basic Case: terza parte del risultato. . . . 58
4.11 Basic Case: quarta parte del risultato. . . . 59
4.12 Twins Case (fonte: Pexels) . . . 59
4.13 Twins Case: prima parte del risultato. . . . 59
4.14 Twins Case: seconda parte del risultato. . . . 60
4.15 Twins Case: terza parte del risultato. . . . 61
4.16 Twins Case: quarta parte del risultato. . . . 62
4.17 Analisi facciale della prima gemella . . . 62
4.18 Richiesta relativa alla prima gemella. . . . 63
4.19 Risposta relativa alla prima gemella: prima parte . . . 63
4.20 Risposta relativa alla prima gemella: seconda parte . . . 64
4.21 Analisi facciale della seconda gemella . . . 65
4.22 Richiesta relativa alla seconda gemella. . . . 65
4.23 Risposta seconda gemella. . . . 66

5.1 Estratto della poesia *The Battle-Field* di Emily Dickinson (Screen preso da Emilydickinson). . . . 69
5.2 Sentiment analysis binaria *The Battle-Field* di Emily Dickinson. . . . 70
5.3 Emotion classification multi-class*The Battle-Field* di Emily Dickinson. . . . 70
5.4 Estratto del romanzo *The Story of a Seagull and the Cat Who Taught Her to Fly* di Luis Sepúlveda . . . 72
5.5 Sentiment analysis binaria *The Story of a Seagull and the Cat Who Taught Her to Fly* di Luis Sepúlveda. . . . 72
5.6 Emotion classification multi-classe *The Story of a Seagull and the Cat Who Taught Her to Fly* di Luis Sepúlveda. . . . 73
5.7 Estratto della poesia *Annabel Lee* di Edgar Allan Poe (Testo preso da Poetry-foundation. Traduzione presa da Oubliettemagazine). . . . 74
5.8 Sentiment analysis binaria *Annabel Lee* di Edgar Allan Poe. . . . 75
5.9 Emotion classification multi-classe *Annabel Lee* di Edgar Allan Poe. . . . 75

5.10 Estratto della poesia *Annabel Lee* di Edgar Allan Poe (Testo preso da Poets. Traduzione presa da Isoladellapoesia). 77
5.11 Sentiment analysis binaria *My Heart Leaps Up* di William Wordsworth. 77
5.12 Emotion classification multi-classe *My Heart Leaps Up* di William Wordsworth. 78

6.1 Esempio di Warning Label su TikTok. (fonte: TikTok) 84
6.2 Esempio del processo di Sentiment Analysis. (fonte: MathWorks) 87

Elenco delle tabelle

1.1 Confronto tra modelli generativi e modelli discriminativi (fonte: elaborazione da HumAI). 15

5.1 Confronto tra la sentiment analysis binaria e l'emotion classification multi-classe di The Battle-Field 71
5.2 Confronto tra la sentiment analysis binaria e l'emotion classification multi-classe The Story of a Seagull and the Cat Who Taught Her to Fly 74
5.3 Confronto tra la sentiment analysis binaria e l'emotion classification multi-classe Annabel Lee 76
5.4 Confronto tra la sentiment analysis binaria e l'emotion classification multi-classe My Heart Leaps Up 79

CAPITOLO 1

Introduzione all'Intelligenza Artificiale

Questo capitolo introduttivo presenta i principali concetti dell'Intelligenza Artificiale, illustrandone la definizione e l'evoluzione storica. Vengono, inoltre, descritti i fondamenti teorici che hanno contribuito allo sviluppo di questa disciplina, dai primi studi fino alla diffusione delle moderne tecniche di apprendimento automatico.
Successivamente vengono analizzati il Machine Learning, il Deep Learning e i principali paradigmi di apprendimento, evidenziandone il ruolo nello sviluppo dell'AI moderna e l'importanza dei Big Data nell'addestramento dei modelli intelligenti.
L'ultima parte del capitolo è dedicata al confronto tra Intelligenza Artificiale discriminativa e Intelligenza Artificiale generativa, descrivendone le caratteristiche, le principali differenze e gli ambiti applicativi. Infine, vengono introdotte le applicazioni dell'AI oggetto dei case study sviluppati nei capitoli successivi.

1.1 Definizione di Intelligenza Artificiale e principi fondamentali

L'Intelligenza Artificiale (AI) è una branca dell'informatica che nasce dal desiderio di emulare, tramite macchina, la mente umana. Numerose discipline hanno portato alla formazione dell'AI tra le quali: psicologia, filosofia, scienze cognitive, matematica e cibernetica. L'uomo definisce l'*intelligenza* come la capacità di pensare, comprendere, ragionare e provare sentimenti, ovvero l'insieme delle caratteristiche che lo distinguono da ogni altro essere vivente. Sebbene al giorno d'oggi l'Intelligenza Artificiale venga associata comunemente a computer, algoritmi e allo sviluppo di tecnologie in grado di supportare o semplificare numerose attività umane, in realtà le sue radici sono molto precedenti alla sua affermazione come disciplina scientifica avvenuta nel XX secolo. Fin dall'antichità filosofi, matematici e scienziati cercarono di comprendere la natura dell'intelligenza e del ragionamento umano, studiando metodi per descriverli attraverso l'uso di regole precise.

Nella filosofia Aristotelica il pensiero è *divino*: è il pensiero stesso a fare l'uomo e a differirlo dagli altri esseri viventi. Egli ideò il primo tipo di ragionamento formale, ovvero quel ragionamento tramite il quale le conclusioni dipendono dalle ipotesi iniziali.

Negli anni molti si posero il dubbio se il ragionamento formale potesse essere trasmesso a una macchina, creando così l'idea di "*macchine pensanti*". Raimondo Lullo, filosofo del 1300, fu uno dei precursori dell'Intelligenza Artificiale. Nella sua opera "*Ars Magna*" immaginava una macchina in grado di svolgere calcoli e ragionamenti, creando l'idea che il ragionamento potesse essere automatizzato. Con Leibniz si è provato a trasformare il ragionamento in

calcolo tramite regole che avrebbero consentito la creazione di un linguaggio universale in grado di rappresentare qualsiasi concetto, riducendo i pensieri umani in simboli elementari e ponendo le basi per l'idea che il pensiero potesse essere elaborato tramite macchina.

Queste teorie portarono alle sperimentazioni di Charles Babbage, che cercò di inventare la macchina analitica, in grado non solo di fare calcoli, ma anche di svolgere programmi. Precursore dei computer moderni, questa conteneva un "*mulino*" (CPU) per svolgere calcoli, il "*magazzino*" (memoria) che conservava i risultati, e un *lettore* costituito da schede perforate tramite il quale era possibile leggere istruzioni. Vennero, quindi, gettate le basi dell'era digitale e l'idea che una macchina potesse essere istruita e programmabile. Nel 1843 Lady Ada Lovelace scrisse quello che ancora oggi viene considerato il primo programma informatico della storia, diventando la prima programmatrice. Mentre traduceva una lezione di Babbage svoltasi a Torino, annotò alcune osservazioni, fra le quali la "Nota G", nella quale scrisse un algoritmo per calcolare i numeri di Bernoulli, il primo della storia dell'informatica. La sua più grande intuizione fu che le macchine potessero manipolare altri simboli oltre ai valori numerici ed essere usate per aiutare l'uomo a comporre musica, scrivere testi e creare opere, mettendo le basi che porteranno alla nascita delle AI attuali. Di particolare rilievo per l'informatica moderna è "l'obiezione di Lady Lovelace", una nota nella quale affermava che le macchine non potessero pensare o creare nulla di nuovo in modo autonomo ma solo eseguire le istruzioni date ad esse da un essere umano, ponendosi già allora una delle questioni etiche che ancora oggi alimentano il dibattito sull'Intelligenza Artificiale. Queste teorie, però, rimasero tali fino al 1936 quando Alan Turing, il padre dell'informatica, teorizzò la Macchina di Turing.

1.1.1 Il test di Turing: criterio di valutazione dell'Intelligenza Artificiale

A Turing si deve il passaggio da computer con comportamento di macchine meccaniche, in grado unicamente di svolgere calcoli e non programmabili, al concetto di macchina universale, ossia una macchina programmabile in grado di eseguire algoritmi che non fanno parte della macchina stessa, introducendo una prima separazione tra l'hardware e il software di un computer. Turing fu uno dei primi a cercare di rispondere alla domanda "una macchina può pensare?" tramite il famoso "Test di Turing".

Turing propone questo test sotto il nome di "*The imitation game*". Il test si comporta come un gioco a tre partecipanti: un giudice, il cui ruolo verrà assegnato all'umano, un uomo e una macchina, posizionati in tre stanze diverse senza la possibilità di vedersi a vicenda. La conclusione è semplice: una macchina "pensa" se durante una conversazione scritta l'osservatore non riesce a riconoscere chi sia la macchina e chi l'essere umano.

Questo test serve a definire l'intelligenza di una macchina: se il giudice, che comunica con l'essere umano e con la macchina, non riesce a riconoscere quest'ultima, allora essa viene definita "*intelligente*".

La Figura 1.1 mostra graficamente il principio alla base del Test di Turing.

Per superare il test è necessario che il computer soddisfi una serie di condizioni:

- *elaborazione del linguaggio naturale*, per la generazione di risposte comprensibili;
- *rappresentazione della conoscenza*, per la memorizzazione di informazioni;
- *ragionamento automatico*, per dare risposte coerenti;
- *apprendimento automatico*, per adattarsi a nuove situazioni.

Queste capacità sono state successivamente ampliate nel cosiddetto *Test di Turing totale*, al quale vengono aggiunte ulteriori competenze che consentono l'interazione con oggetti e persone, quali:

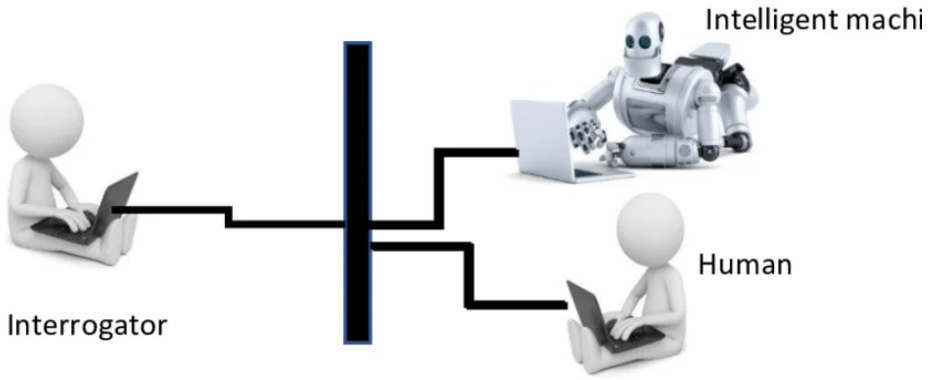


Figura 1.1: Rappresentazione del Test di Turing (fonte: ResearchGate).

- *visione artificiale e riconoscimento vocale*, per avere percezioni visive e uditive nell'ambiente circostante;
- *robotica*, per le interazioni con l'ambiente.

Definendo i pilastri dell'Intelligenza Artificiale e gli obiettivi che ogni macchina deve raggiungere. Turing fece una previsione: nel giro di 50 anni sarebbero esistiti computer così bravi nell'imitation game che il giudice avrebbe avuto solo una probabilità del 30% di vincere il gioco.

Tuttavia, nel corso degli anni, questo test è stato oggetto di numerose critiche e sono stati evidenziati i suoi limiti come criterio per valutare l'intelligenza di una macchina. Per questo motivo, oggi non è più considerato un metodo di studio sufficiente per riconoscere le capacità dell'Intelligenza Artificiale.

1.1.2 Il modello cognitivo umano applicato all'Intelligenza Artificiale

Il modello cognitivo si basa sull'idea che la mente umana possa essere descritta come un *elaboratore di informazioni*. Secondo questo approccio, l'essere umano riceve stimoli dall'ambiente esterno, li elabora attraverso le percezioni, o *processi cognitivi*, e produce una risposta o azione. L'obiettivo dell'Intelligenza Artificiale è comprendere tali meccanismi e riprodurli tramite modelli computazionali, così da sviluppare dei sistemi in grado di emulare il comportamento umano. Per raggiungere questo obiettivo è necessario comprendere il funzionamento della mente umana. A tal fine, il modello cognitivo si basa sia sull'*introspezione*, che rappresenta l'osservazione dei propri processi mentali, sia sugli *esperimenti condotti nell'ambito della psicologia cognitiva*, che consentono di analizzare il modo in cui gli esseri umani percepiscono, apprendono e ragionano.

Il modello cognitivo si sviluppò nello stesso periodo in cui Turing proponeva il suo test, grazie anche ai progressi della psicologia cognitiva e delle scienze cognitive. Diversi studi nell'ambito della psicologia cognitiva descrivono la mente come un sistema di elaborazione delle informazioni che opera in modo analogo a un computer, ricevendo input dall'ambiente e, successivamente all'elaborazione, producendo diversi output. Tali osservazioni sul comportamento umano hanno posto le basi per la realizzazione di modelli computazionali in grado di riprodurre alcuni processi cognitivi umani.

Le principali capacità cognitive legate al processo di elaborazione delle informazioni che caratterizzano l'essere umano sono:

- *memoria*; consente di acquisire, conservare e recuperare informazioni;

- *attenzione*; consente di rilevare le informazioni e riconoscere i pericoli;
- *percezione*; consente di ricevere e interpretare gli stimoli provenienti dall'ambiente;
- *linguaggio*; consente di comunicare e comprendere le informazioni;
- *funzioni esecutive*; consentono di pianificare, prendere decisioni e controllare il comportamento;
- *orientamento*; consente di muoversi nello spazio;
- *ragionamento*; consente di analizzare le situazioni e risolvere problemi;
- *motivazione*; nasce dal desiderio di soddisfare un bisogno o un obiettivo.

Lo studio dei processi cognitivi ha fornito importanti basi teoriche per la progettazione di modelli computazionali ispirati al funzionamento della mente umana. Tali modelli hanno contribuito allo sviluppo di sistemi capaci di apprendere, ragionare e prendere decisioni, pur senza riprodurre fedelmente il funzionamento biologico del cervello.

1.1.3 Approccio razionale allo studio dell'Intelligenza Artificiale

L'approccio razionale si basa sulle cosiddette "leggi del pensiero" (*laws of thought*), secondo cui il ragionamento può essere descritto attraverso regole logiche. Tali regole derivano dal pensiero Aristotelico, il quale si sviluppò attraverso l'uso di sillogismi, ovvero schemi di ragionamento capaci di condurre a conclusioni corrette partendo da premesse corrette, ponendo le basi della logica.

Lo sviluppo della logica formale rese possibile rappresentare problemi e relazioni mediante notazioni precise, favorendo, così, la realizzazione di programmi in grado di risolvere problemi espressi in termini logici.

Questo approccio presenta alcuni limiti. Il principale è rappresentato dal fatto che il pensiero umano è spesso caratterizzato da informazioni incomplete, incerte o contraddittorie, difficili da rappresentare mediante regole logiche. Inoltre, la risoluzione di problemi complessi può richiedere un elevato impiego di risorse computazionali.

1.1.4 Approccio basato sull'agente razionale nell' Intelligenza Artificiale

L'approccio basato sull'agente razionale considera l'Intelligenza Artificiale come lo studio e la progettazione di sistemi in grado di percepire l'ambiente e agire in modo da raggiungere i propri obiettivi. In questo approccio, un agente è un'entità capace di percepire ciò che lo circonda e di scegliere l'azione più adatta in base alle informazioni disponibili.

A differenza dell'approccio razionale basato sulle *leggi del pensiero*, l'obiettivo non è soltanto formulare inferenze logiche corrette, ma anche quello di individuare il comportamento più efficace per raggiungere un determinato obiettivo. In alcune situazioni, infatti, non esiste un'unica soluzione corretta; pertanto, è necessario prendere una decisione anche in presenza di informazioni incomplete o incerte.

Per questo motivo un agente razionale deve possedere diverse capacità, come la rappresentazione della conoscenza, il ragionamento, l'apprendimento, la percezione e la comprensione del linguaggio naturale. Tuttavia, in ambienti complessi non è sempre possibile raggiungere una razionalità perfetta, poiché le risorse computazionali e le informazioni disponibili possono essere limitate. Per questa ragione si parla di *razionalità limitata*, secondo cui un agente deve scegliere la migliore azione possibile nelle condizioni in cui opera.

1.2 Evoluzione storica e sviluppo dell'Intelligenza Artificiale

L'Intelligenza Artificiale odierna è il risultato di un lungo percorso di evoluzione scientifica e tecnologica, partendo dalle prime riflessioni teoriche sulla possibilità di costruire macchine intelligenti fino agli attuali sistemi basati su reti neurali e apprendimento automatico. La disciplina ha attraversato fasi di sviluppo, periodi di crisi e importanti progressi che ne hanno determinato l'affermazione.

La Figura 1.2 mostra una rappresentazione grafica della storia dell'Intelligenza Artificiale.



Figura 1.2: Storia dell'Intelligenza Artificiale (fonte: Noovle).

1.2.1 Dalle origini al "primo inverno" dell'Intelligenza Artificiale (1943–1979)

Nel 1943 Warren S. McCulloch e Walter Pitts pubblicarono una delle opere fondamentali che contribuirono alla nascita dell'Intelligenza Artificiale: "*A Logical Calculus of the Ideas Immanent in Nervous Activity*". In questo studio venne ipotizzato per la prima volta che il cervello umano potesse essere interpretato come un sistema di elaborazione e venne introdotto il primo modello matematico di neurone artificiale ponendo le basi teoriche delle future reti neurali e, più in generale, del deep learning. Questo, però, rimase un modello matematico puramente teorico a causa delle limitate capacità tecnologiche e computazionali dell'epoca. Ulteriori passi avanti nella creazione dell'Intelligenza Artificiale vennero fatti nel 1950 con l'articolo "*Computing Machinery and Intelligence*" scritto dal matematico britannico Alan Turing che introduceva il Test di Turing, un test utilizzato per stabilire l'intelligenza delle macchine basato sulle loro capacità di imitare comportamenti umani.

Una svolta estremamente importante nella storia dell'Intelligenza Artificiale si ebbe nel 1951, quando Marvin Minsky e Dean Edmunds realizzarono la prima rete neurale, lo *Stochastic Neural Analog Reinforcement Calculator* (SNARC). I due ricercatori cercarono di istruire le macchine a riprodurre comportamenti umani tramite l'apprendimento per rinforzo, ossia mediante un sistema di premi e punizioni

Nel 1952 venne sviluppato da Arthur Samuel la prima macchina in grado di competere con l'uomo nel gioco della dama, poi migliorata nel 1955. In quest'ultima versione la macchina non era solo in grado di giocare ma riusciva ad apprendere nuove mosse, analizzare gli errori e non ripeterli.

Nel 1955 John McCarthy utilizzò per la prima volta il termine "Intelligenza Artificiale", successivamente proposto durante la conferenza di Dartmouth del 1956.

Sebbene gli anni 60 si avvicinarono al primo periodo di perdita di fiducia nell'AI furono anche quelli in cui venne creato il primo *chatbot*[1], ELIZA, sviluppato nel 1966.

La Figura 1.3 mostra un esempio del chatbot ELIZA.

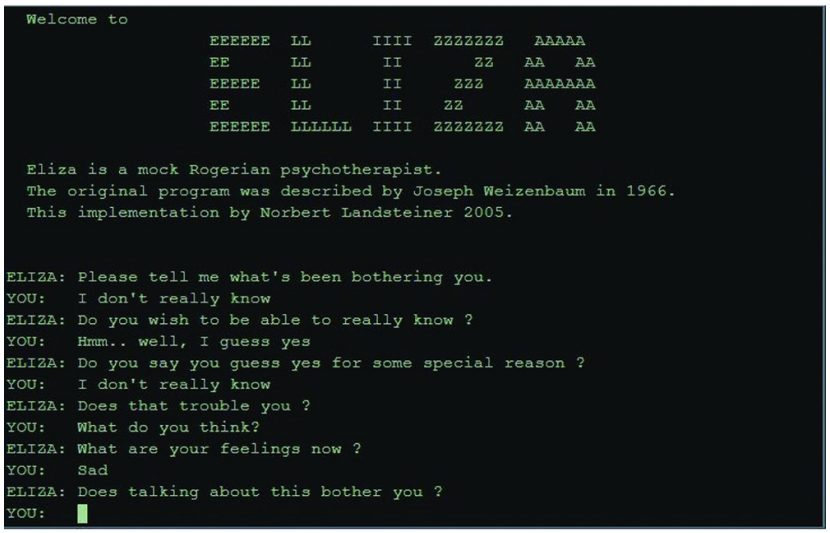


Figura 1.3: ELIZA (fonte: NeonTools).

ELIZA fu sviluppato da Joseph Weizenbaum ed è considerato il primo chatbot della storia. Quest'ultimo applicava semplici regole per analizzare le frasi digitate dall'utente e forniva risposte coerenti alle sue affermazioni. Fra i vari script il più noto fu "DOCTOR", il quale aveva il compito di simulare il comportamento di uno psicologo e formulare risposte pertinenti a determinate domande riguardanti stati d'animo, come tristezza, ansia e depressione. La cosa che più impressionò Weizenbaum di "DOCTOR" fu come le persone cominciarono a considerarlo "umano" e, molti, addirittura si affezionarono al computer e instaurarono un vero e proprio legame emotivo con quest'ultimo. Da questo programma nasce l'"*effetto ELIZA*" che si verifica quando le persone attribuiscono a un computer capacità che, in realtà, non ha, come, per l'appunto, la comprensione delle emozioni umane.

Negli anni settanta comincia ad esserci un primo punto di rottura; infatti nonostante i fondi e l'entusiasmo dei ricercatori le tecnologie dell'epoca non risultarono abbastanza avanzate per riuscire a creare programmi di Intelligenza Artificiale in grado di rispettare le aspettative. Nel 1973 iniziò ufficialmente il primo inverno dell'AI. Nel rapporto Lighthill vennero messi in discussione i progressi compiuti fino a quel momento e il governo britannico decise di interrompere gran parte dei finanziamenti destinati alla ricerca. In realtà, questa crisi era iniziata già alcuni anni prima: nel 1969 Marvin Minsky e Seymour Papert, nel loro libro Perceptrons, evidenziarono i limiti del *Perceptron* di Rosenblatt, che rappresentava uno dei primi esempi di rete neurale. Quest'ultimo, essendo costituito da un unico strato, poteva compiere soltanto operazioni elementari, ma non quelle più complesse. L'aggiunta di ulteriori strati di neuroni avrebbe consentito di ottenere risultati migliori; tuttavia, le tecnologie dell'epoca non erano ancora in grado di gestirli. Queste intuizioni anticiparono alcuni concetti fondamentali del Deep Learning.

La situazione rimase tale fino agli anni ottanta, quando, seppur per breve tempo, si riscoprì un interesse per l'Intelligenza Artificiale.

[1] Un *chatbot* è un programma o un'applicazione utilizzabili dall'utenza per conversare tramite voce o testo (definizione tratta da Amazon Web Services (AWS)).

1.2.2 L'affermazione dei sistemi esperti (1980–1987)

Con il rinnovato entusiasmo per l'Intelligenza Artificiale negli anni 80 vennero creati i sistemi esperti.

Essi rappresentano quel campo dell'Intelligenza Artificiale in grado di rispondere a domande e risolvere problemi riguardanti specifici campi umani; essi sono programmi software che lavorano in ambienti applicativi limitati superando, talvolta, le capacità umane in quello specifico dominio. Un esempio di sistemi esperti molto diffusi sono i programmi diagnostici in medicina. Una volta acquisita la conoscenza del dominio questo tipo di sistemi sono in grado di acquisire informazioni in input e rilasciare output coerenti. La particolarità è che si possono ripercorre tutti i passaggi logici usati per arrivare alla decisione finale. Questa tecnica, basata sulla totale trasparenza nei passaggi di programmazione, viene chiamata Glass Box. Questo è un approccio diverso da quello utilizzato dalle reti neurali e dal deep learning nel quale non è possibile ripercorrere le tappe, chiamato Black Box.

I sistemi esperti sono costituiti da tre componenti fondamentali:

- *Base di conoscenza (Knowledge Base)*; rappresenta il database delle regole utilizzate per iniziare il ragionamento logico e fornire risposte adeguate al problema.
- *Motore inferenziale*; elabora le conoscenze contenute nella knowledge base, le interpreta e fornisce la soluzione. Inoltre, al suo interno, applica una serie di regole if-then, le quali indicano i legami che si trovano fra i vari eventi. Al verificarsi di una condizione ci sarà una soluzione che può essere rappresentata da un'altra domanda o dall'esecuzione di un'azione. Il motore inferenziale utilizza due metodi:
  - *Foreward chaining (catena in avanti)*; consiste nel partire da fatti semplici, applicarvi regole di inferenza e estrarre dati finché non si è raggiunto l'obiettivo;
  - *Backward chaining (catena all'indietro)*; consiste nel procedere a ritroso per trovare tutte le regole che hanno condotto all'obiettivo finale.
- *Interfaccia utente*; mette in comunicazione l'utente con il motore inferenziale, consentendo ad esso di utilizzare il sistema esperto. Questo può essere rappresentato da un'interfaccia input/output; all'interno di quest'ultima l'utente scrive una domanda da tastiera e poi visualizza la risposta elaborata dal sistema sullo schermo, o su un'altra interfaccia evoluta che permette un'interazione più naturale e intuitiva.

I primi sistemi esperti nacquero negli anni Settanta e rappresentarono la prima generazione. Quest'ultima utilizzava la logica booleana e un modello deterministico basato sul rapporto causa-effetto. Uno dei sistemi esperti più importanti fu MYCIN, sviluppato per diagnosticare infezioni batteriche del sangue e suggerire la terapia antibiotica più appropriata. Tuttavia, questo approccio risultava limitato a causa dei conflitti logici che potevano entrare in contrasto con il ragionamento artificiale. Per superare tale limite nacque la seconda generazione di sistemi esperti, basata su un modello probabilistico che considerava le cause e i loro possibili effetti. Sebbene questi sistemi gestissero meglio l'incertezza, non garantivano che la risposta più probabile fosse anche la più utile.

Nonostante l'idea di sistemi in grado di superare l'uomo in ambiti specifici fosse stata accolta con grande entusiasmo negli anni Ottanta, emerse ben presto un problema fondamentale per il loro sviluppo: il knowledge acquisition bottleneck (collo di bottiglia dell'acquisizione della conoscenza). Infatti, estrarre e formalizzare la conoscenza umana si dimostrò un processo lento, costoso e caratterizzato da un elevato margine di errore, poiché la conoscenza umana è spesso intuitiva e non basata su regole rigide. Inoltre, questo approccio non risultava

scalabile, in quanto ogni nuovo dominio richiedeva di essere costruito da zero, senza la possibilità di riutilizzare componenti già esistenti. Questi limiti determinarono un progressivo declino dei sistemi esperti, favorendo il ritorno dell'interesse verso le reti neurali artificiali

1.2.3 Il ritorno delle reti neurali (1987–2000)

A partire dal 1987 si assistette a un rinnovato interesse per le reti neurali artificiali passando alla teoria delle reti multistrato e segnando, ufficialmente, una nuova era dell'Intelligenza Artificiale. Nel 1986 David E. Rumelhart, Geoffrey E. Hinton e Ronald J. Williams pubblicarono l'articolo "*Learning Representations by Back-Propagating Errors*", all'interno del quale scrissero l'algoritmo di backpropagation che consentiva di addestrare reti a più strati. In questo algoritmo, tutte le connessioni fra nodi sono caratterizzate da un peso che viene aggiornato durante l'addestramento.

L'obiettivo era quello di modificare i pesi fino a ridurre al minimo la differenza tra l'output prodotto dalla rete e quello desiderato; per fare ciò viene utilizzata una funzione di perdita (loss function), la quale riporta la misura precisa dell'errore. Il peso di una connessione stabilisce la quantità di input da trasmettere e questo avrà influenza sull'output del nodo successivo. Ogni nodo ricava, quindi, l'input dal nodo precedente ma non su tutti questo avrà lo stesso peso.

Nella Figura 1.4 viene riportato un esempio di reti neurali multistrato.

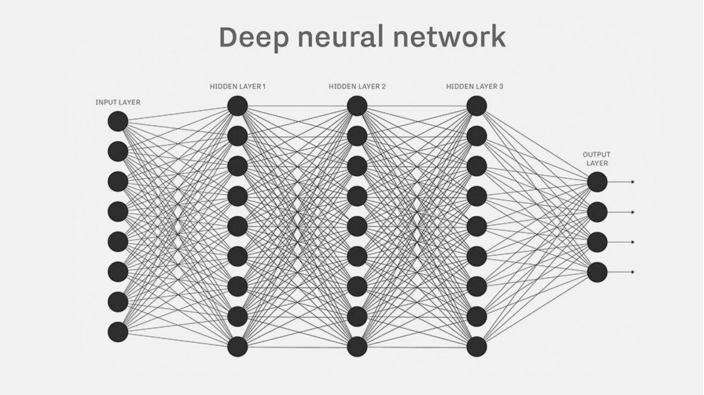


Figura 1.4: Reti neurali (fonte: ZeroUno).

L'algoritmo di backpropagation si sviluppa in due fasi:

- *Forward pass*: l'informazione viene propagata dallo strato di input fino allo strato di output, ottenendo la risposta della rete.
- *Backward pass*: l'errore calcolato sull'output viene propagato all'indietro attraverso tutti gli strati della rete, consentendo di calcolare quanto ogni peso abbia contribuito all'errore e facendo capire come andarlo a modificare.

La funzione di perdita può presentare dei minimi locali, nei quali l'algoritmo potrebbe arrestarsi prima di raggiungere il minimo globale. Dal punto di vista sperimentale questo problema si verifica raramente e può essere risolto semplicemente aggiungendo nuove connessioni.

L'algoritmo di backpropagation è alla base di gran parte dei moderni sistemi di Machine Learning basati su reti neurali, in particolare del Deep Learning.

1.2.4 L'affermazione del Machine Learning (anni '90-oggi)

Tra la fine degli anni Ottanta e l'inizio degli anni Novanta l'Intelligenza Artificiale attraversò il suo secondo inverno. A differenza del primo in quest'ultimo la ricerca non si fermò completamente nonostante la carenza di fondi. Di fondamentale importanza per superare questo periodo fu l'atteggiamento dei ricercatori, i quali, per recuperare la fiducia dei finanziatori, si videro costretti a costruire basi solide, non solo teoriche, per le ricerche sull'Intelligenza Artificiale. I sistemi esperti, che nella prima metà degli anni Ottanta sembrarono rappresentare una grande innovazione, vennero progressivamente rivalutati, poiché si dimostrarono poco convenienti. Questo favorì un rinnovato interesse nei confronti delle reti neurali, considerate un approccio più robusto e promettente.

Si deve a questi anni una rinnovata attenzione nei confronti del Machine Learning, che cominciò ad affermarsi nel corso degli anni Novanta grazie anche alla diffusione di Internet. Ciò consentì di condividere quantità di dati mai viste prima. Questa possibilità di condivisione era destinata ad aumentare in maniera esponenziale nel corso degli anni, modificando definitivamente l'approccio nei confronti dell'Intelligenza Artificiale.

Il Machine Learning è una categoria dell'Intelligenza Artificiale che sviluppa algoritmi capaci di apprendere nozioni attraverso l'acquisizione di dati e l'esperienza, migliorando progressivamente le proprie prestazioni senza la necessità di interventi di programmazione da parte del programmatore umano.

Negli anni 2000 si iniziò ad utilizzare l'apprendimento per rinforzo, paradigma che si basa sull'idea che questi agenti possano essere istruiti tramite un sistema di ricompense e punizioni, approccio successivamente adottato anche in altri settori, fra i quali la robotica. In contemporanea con la diffusione del World Wide Web e, successivamente, dei social media si entrò nell'epoca dei Big Data. La disponibilità di una quantità sempre più crescente di dati contribuì allo sviluppo di modelli di Intelligenza Artificiale sempre più accurati e capaci di affrontare problemi complessi.

1.2.5 I Big Data (2001–oggi)

> *I big data si riferiscono a set di dati enormi e complessi che i tradizionali sistemi di gestione dei dati non sono in grado di gestire. Se raccolti, gestiti e analizzati correttamente, i big data possono aiutare le organizzazioni a scoprire nuovi insight e prendere decisioni aziendali migliori.*(*Fonte:* IBM Think – What is Big Data?)

Sebbene l'uomo, fin dall'inizio della sua storia, abbia sempre raccolto dati, inizialmente in forma manuale e, successivamente, grazie all'evoluzione degli strumenti informatici, in forma digitale, con l'arrivo di Internet e la possibilità di condividere informazioni la quantità di dati raccolti e condivisi è cresciuta significativamente, dando vita all'era dei Big Data.

Nel corso degli anni 2000 le tecnologie della comunicazione si svilupparono rapidamente. Grazie all'utilizzo di satelliti e reti di comunicazione divenne possibile connettere l'intero pianeta e, con la diffusione degli smartphone, chiunque poté accedere ai Big Data. Per organizzare e dare un senso alle grandi quantità di dati contenute nei Big Data viene utilizzata la Data Science; questa scienza combina Intelligenza Artificiale e Machine Learning alla matematica e statistica per scoprire modelli e prevedere risultati. L'avvento dei Big Data rappresenta uno dei principali fattori che favorì la nascita del Deep Learning.

1.3 Machine Learning come fondamento dell'AI moderna

Nella Figura 1.5 vengono mostrate le differenze principali tra Intelligenza Artificiale, Machine Learning e Deep Learning.

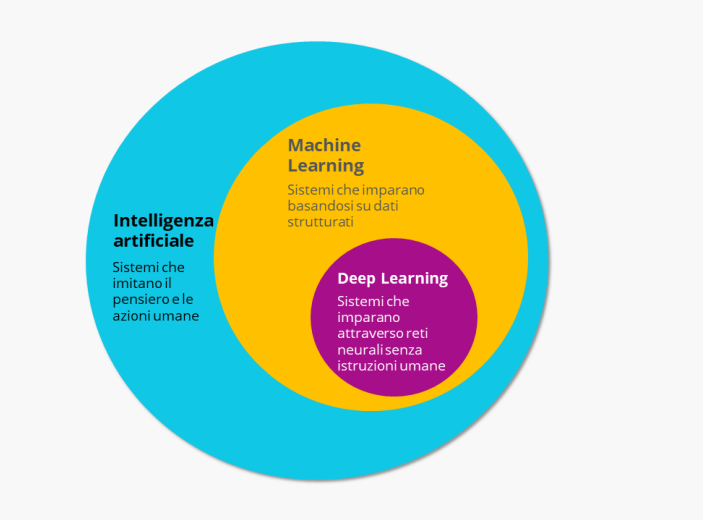


Figura 1.5: Le gerarchie dell'Intelligenza Artificiale (fonte: IONOS Digital Guide).

1.3.1 Definizione di Machine Learning

> *Il Machine Learning è il sottoinsieme dell'Intelligenza Artificiale (AI) basato su algoritmi in grado di "apprendere" i modelli dei dati di addestramento e, successivamente, fare inferenze accurate sui nuovi dati. Questa capacità di riconoscimento dei modelli consente ai modelli di Machine Learning di prendere decisioni o fare previsioni senza istruzioni esplicite e codificate* (*Fonte:* IBM Think – Machine Learning)

Già nel 1952 Arthur Samuel anticipò il concetto di Machine Learning con un software per giocare a dama, in grado di migliorare le proprie prestazioni sulla base dell'esperienza. Il Machine Learning viene spesso usato come sinonimo di Intelligenza Artificiale; tuttavia i due termini non coincidono. Infatti, mentre l'AI è la disciplina che contiene i metodi utilizzati per la costruzione dei sistemi di Intelligenza Artificiale, il Machine Learning rappresenta uno dei suoi principali approcci.

Quest'ultimo è un tipo di AI che non si basa sull'utilizzo di regole di programmazione prestabilite per la creazione di modelli, ma utilizza grandi quantità di dati per individuare schemi, effettuare previsioni e generare risposte; in generale, più dati verranno acquisiti, più le previsioni risulteranno affidabili.

Gli algoritmi di Machine Learning apprendono tramite un processo di addestramento durante il quale analizzano dati, ricercano relazioni e aggiornano i propri parametri interni per fare previsioni e cercare di minimizzare gli errori, migliorando così le prestazioni. Tra le varie branche del Machine Learning la più importante è il Deep Learning, che ha assunto un ruolo fondamentale negli sviluppi dell'Intelligenza Artificiale odierna.

1.3.2 Paradigmi di apprendimento

Dopo aver definito il Machine Learning, è necessario comprendere come avvengono i suoi processi di apprendimento.

I paradigmi di apprendimento descrivono gli obiettivi, il tipo di addestramento attuato e le tecniche utilizzate per ottimizzare le prestazioni. Ne esistono tre tipi: l'apprendimento supervisionato (Supervised Learning), l'apprendimento non supervisionato (Unsupervised Learning) e l'apprendimento per rinforzo (Reinforcement Learning).

Nella Figura 1.6 viene presentato un grafico concettuale sui paradigmi di apprendimento.

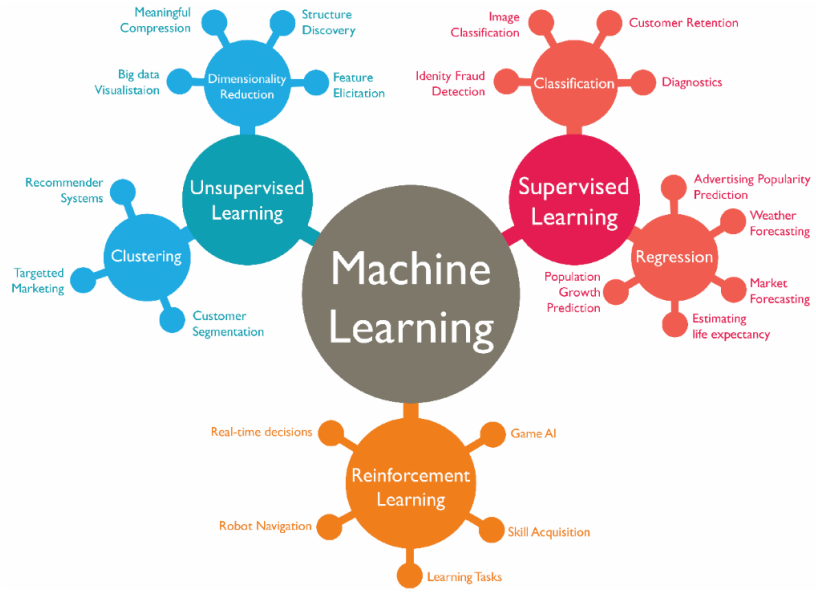


Figura 1.6: Paradigmi di apprendimento (fonte: Deep Learning Italia su Medium).

- *Apprendimento supervisionato*: rappresenta il paradigma di Machine Learning maggiormente diffuso. In questo approccio i modelli vengono addestrati utilizzando set di dati etichettati, grazie ai quali migliorano progressivamente la propria accuratezza. Durante la fase di addestramento, l'algoritmo apprende la relazione tra le variabili indipendenti (input) e le variabili dipendenti (output o target). Questo paradigma viene impiegato principalmente per risolvere problemi di classificazione e regressione e si basa sull'utilizzo di dati etichettati, che forniscono al modello le informazioni necessarie per apprendere le corrette associazioni tra input e output.
- *Apprendimento non supervisionato*: in questo paradigma il modello viene addestrato utilizzando set di dati non etichettati. L'algoritmo analizza autonomamente i dati con l'obiettivo di individuare schemi, dipendenze, correlazioni e tendenze che non risultano immediatamente evidenti. A differenza dell'apprendimento supervisionato, non è presente una risposta corretta o un riferimento esterno con cui confrontare i risultati ottenuti.
- *Apprendimento rinforzato*: in questo paradigma il modello viene addestrato attraverso tentativi ed errori. Durante l'apprendimento, le azioni compiute vengono valutate mediante un sistema di ricompense e penalità: le azioni corrette vengono premiate, mentre quelle non corrette vengono penalizzate. A differenza dell'apprendimento supervisionato, il cui obiettivo è ridurre al minimo l'errore, nell'apprendimento per rinforzo l'algoritmo ottimizza i propri parametri per massimizzare la ricompensa ottenuta.

Questi paradigmi di apprendimento rappresentano le principali modalità con cui gli algoritmi di Machine Learning acquisiscono dati. Tuttavia, nella pratica, tali paradigmi non sono sempre utilizzati in modo isolato. Esistono tecniche che combinano le caratteristiche di più approcci:

- *apprendimento semi-supervisionato*, che unisce elementi dell'apprendimento supervisionato e di quello non supervisionato;
- *apprendimento auto-supervisionato*, che sfrutta dati non etichettati per l'addestramento del modello.

Uno stesso modello può essere addestrato in più fasi, impiegando differenti paradigmi di apprendimento. Questo accade nei modelli linguistici di grandi dimensioni, inizialmente pre-addestrati mediante apprendimento auto-supervisionato e successivamente perfezionati attraverso algoritmi di apprendimento supervisionato oppure di apprendimento per rinforzo. Questi paradigmi sono gli stessi utilizzati nel Deep Learning.

1.3.3 L'evoluzione del Deep Learning

Dal 2010 l'introduzione di nuove tecniche nelle reti neurali ha dato origine al più importnante sottoinsieme del Machine Learning: il Deep Learning. Quest'ultimo si basa sul concetto di reti neurali multistrato, costituite da numerosi strati di neuroni interconnessi ispirati al funzionamento del cervello umano. Il Deep Learning risulta fondamentale per gran parte dell'Intelligenza Artificiale, come, ad esempio, la *computer vision*[2]e l'AI generativa.

Il Deep Learning può addestrare solo modelli che siano rappresentati da almeno 4 livelli nelle reti neurali; questa struttura flessibile e regolabile spiega la sua potenza. Le differenze principali fra Machine Learning e Deep Learning sono le seguenti:

- *Intervento umano*: il Deep Learning necessita di un intervento del programmatore molto inferiore rispetto al Machine Learning.
- *Tempi e risorse computazionali*: il Machine Learning richiede una quantità minore di risorse computazionali e ha tempi di addestramento ridotti, mentre il Deep Learning richiede più risorse così da avere una maggiore potenza di calcolo, necessita di tempi di addestramento più lunghi, garantendo, però, maggiori prestazioni.
- *Metodo di apprendimento*: il Machine Learning apprende dai dati tramite algoritmi precisi e solo dopo prende decisioni, mentre il Deep Learning apprende attraverso reti neurali che filtrano in autonomia i dati di interesse.

Oggi il Deep Learning rappresenta il principale protagonista dell'Intelligenza Artificiale e, grazie a queste sue caratteristiche, sta contribuendo allo sviluppo dei nuovi sistemi.

1.4 Modelli discriminativi e generativi

Con l'evoluzione degli studi sull' Intelligenza Artificiale sono cambiati i metodi tramite i quali si valuta la sua affidabilità. Come visto in precedenza nel test di Turing, l'intelligenza di una macchina dipendeva dalle sue capacità di imitare il comportamento umano; oggi per valutarne la bravura si verifica la sua capacità di prendere decisioni e formulare previsioni.

I modelli di Intelligenza Artificiale basati sul Machine Learning possono essere classificati in base alla metodologia con cui elaborano i dati. Secondo questo criterio, si distinguono in modelli generativi e modelli discriminativi.

1.4.1 Modelli discriminativi

I modelli discriminativi di Intelligenza Artificiale sono stati progettati con lo scopo di distinguere e riconoscere elementi appartenenti a classi differenti. Questa tipologia di modelli, generalmente, utilizza l'apprendimento supervisionato e vengono addestrati con l'obiettivo di avere un'AI capace di apprendere il confine decisionale tra le diverse classi di dati. In particolare, per farli funzionare correttamente, quando si costruiscono, si ha la

[2]La *computer vision* è un sottocampo dell'Intelligenza Artificiale che consente ai sistemi informatici di elaborare, analizzare e interpretare immagini e video (definizione rielaborata da IBM Think).

necessità di raccogliere una grande quantità di dati e si istruisce il modello a riconoscerli così da classificare correttamente un nuovo input sulla base delle caratteristiche osservate. In generale, più informazioni vengono acquisite più il sistema risulta affidabile e in grado di fornire previsioni corrette. Per questo motivo l'Intelligenza Artificiale Discriminativa risulta efficace nelle attività di classificazione, previsione e regressione.

Le principali applicazioni dei modelli discriminativi comprendono il riconoscimento di immagini, il riconoscimento del parlato, il rilevamento delle e-mail di spam e l'analisi del sentiment. Queste loro capacità di classificare e riconoscere i dati fanno si che questi modelli possano essere impiegati in numerosi dispositivi e servizi, come smartphone, fotocamere e assistenti vocali.

Un ulteriore esempio di applicazione di Intelligenza Artificiale discriminativa è rappresentato dalle cuffie bluetooth dotate di cancellazione del rumore: queste utilizzano modelli discriminativi per riconoscere i rumori esterni e ne riducono l'intensità. Diversamente dalle AI discriminative quelle generative non vengono solo usate per la classificazione e il riconoscimento di dati ma possono essere impiegate anche per la generazione di nuovi elementi.

1.4.2 Modelli generativi

I modelli di Intelligenza Artificiale generativi sono tecnologie in grado di creare contenuti originali, come testi, immagini, video, audio o codice, in risposta a una richiesta dell'utente. Essi si basano su modelli di Machine Learning, in particolare di Deep Learning, che apprendono identificando schemi e relazioni presenti in grandi quantità di dati. Le informazioni acquisite durante l'addestramento vengono successivamente utilizzate per comprendere le richieste ricevute e generare nuovi contenuti coerenti e pertinenti.

L'AI generativa è costituita da tre fasi principali:

- *formazione*: si basa su *foundation model*[3], modelli di Deep Learning addestrati su grandi quantità di dati per generare contenuti in risposta a delle richieste degli utenti;
- *ottimizzazione*: avviene attraverso la messa a punto (fine-tuning) e, in alcuni casi, mediante l'apprendimento per rinforzo con feedback umano (*Reinforcement Learning from Human Feedback, RLHF*)[4], che ne migliora progressivamente la qualità delle risposte;
- *generazione, valutazione e riottimizzazione*: dopo l'addestramento, il modello viene continuamente valutato e ulteriormente ottimizzato per verificare la sua pertinenza e la qualità delle risposte alle domande di un utente.

Nella Figura 1.7 sono rappresentate alcune applicazioni dell'AI generativa.

1.4.3 Confronto fra Intelligenza Artificiale Discriminativa e Generativa

Come si è potuto comprendere nei paragrafi precedenti, l'Intelligenza Artificiale generativa e quella discriminativa svolgono due compiti ben distinti. Un modello generativo analizza i dati forniti durante l'addestramento, apprendendone le caratteristiche al fine di generare

[3] I foundation model sono modelli di Intelligenza Artificiale addestrati su grandi quantità di dati e progettati per svolgere un'ampia gamma di compiti generali. Essi costituiscono la base per lo sviluppo di applicazioni e modelli più specializzati (definizione rielaborata da IBM Think).

[4] Il Reinforcement Learning from Human Feedback (RLHF) è una tecnica di Machine Learning che integra il feedback umano nel processo di apprendimento per migliorare le prestazioni del modello e allinearne il comportamento agli obiettivi e alle preferenze degli utenti. Questa tecnica è ampiamente utilizzata nei modelli di Intelligenza Artificiale generativa (definizione rielaborata da Amazon Web Services (AWS)).

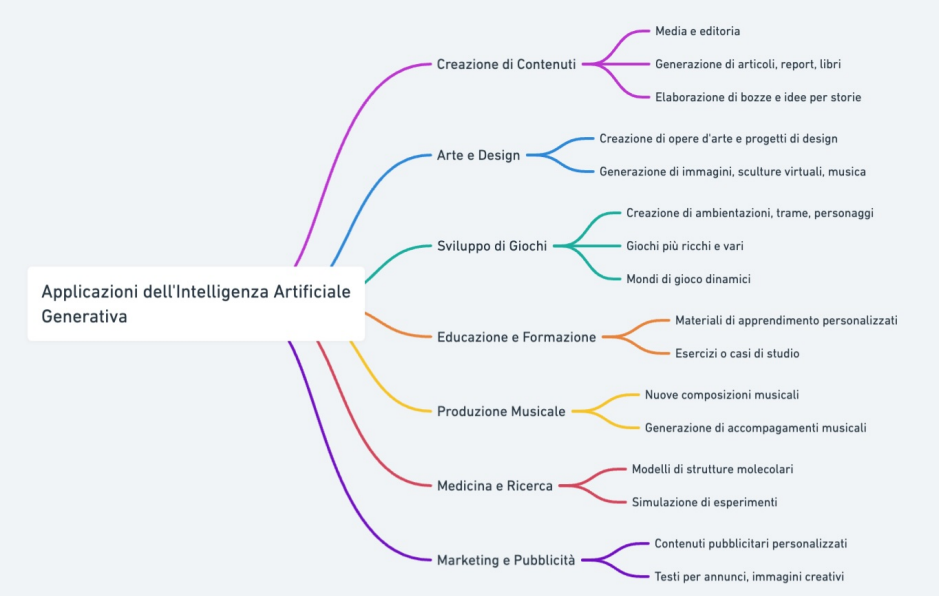
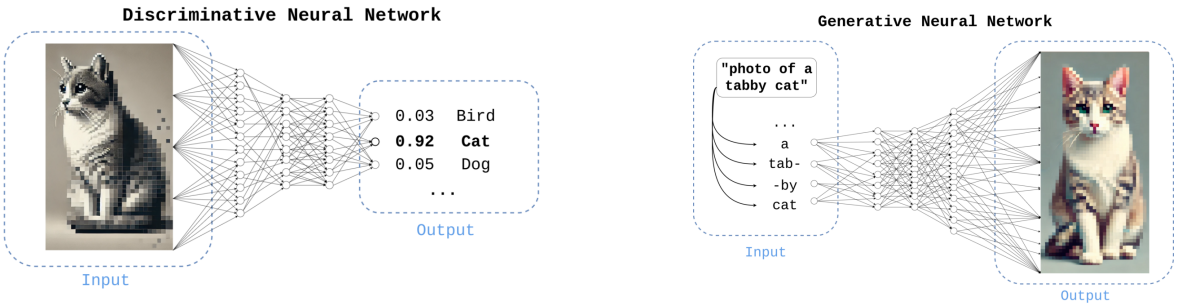


Figura 1.7: Applicazioni dell'Intelligenza Artificiale Generativa (fonte: Cittadinanza Digitale).

nuovi contenuti. Al contrario, un modello discriminativo apprende le differenze tra i dati ricevuti,con l'obiettivo di riconoscerli e classificarli correttamente in presenza di un nuovo input.

Nella Figura 1.8 è riportato un confronto tra un modello di Intelligenza Artificiale discriminativa e un modello di Intelligenza Artificiale generativa.



(a) Modello discriminativo

(b) Modello generativo

Figura 1.8: Confronto tra modello discriminativo e modello generativo (fonte: Wikipedia).

Nella Tabella 1.1 vengono confrontate alcune caratteristiche dei due modelli.

1.5 Applicazioni dell'Intelligenza Artificiale discriminativa oggetto dei case study

Per comprendere meglio il funzionamento dell'Intelligenza Artificiale Discriminativa sono stati svolti dei case study, che verranno riportati nei capitoli successivi. Per la loro realizzazione è stato utilizzato Amazon Web Services (AWS), una piattaforma di cloud computing che mette a disposizione numerosi servizi basati sull'Intelligenza Artificiale.

Prima di spiegare Amazon Web Services e i servizi di cui si è usufruito, risulta necessario illustrare le tipologie di Intelligenza Artificiale utilizzate.

Caratteristica	Modelli generativi	Modelli discriminativi
Obiettivo	Generare nuovi contenuti	Classificare e riconoscere i dati
Funzionamento	Apprendono le caratteristiche dei dati per creare nuovi contenuti	Apprendono le differenze tra le classi di dati per classificarle
Output	Testi, immagini, audio, video e codice	Categoria, previsione o classe di appartenenza
Principali applicazioni	Chatbot, generazione di immagini, contenuti multimediali	Riconoscimento di immagini, spam, analisi del sentiment, riconoscimento del parlato
Risorse computazionali	Generalmente più elevate	Generalmente inferiori

Tabella 1.1: Confronto tra modelli generativi e modelli discriminativi (fonte: elaborazione da HumAI).

1.5.1 Moderazione dei contenuti

La moderazione dei contenuti rappresenta una delle principali sfide per le piattaforme digitali, a causa dell'enorme quantità di informazioni generate quotidianamente dagli utenti. La diffusione degli smartphone e dei social media consente, infatti, di pubblicare contenuti in qualsiasi momento e da qualsiasi luogo, rendendo impraticabile un controllo esclusivamente umano.

Nella Figura 1.9 sono rappresentati alcuni esempi di moderazione dei contenuti.

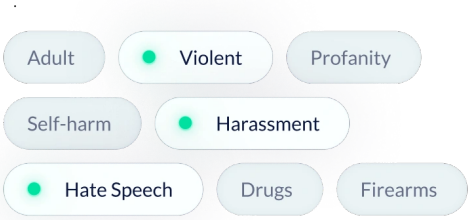


Figura 1.9: Moderazione dei contenuti (fonte: Copyleaks).

La complessità della moderazione è accentuata anche dalla varietà dei contenuti condivisi, che comprendono testi, immagini, video e dirette streaming. Ciascun formato richiede tecniche di analisi specifiche per individuare eventuali contenuti dannosi o non conformi alle politiche della piattaforma.

A ciò si aggiunge l'impiego di strategie sempre più sofisticate per eludere i sistemi di controllo, come l'utilizzo di account bot e dei servizi di messaggistica privata, che rendono più difficile individuare tempestivamente contenuti inappropriati e garantire un'efficace moderazione.

Molte aziende affiancano ai moderatori umani strumenti basati sull'Intelligenza Artificiale. Con l'utilizzo di algoritmi di Machine Learning questi sistemi diventano in grado di identificare automaticamente i contenuti dannosi o inappropriati, contribuendo a preservare la sicurezza delle community e delle piattaforme online così da proteggere gli utenti. Inoltre, l'automazione di una parte del processo di moderazione consente di ridurre il carico di lavoro dei moderatori, permettendo loro di concentrarsi sui casi più complessi. La combinazione tra

l'esperienza umana e le tecnologie di Intelligenza Artificiale rende il processo di moderazione più efficace nella gestione del crescente volume di contenuti generati dagli utenti.

1.5.2 Riconoscimento visivo

Le attuali tecnologie di riconoscimento facciale trovano le loro fondamenta nell'Intelligenza Artificiale e nelle reti neurali multistrato. Queste consentono di identificare o verificare l'identità di una persona in base al suo volto.

Nella Figura 1.10 troviamo un esempio di riconoscimento facciale.

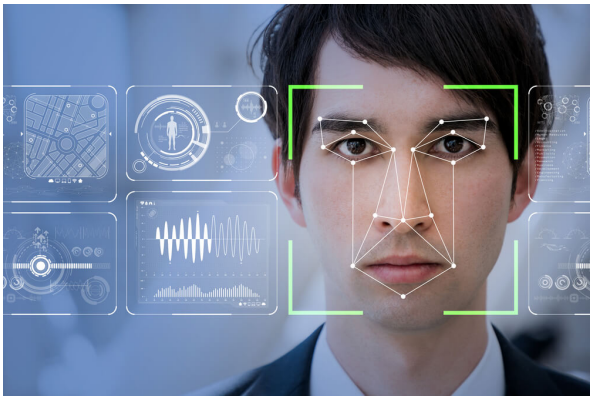


Figura 1.10: Riconoscimento facciale (fonte: AI4Business).

Il riconoscimento facciale rappresenta una delle applicazioni più diffuse dell'Intelligenza Artificiale discriminativa. In questa tecnologia è fondamentale l'utilizzo della *biometria facciale* così da consentire la verifica e l'autenticazione dell'identità di una persona attraverso il riconoscimento delle caratteristiche del volto, attuando la stessa logica associativa utilizzata dall'uomo per riconoscere le persone. Il suo funzionamento si basa sul confronto tra l'immagine acquisita e quella precedentemente registrata per verificare che entrambe appartengano alla stessa persona. I sistemi di riconoscimento facciale svolgono due funzioni principali:

- *verifica dell'identità*: si svolge durante la fase di registrazione di un utente, confrontando la fotografia del documento di identità con quella acquisita durante il processo;
- *l'autenticazione degli utenti già registrati*: questi possono accedere a servizi o aree riservate semplicemente attraverso il riconoscimento facciale, senza dover utilizzare password o altre credenziali.

Per questo motivo il riconoscimento facciale viene oggi impiegato in numerosi servizi e applicazioni, come l'accesso allo smartphone, l'autorizzazione di pagamenti e il controllo degli accessi.

1.5.3 Analisi dei sentimenti

L'analisi dei sentimenti, nota anche come opinion mining, rappresenta una delle principali applicazioni dell'Intelligenza Artificiale discriminativa. Questa tecnica consiste nell'analizzare grandi quantità di dati testuali con l'obiettivo di determinare se il tono emotivo di un testo sia positivo, negativo o neutro.

Il funzionamento dell'analisi dei sentimenti si basa sull'utilizzo di tecnologie di elaborazione del linguaggio naturale (*Natural Language Processing, NLP*)[5] e di algoritmi di Machine Learning, che consentono al software di analizzare e interpretare automaticamente il testo in modo simile agli esseri umani. Il software può utilizzare un approccio basato su regole, un approccio basato sul Machine Learning oppure una combinazione dei due, definita approccio ibrido. Grazie a queste tecnologie è possibile classificare automaticamente le opinioni espresse dagli utenti e analizzare grandi quantità di dati testuali in modo rapido ed efficiente.

Esistono diversi tipi di analisi dei sentimenti che vengono classificati in base alla polarità. Quest'ultima valuta se il sentimento espresso sia positivo, negativo o neutro.

Ne esistono tre tipi:

- *dettagliata (graduata)*: misura il livello di intensità dell'emozione espressa;
- *basata sugli aspetti (ABSA)*: limita l'analisi a uno specifico aspetto di un prodotto o di un servizio, valutando separatamente il sentimento espresso per ciascuna caratteristica;
- *rilevamento delle emozioni*: cerca di identificare emozioni specifiche (come frustrazione o indifferenza).

L'analisi dei tre case study riguardanti queste tipologie di Intelligenza Artificiale discriminativa è stata condotta utilizzando i servizi messi a disposizione dalla piattaforma Amazon Web Services (AWS).

[5] Il *Natural Language Processing* (*NLP*) è un sottocampo dell'informatica e dell'Intelligenza Artificiale che, attraverso tecniche di Machine Learning, consente ai computer di comprendere, interpretare e generare il linguaggio umano (definizione rielaborata da IBM Think).

CAPITOLO 2

Amazon Web Services

In questo capitolo viene presentata la piattaforma Amazon Web Services (AWS). Nei primi paragrafi vengono descritti i principi fondamentali del cloud computing e i principali modelli di servizio e distribuzione. Successivamente vengono analizzate l'infrastruttura globale e l'architettura della piattaforma, con particolare attenzione alle modalità di accesso e interazione con i servizi AWS. Infine, vengono descritti alcuni dei principali servizi di storage, calcolo e Intelligenza Artificiale, approfondendo, in particolare, Amazon Comprehend e Amazon Rekognition, che verranno utilizzati nelle sperimentazioni descritte nei capitoli successivi.

2.1 Introduzione ad Amazon Web Services

Amazon Web Services (AWS) è la piattaforma di cloud computing sviluppata da Amazon nel 2006 per offrire alle aziende servizi di infrastruttura IT tramite Internet, oggi comunemente conosciuti come servizio di cloud computing. AWS permette ad aziende e organizzazioni di utilizzare risorse informatiche on-demand, evitando gli elevati costi iniziali necessari per l'acquisto e la gestione di infrastrutture hardware proprie. Grazie a questo modello è possibile accedere rapidamente alle risorse necessarie e pagare esclusivamente quelle effettivamente utilizzate.

AWS non è un singolo servizio, ma una piattaforma composta da numerosi servizi cloud gestiti attraverso un'unica interfaccia. L'offerta comprende servizi di tipo *Infrastructure as a Service (IaaS)* e *Platform as a Service (PaaS)*, che consentono alle organizzazioni di sviluppare, distribuire e gestire applicazioni in modo semplice ed efficiente.

Nel corso degli anni AWS ha ampliato la propria offerta, arrivando a mettere a disposizione oltre duecento servizi cloud. Tra questi rientrano servizi di elaborazione, archiviazione dei dati, database, reti, sicurezza, analisi dei dati, *Internet of Things (IoT)*[1], strumenti per sviluppatori, strumenti di gestione, applicazioni aziendali e servizi di Intelligenza Artificiale. Tutti i servizi sono disponibili on-demand e possono essere attivati in pochi minuti, permettendo alle organizzazioni di rispondere rapidamente alle esigenze operative.

Attualmente AWS è utilizzata da startup, piccole e medie imprese, grandi aziende, organizzazioni no profit ed enti pubblici. Uno dei principali vantaggi della piattaforma è la

[1] L'*IoT* (*Internet of Things*) si riferisce a una rete di dispositivi fisici, veicoli, apparecchi e altri oggetti incorporati con sensori, software e connettività che consentono di raccogliere e condividere dati (definizione presa da IBM IoT).

possibilità di sostituire gli elevati investimenti iniziali richiesti dalle infrastrutture tradizionali con costi variabili, consentendo alle organizzazioni di utilizzare solo le risorse necessarie in base alle proprie esigenze.

I principali servizi offerti da Amazon Web Services saranno analizzati nei paragrafi successivi.

Nella Figura 2.1 troviamo uno schema dei vari servizi AWS.



Figura 2.1: Schema AWS (fonte: Maceforce).

2.1.1 Fondamenti del cloud computing: IaaS, PaaS e SaaS

Il cloud computing è un modello che consente l'accesso alle risorse di computing quali server fisici e virtuali, data storage (sistemi di archiviazione dati), funzionalità di rete, strumenti per lo sviluppo di applicazioni, software, piattaforme di analisi basate sull'Intelligenza Artificiale, tramite l'utilizzo di internet. L'accesso a tali risorse avviene tramite un metodo di pagamento *pay-per-use*, che consente all'utente di pagare solo i servizi effettivamente utilizzati. Il cloud computing nasce con l'obiettivo di offrire un'infrastruttura *scalabile*[2] e più flessibile rispetto a quelle tradizionali.

Secondo il *National Institute of Standards and Technology (NIST)*, il modello di cloud computing è caratterizzato da *cinque caratteristiche essenziali*, *tre modelli di servizio* e *quattro modelli di distribuzione* (deployment model).

Le cinque caratteristiche essenziali sono rappresentate da:

1. *Self-service su richiesta (on-demand self-service)*: l'utente può accedere alle risorse in maniera automatica senza chiedere permessi al fornitore.

2. *Ampio accesso alla rete (broad network access)*: le risorse sono accessibili tramite la rete e gli utenti possono usufruirne attraverso diversi dispositivi elettronici.

3. *Condivisione delle risorse (resource pooling)*: il cloud utilizza un modello di condivisione multi-tenant, tramite il quale le risorse vengono assegnate dinamicamente in base alle richieste.

4. *Elasticità rapida (rapid elasticity)*: le risorse vengono fornite e rilasciate rapidamente; tendenzialmente queste operazioni vengono gestite in maniera automatica per adattare le risorse in base alla domanda (scalabilità).

[2] La *scalabilità* in ambito cloud computing è la capacità di un'infrastruttura di aumentare o ridurre le proprie risorse in base alle esigenze operative, mantenendo costanti le performance (definizione presa da Timenterprise).

5. *Servizio misurato (measured service)*: i sistemi cloud ottimizzano l'utilizzo di risorse mediante la valutazione del servizio offerto; un esempio è rappresentato dallo spazio di archiviazione. L'utilizzo delle risorse può essere monitorato così che l'utente abbia sempre la situazione sotto controllo.

I servizi sono rappresentati da *IaaS (Infrastructure as a Service)*, *PaaS (Platform as a Service)* e *SaaS (Software as a Service)*:

- L'*Infrastructure as a Service (IaaS)* rappresenta un modello di servizio cloud che consente all'utente di utilizzare risorse informatiche, come capacità di elaborazione, sistemi di archiviazione, reti e altre risorse di calcolo, sulle quali è possibile installare ed eseguire qualsiasi software, sistema operativo e applicazione.

  In questo modello, l'utente non ha capacità di gestione o controllo sull'infrastruttura cloud sottostante; tuttavia, può mantenere il controllo sui sistemi operativi, spazio di archiviazione, applicazioni e, in alcuni casi, componenti della rete, come i firewall.

  IaaS fornisce l'accesso on-demand e permette di accedere alle risorse di calcolo tramite internet con un modello di pagamento basato sul consumo (pay-per-use), consentendo di aumentare o ridurre rapidamente le risorse in base alle esigenze.

- Il *Platform as a Service (PaaS)* è un modello di servizio cloud che consente agli utenti di sviluppare, distribuire e gestire applicazioni su un'infrastruttura messa a disposizione dal provider. Le applicazioni possono essere realizzate utilizzando linguaggi di programmazione, librerie, servizi e strumenti supportati dalla piattaforma.

  In questo modello è il provider a gestire l'infrastruttura cloud sottostante, comprendente rete, server, sistemi operativi, sistemi di archiviazione e gli altri componenti necessari al funzionamento della piattaforma. L'utente, invece, mantiene il controllo sulle applicazioni sviluppate e distribuite e, in alcuni casi, sulle impostazioni di configurazione dell'ambiente che le ospita.

  Il modello PaaS mette, inoltre, a disposizione un ambiente di sviluppo completo, accessibile on-demand, che consente agli sviluppatori di creare, testare, distribuire, aggiornare e gestire le applicazioni senza dover installare o mantenere un'infrastruttura hardware e software dedicata.

- Il *Software as a Service (SaaS)* è un modello di servizio cloud che mette a disposizione degli utenti applicazioni ospitate sull'infrastruttura del provider e permette l'accesso tramite internet.

  L'utilizzo dei servizi può avvenire attraverso un browser web oppure mediante interfacce di programmazione (API), permettendo l'accesso da diversi dispositivi.

  In questo modello è il provider ad occuparsi della gestione dell'infrastruttura, mentre l'utente utilizza e dispone esclusivamente di limitate possibilità di configurazione delle applicazioni personalizzabili. Quindi, la manutenzione, gli aggiornamenti e l'archiviazione dei dati sono gestiti direttamente dal provider, riducendo il rischio di perdita delle informazioni in caso di malfunzionamenti dei dispositivi utilizzati.

Esistono, poi, quattro modelli di distribuzione di cloud:

1. *private cloud*: si parla di cloud privato quando l'infrastruttura è destinata all'utilizzo di una singola organizzazione;

2. *community cloud*: si parla di community cloud quando l'infrastruttura è destinata ad essere utilizzata da più organizzazioni che, però, appartengono alla stessa comunità, quindi che condividono esigenze comuni;

3. *public cloud*: si parla di public cloud quando l'infrastruttura è disponibile a uso pubblico;
4. *hybrid cloud*: l'hybrid cloud viene integrato utilizzando due o più infrastrutture cloud.

La comprensione dei principi del cloud computing risulta fondamentale per comprendere le varie tecnologie di Amazon Web Services che verranno trattate in questo capitolo.

2.1.2 AWS nel panorama degli hyperscaler

Il termine hyperscaler deriva da hyperscale computing, un metodo di elaborazione dei dati progettato per garantire la scalabilità e l'espansione delle infrastrutture informatiche in risposta all'aumento della domanda di risorse. Per garantire l'accesso ai servizi su scala globale, i provider hyperscaler utilizzano numerosi data center distribuiti geograficamente. I data center sono composti da migliaia di server fisici che eseguono milioni di macchine virtuali. L'infrastruttura hyperscale utilizza la scalabilità orizzontale, permettendo di aumentare rapidamente le risorse disponibili man mano che cresce la domanda. Inoltre, tali infrastrutture devono garantire elevati livelli di prestazioni, produttività e ridondanza, assicurando fault tolerance e alta disponibilità. Queste caratteristiche rendono gli hyperscaler particolarmente adatti all'ambiente aziendale, dove applicazioni e dati richiedono infrastrutture flessibili, affidabili e facilmente scalabili.

Gli hyperscaler, in quanto provider di servizi cloud, offrono modelli di servizio quali IaaS, PaaS e SaaS, oltre a funzionalità di distribuzione, sviluppo, archiviazione cloud, networking, Intelligenza Artificiale e calcolo per le applicazioni. Tendenzialmente gli hyperscaler offrono funzionalità di cloud pubblico, molti fornitori offrono anche funzionalità di cloud privato.

Nella Figura 2.2 è rappresentato uno schema che mostra le differenze fra hyperscaler e data center tradizionali.

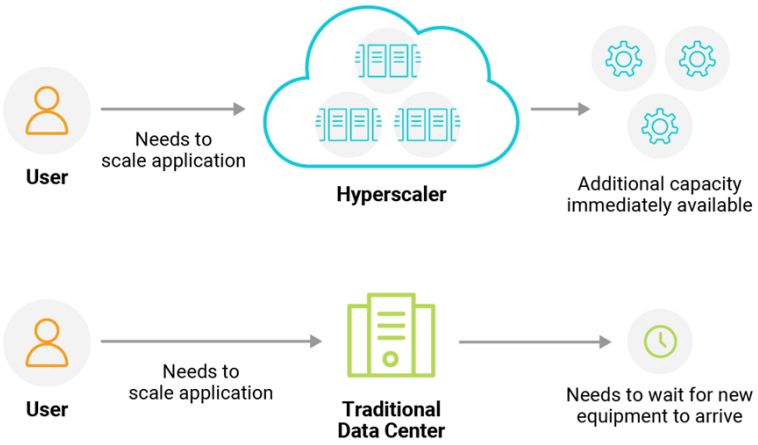


Figura 2.2: Schema differenziazione hyperscaler e data center tradizionali (fonte: Solarwinds).

Tra i principali hyperscaler rientrano Amazon Web Services (AWS), Microsoft Azure, Google Cloud, IBM Cloud e Alibaba Cloud. Dal 2023 Amazon Web Services è diventato il più grande servizio di hyperscaler fornendo opzioni di storage, potenza di elaborazione, automazione, database e data analytics.

2.1.3 Storia ed evoluzione di AWS

Lo sviluppo di Amazon Web Services (AWS) prese forma tra il 2003 e il 2006, periodo in cui Amazon iniziò a progettare un'infrastruttura informatica standardizzata e automatizzata

da offrire anche ad aziende esterne. Il primo servizio rilasciato da AWS fu Amazon Simple Storage Service (Amazon S3), presentato il 14 marzo 2006, che consentiva l'archiviazione dei dati garantendo elevati livelli di sicurezza, privacy e controllo. Nell'agosto dello stesso anno venne introdotto Amazon Elastic Compute Cloud (Amazon EC2), che rese disponibile capacità di elaborazione on-demand, permettendo agli utenti di utilizzare risorse di calcolo da remoto senza dover disporre di un'infrastruttura fisica dedicata. L'introduzione di questi servizi rappresentò un momento fondamentale per la crescita della piattaforma e segnò l'inizio dell'evoluzione di AWS nel settore del cloud computing.

Nel corso degli anni AWS ampliò progressivamente la propria offerta introducendo nuovi servizi dedicati ai database, al networking, al Machine Learning e a numerosi altri ambiti del cloud computing. Già nel 2015 AWS era diventata l'attività più redditizia di Amazon, consolidando nel tempo il proprio ruolo come una delle principali piattaforme di cloud computing a livello mondiale.

Nella Figura 2.3 viene riportato un grafico che descrive le fasi di crescita della piattaforma AWS; come si può notare, dal 2015 la crescita è esponenziale.

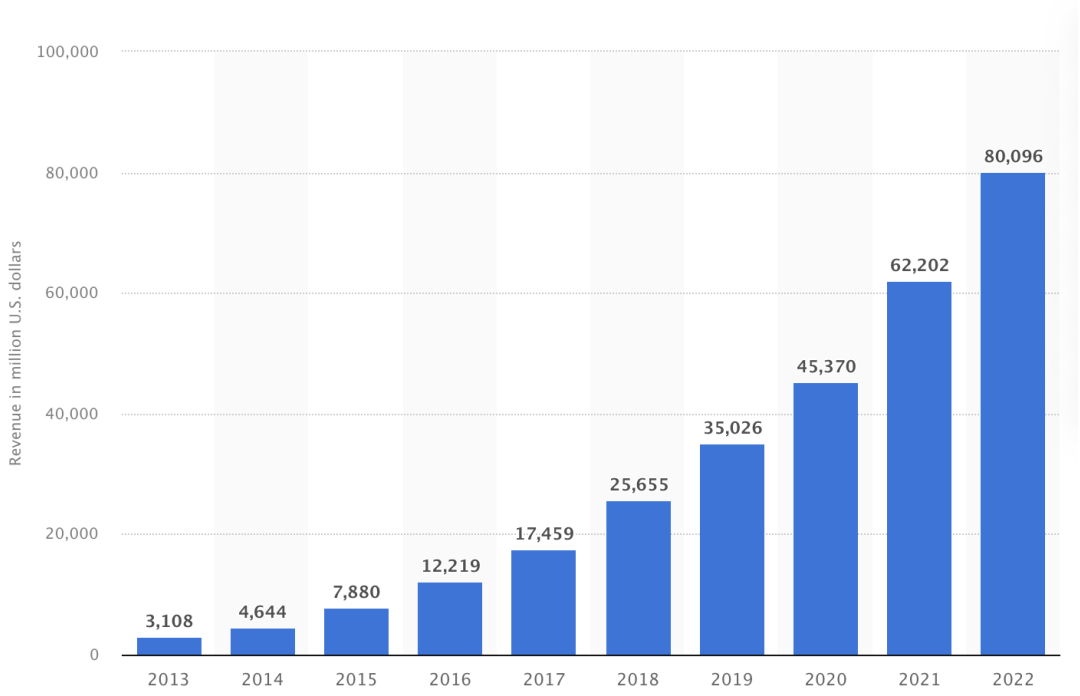


Figura 2.3: Schema crescita AWS (fonte: Kinsta).

2.2 Architettura e infrastruttura di AWS

L'architettura di Amazon Web Services è progettata per offrire un'infrastruttura cloud scalabile, affidabile e altamente disponibile, composta da servizi distribuiti a livello globale e organizzati secondo principi architetturali *modulari*[3].

Nei paragrafi seguenti vengono descritti gli elementi fondamentali che caratterizzano tale architettura, ovvero:

[3]La *modularizzazione* di un sistema (complesso) consente di suddividerlo in diversi sottosistemi (o moduli) interfacciati tra loro. In un sistema modulare le interazioni avvengono tra elementi dello stesso sottosistema, mentre i diversi sottosistemi sono connessi attraverso interfacce (definizione presa da Treccani).

- l'infrastruttura globale di AWS;
- l'organizzazione tecnica dei servizi;
- le modalità di accesso e interazione con la piattaforma.

2.2.1 Infrastruttura globale

L'infrastruttura globale di Amazon Web Services (AWS) è costituita da una rete di data center distribuiti a livello mondiale, progettata per garantire elevati livelli di affidabilità, resilienza e prestazioni. Ogni Regione AWS è composta da più Availability Zone (AZ) fisicamente separate e indipendenti, progettate per assicurare continuità operativa e tolleranza ai guasti. Questa organizzazione consente di distribuire applicazioni e dati su più zone, aumentando la disponibilità del servizio. Per rispondere alla crescente domanda di servizi cloud, AWS continua ad ampliare la propria infrastruttura attraverso l'apertura di nuove Regioni, consentendo ai clienti di scegliere l'area geografica in cui archiviare ed elaborare dati e applicazioni, in funzione delle proprie esigenze operative e dei requisiti di localizzazione dei dati. AWS memorizza ed elabora i contenuti esclusivamente nelle Regioni selezionate dal cliente, mantenendone la localizzazione, salvo nei casi previsti dalla normativa vigente.

Nella Figura 2.4 vengono mostrate le regioni AWS.

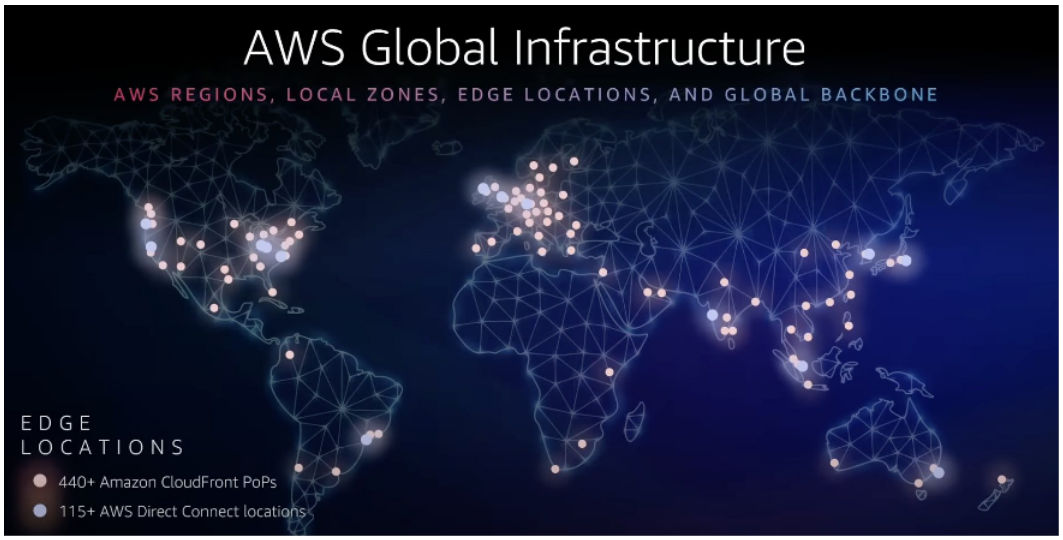


Figura 2.4: Infrastruttura globale (fonte: Amazon).

La connessione tra le diverse regioni è garantita dalla rete globale AWS, che si estende per oltre 20 milioni di chilometri di cavi in *fibra ottica*[4], contribuendo a ridurre la *latenza* [5]e a migliorare le prestazioni delle applicazioni.

Nella Figura 2.5 viene riportato un grafico di come sono distribuite le fibre ottiche.

Le Availability Zone sono costituite da uno o più data center dotati di sistemi indipendenti di alimentazione, raffreddamento, rete e sicurezza fisica. Ogni AZ è progettata per operare in modo autonomo, pur rimanendo interconnessa alle altre Availability Zone della stessa Regione tramite una rete dedicata, ridondante, ad alta larghezza di banda e a bassa latenza.

[4]La *fibra ottica* è una tecnologia di trasmissione dati che utilizza filamenti di vetro o plastica per trasportare informazioni tramite impulsi di luce (definizione presa da Windtre).

[5]La *latenza* è il tempo di ritardo che intercorre tra una richiesta e la relativa risposta di un sistema. La latenza di rete indica il tempo necessario affinché i dati vengano trasmessi da un punto all'altro di una rete (definizione presa da IBM).

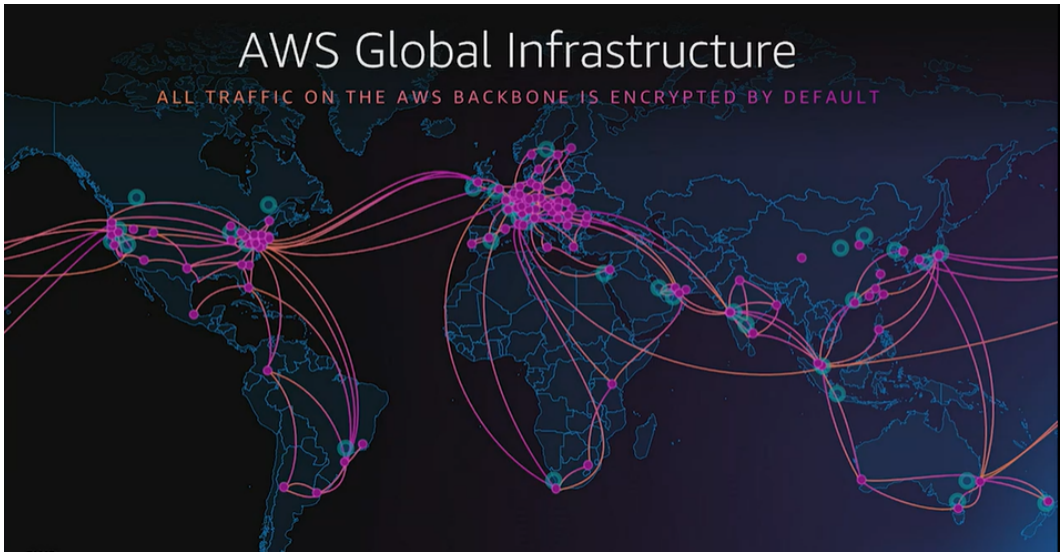


Figura 2.5: Fibre ottiche (fonte: Linkedin).

Inoltre, tutto il traffico tra le Availability Zone è *crittografato*[6], consentendo di realizzare architetture altamente disponibili e resilienti, in grado di garantire la continuità operativa anche in caso di guasti localizzati.

Per soddisfare le esigenze delle applicazioni che richiedono tempi di risposta molto ridotti, AWS mette a disposizione le Local Zone, estensioni delle Regioni che portano servizi di elaborazione, archiviazione, database e rete in aree geografiche più vicine agli utenti finali. Le Local Zone consentono di ridurre la latenza mantenendo un collegamento sicuro e ad alta velocità con la Regione di appartenenza e permettono di utilizzare gli stessi strumenti, le stesse API e i medesimi servizi disponibili nell'infrastruttura AWS.

2.2.2 Architettura tecnica dei servizi

L'architettura dei servizi di Amazon Web Services (AWS) può essere descritta attraverso i principi dell'architettura orientata ai servizi (Service-Oriented Architecture, SOA), un modello di sviluppo software basato sull'impiego di componenti indipendenti, chiamati servizi, per la realizzazione di applicazioni aziendali. In questo modello ogni servizio svolge una funzione specifica ed è progettato per operare in modo autonomo rispetto agli altri, pur potendo comunicare con servizi sviluppati mediante piattaforme o linguaggi di programmazione differenti. La SOA consente di integrare servizi indipendenti per realizzare funzionalità più complesse e, a differenza delle architetture monolitiche, nelle quali i processi sono eseguiti come un'unica unità, favorisce la suddivisione delle applicazioni in componenti modulari e riutilizzabili.

I principi base dell'architettura SOA sono i seguenti:

- *Interoperabilità*; ogni servizio è descritto da documenti che ne definiscono funzionalità e modalità di utilizzo. I servizi possono comunicare tra loro anche se sviluppati con piattaforme o linguaggi di programmazione differenti.
- *Accoppiamento debole*; i servizi sono progettati per ridurre al minimo le dipendenze reciproche e le dipendenze da risorse esterne. Di conseguenza, le modifiche apportate

[6] La *crittografia* è la pratica di sviluppare e utilizzare algoritmi codificati per proteggere e oscurare le informazioni trasmesse, in modo che possano essere lette solo da coloro che hanno il permesso e la capacità di decifrarle (definizione presa da IBM).

a un servizio hanno un impatto limitato sulle applicazioni e sugli altri servizi che lo utilizzano.

- *Astrazione*; consente agli utenti di utilizzare un servizio senza conoscerne l'implementazione interna. Le modalità di utilizzo sono descritte nella documentazione e nei contratti di servizio.
- *Granularità*; prevede che ogni servizio svolga una funzione specifica e possa essere combinato con altri servizi per realizzare funzionalità più complesse.

Nell'architettura orientata ai servizi (SOA), i servizi operano in modo indipendente e forniscono specifiche funzionalità, oppure consentono lo scambio di dati tra i diversi componenti dell'applicazione. Quando un utente o un'applicazione invia una richiesta, il servizio elabora i dati, esegue l'operazione richiesta e restituisce una risposta. La comunicazione tra i servizi avviene attraverso *protocolli di comunicazione*[7], ossia un insieme di regole che disciplinano la trasmissione dei dati sulla rete.

Dall'evoluzione dell'architettura SOA deriva l'architettura a microservizi, sviluppata per rispondere alle esigenze delle moderne applicazioni cloud. A differenza della SOA, in cui un servizio può rappresentare un'intera funzionalità aziendale, nei microservizi ogni componente è progettato per svolgere un compito specifico e operare in modo autonomo. Questa architettura garantisce maggiore flessibilità, scalabilità e semplicità nella gestione delle applicazioni. Per queste caratteristiche, l'architettura a microservizi rappresenta il modello di riferimento per lo sviluppo delle moderne applicazioni cloud su AWS.

La Figura 2.6 illustra la differenza fra le architetture monolitiche e i microservizi.

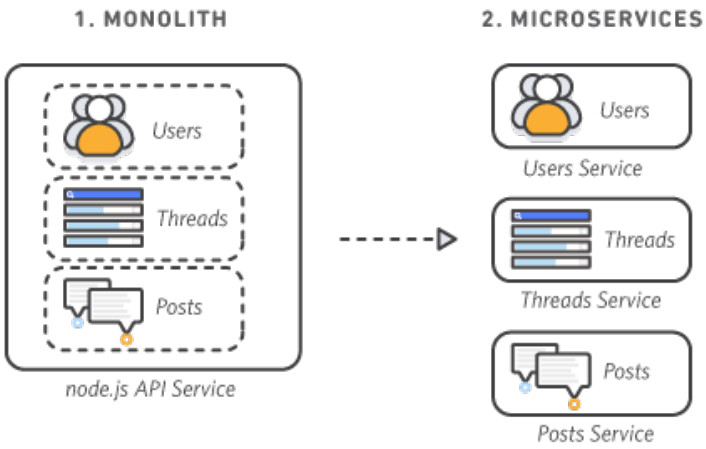


Figura 2.6: Microservizi (fonte: Amazon).

Per supportare la progettazione di queste architetture cloud, AWS mette a disposizione il framework AWS Well-Architected, costituito da linee guida che aiutano gli architetti del cloud a realizzare infrastrutture sicure, affidabili, resilienti, ad alte prestazioni ed economicamente sostenibili. Il framework si basa su sei pilastri fondamentali:

1. *eccellenza operativa*: esecuzione e miglioramento continuo dei processi e delle operazioni;

[7] La *comunicazione di rete*, o interconnessione di reti, definisce un insieme di protocolli (ovvero regole e standard) che consentono ai programmi applicativi di comunicare tra loro indipendentemente dall'hardware e dai sistemi operativi su cui vengono eseguiti. L'interconnessione di reti permette ai programmi applicativi di comunicare a prescindere dalle loro connessioni di rete fisiche (definizione presa da IBM).

2. *sicurezza*: protezione di dati, sistemi e identità;
3. *affidabilità*: capacità di mantenere il corretto funzionamento e di ripristinarsi in caso di guasto;
4. *efficienza delle prestazioni*: utilizzo ottimale delle risorse di calcolo;
5. *ottimizzazione dei costi*: controllo della spesa eliminando gli sprechi;
6. *sostenibilità*: riduzione dell'impatto ambientale delle infrastrutture cloud.

2.2.3 Accesso e interfaccia utente

Per accedere e utilizzare i servizi messi a disposizione da AWS è necessario registrarsi al sito e creare un account tramite il quale è possibile usufruire dei diversi servizi offerti da Amazon Web Services. AWS mette a disposizione strumenti per la gestione dell'identità, delle autorizzazioni e delle credenziali di sicurezza.

Per la gestione dell'identità e dell'accesso (IAM) all'ambiente AWS possono essere utilizzati diversi tipi di interfacce; la più comune è la AWS Management Console, un'interfaccia web che consente di svolgere diverse attività amministrative, tra cui la creazione di utenti e ruoli e la configurazione delle relative autorizzazioni. La pagina iniziale della AWS Management Console fornisce l'accesso alle diverse console dei servizi AWS.

Inoltre, AWS mette a disposizione due set di strumenti: AWS Command Line Interface e AWS Tools for Windows PowerShell, che consentono di eseguire comandi direttamente dal terminale e possono risultare più efficienti rispetto alla navigazione tramite AWS Management Console.

Per integrare le funzionalità IAM direttamente nelle applicazioni, AWS mette a disposizione kit di sviluppo software (SDK) per diversi linguaggi di programmazione.

2.3 Principali servizi e tool di AWS

AWS comprende numerosi servizi che consentono di utilizzare le risorse del cloud per svolgere differenti operazioni. Ogni servizio presenta caratteristiche e funzionalità specifiche e può essere utilizzato in base alle esigenze dell'applicazione. In questa sezione verranno analizzati alcuni dei principali servizi AWS, partendo da quelli dedicati allo storage e al calcolo, fino ad arrivare ai servizi di Intelligenza Artificiale e Machine Learning, alcuni dei quali verranno successivamente utilizzati nella parte sperimentale.

Nella Figura 2.7 sono presentati alcuni tool e servizi AWS.

2.3.1 Servizi di storage

Lo storage consente di archiviare e conservare i dati rendendoli disponibili quando necessario. Nel cloud computing, i dati vengono memorizzati su infrastrutture gestite da un provider cloud e possono essere resi accessibili attraverso una connessione Internet.

AWS offre numerosi servizi di storage affidabili, scalabili e sicuri per l'archiviazione, l'accesso, la protezione e l'analisi dei dati. I servizi disponibili comprendono soluzioni di storage a blocchi, file e oggetti, oltre a opzioni per la migrazione dei dati nel cloud. La scelta del servizio più adatto dipende dalle caratteristiche del carico di lavoro e dalle specifiche esigenze aziendali.

Le principali tipologie di storage descritte da AWS sono:

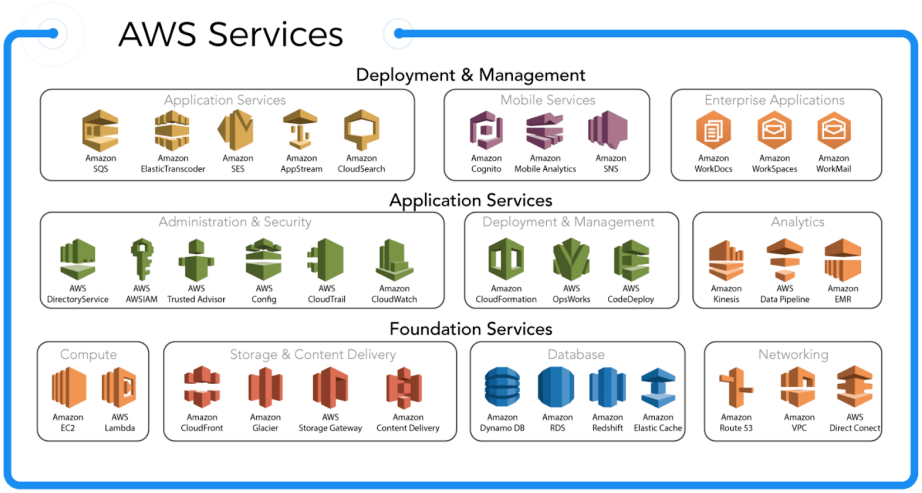


Figura 2.7: Alcuni tool e servizi AWS (fonte: Zerounoweb).

- *Block (storage a blocchi)*: lo storage a blocchi è una tecnologia di archiviazione che suddivide i dati in blocchi di dimensioni uniformi. I blocchi vengono memorizzati sullo storage fisico sottostante in modo da rendere più efficienti le operazioni di accesso e recupero dei dati.
- *File system (storage di file)*: i dati vengono organizzati secondo una struttura gerarchica di file e cartelle. Negli ambienti di rete, questa tipologia di storage può utilizzare la tecnologia *NAS (Network-Attached Storage)*[8], che consente di accedere ai dati in modo simile a un disco rigido locale.
- *Oggetti*: in questa tipologia di storage, i dati non strutturati vengono memorizzati sotto forma di oggetti. Ciascun oggetto è associato a un identificatore univoco e a metadati che forniscono informazioni sul relativo contenuto.
- *Cache*: è un livello di archiviazione dei dati ad alta velocità. Tramite la cache vengono archiviati temporaneamente i dati a cui si accede con maggiore frequenza o che sono stati utilizzati di recente, così da mantenerli più vicini al punto di accesso e da velocizzarne il recupero.
- *Hybrid/Edge (storage ibrido/edge)*: combina l'infrastruttura di archiviazione locale con i servizi di storage cloud, permettendo lo spostamento dei dati tra i due ambienti. Tra i principali vantaggi rientrano l'accesso a bassa latenza, l'ottimizzazione dei costi, la sovranità dei dati, la scalabilità del cloud e la continuità aziendale.

La Figura 2.8 mostra alcuni dei principali servizi di storage messi a disposizione da AWS e la loro suddivisione in base alle diverse modalità di archiviazione dei dati.

Fra i principali servizi di storage offerti da AWS rientra Amazon Simple Storage Service (Amazon S3), un servizio di archiviazione a oggetti che consente di archiviare e recuperare qualsiasi quantità di dati. S3 permette di gestire, analizzare e proteggere grandi quantità di dati ed è caratterizzato da scalabilità, disponibilità, sicurezza e prestazioni elevate. In particolare, si ha:

[8] Il *Network Attached Storage (NAS)* è un server centralizzato che consente a più utenti di memorizzare e condividere file su una rete Transmission Control Protocol/Internet Protocol (TCP/IP) tramite Wi-Fi o un cavo Ethernet (definizione presa da IBM).

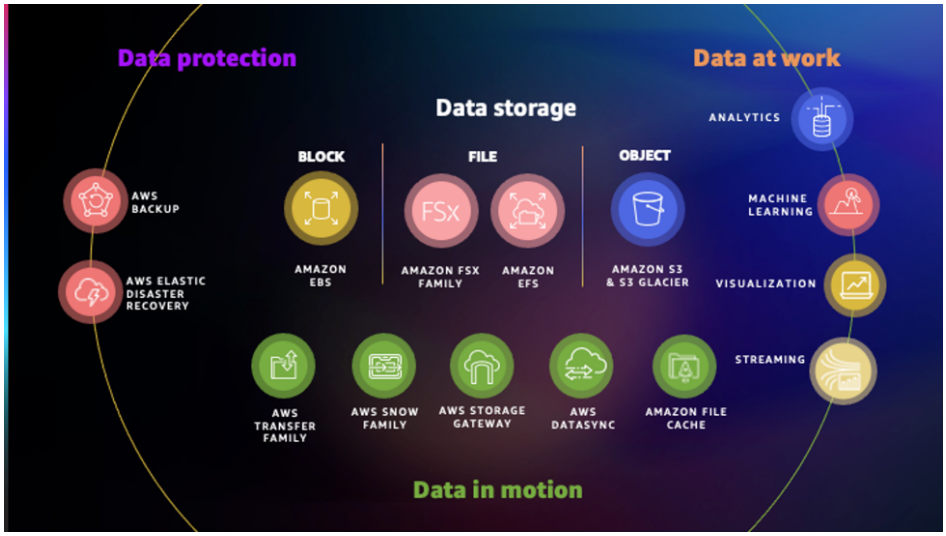


Figura 2.8: Servizi di storage di AWS (fonte: Amazon).

1. *Scalabilità*: con S3 è possibile archiviare qualsiasi quantità di dati; il servizio è completamente elastico e la capacità di storage aumenta o diminuisce automaticamente in base alla quantità di dati aggiunti o rimossi.

2. *Durabilità e disponibilità*: S3 è progettato per garantire una durabilità dei dati del 99,999999999% e una disponibilità del 99,99% per impostazione predefinita.

3. *Sicurezza e protezione dei dati*: S3 è progettato per essere sicuro, privato e crittografato per impostazione predefinita. Inoltre, S3 dispone di funzionalità di audit per il monitoraggio delle richieste di accesso alle risorse.

Amazon Elastic Block Store (Amazon EBS) è un servizio di storage a blocchi che consente di creare volumi di archiviazione collegabili alle istanze *Amazon EC2*. I volumi possono essere utilizzati per creare file system, eseguire database o archiviare direttamente dati a blocchi. Ogni volume viene eseguito all'interno di una specifica Availability Zone (AZ), dove i dati vengono replicati per proteggerli da eventuali guasti delle singole componenti. EBS supporta, inoltre, la creazione di snapshot per la protezione dei dati. Il servizio è caratterizzato da scalabilità, prestazioni elevate e funzionalità dedicate alla sicurezza e alla protezione dei dati.

AWS offre ulteriori servizi di storage, tra cui *Amazon Elastic File System (EFS)*, un file system serverless ed elastico che consente di condividere i dati senza dover gestire l'infrastruttura di archiviazione; *Amazon FSx*, un servizio che gestisce completamente l'archiviazione dei dati e supporta file system commerciali e open source; *Amazon File Cache*, una cache ad alta velocità per set di dati archiviati, utilizzata per accelerare i carichi di lavoro nel cloud; e *AWS Storage Gateway*, un servizio di storage cloud ibrido che consente l'accesso da ambienti *on-premises*[9]allo storage nel cloud.

2.3.2 Servizi di calcolo

Il calcolo, nel cloud computing, è un termine che viene utilizzato per indicare l'insieme delle risorse necessarie per l'esecuzione dei programmi, tra cui la potenza di elaborazione, la memoria, le reti e l'archiviazione, fondamentali per l'esecuzione delle operazioni di calcolo software.

[9]Una *licenza on-premises* è una licenza software installata e utilizzata a livello locale. Definizione presa da Thalesgroup.

I servizi di calcolo rientrano nel modello Infrastructure-as-a-Service (IaaS), che permette agli utenti di utilizzare e gestire risorse informatiche virtualizzate in base alle proprie esigenze. Per ciascuna risorsa di calcolo è possibile determinare la quantità di potenza necessaria allo svolgimento delle operazioni. Tra le principali risorse di calcolo ci sono la CPU e la memoria.

Amazon EC2 rappresenta uno dei principali servizi di calcolo AWS. EC2 fornisce diversi tipi di istanze, ovvero *macchine virtuali*[10], con differenti configurazioni, permettendo agli utenti di personalizzare le risorse di calcolo in base alle esigenze dell'applicazione.

Esistono cinque tipologie di istanze:

1. *istanze per uso generico*: permettono di utilizzare in maniera bilanciata capacità di elaborazione, memoria e rete, adattandosi a differenti tipologie di utilizzo;
2. *istanze ottimizzate per il calcolo*: sono utilizzate per applicazioni che necessitano di un'elevata capacità di elaborazione e di prestazioni di rete elevate;
3. *istanze di calcolo accelerate*: utilizzano componenti hardware specializzati, chiamati acceleratori, che permettono di eseguire più velocemente alcune funzioni generalmente svolte dalla CPU;
4. *istanze ottimizzate per la memoria*: sono progettate per applicazioni che richiedono una maggiore quantità di memoria e permettono un accesso rapido ai dati;
5. *istanze ottimizzate per l'archiviazione*: vengono utilizzate per l'archiviazione locale e servono per migliorare le prestazioni nelle operazioni di lettura e scrittura.

Oltre ai servizi basati sull'utilizzo di istanze, AWS fornisce anche servizi *serverless*[11], come AWS Lambda, che permette di adattare automaticamente le risorse alle necessità dell'applicazione, garantendo un'elevata disponibilità e prevedendo un modello di fatturazione *pay-for-value*.

Inoltre, AWS mette a disposizione risorse per il *bilanciamento del carico di lavoro*[12]. Un esempio è Elastic Load Balancing, il quale consente agli utenti di migliorare le prestazioni e l'affidabilità delle applicazioni.

Tra i servizi di calcolo AWS rientra anche Amazon Elastic Container Service (ECS), attraverso il quale è possibile gestire ed eseguire applicazioni basate su container *Docker*[13], utilizzando cluster di istanze Amazon EC2.

La Figura 2.9 riporta una panoramica dei principali servizi di calcolo messi a disposizione da AWS, suddivisi in base alle diverse categorie di utilizzo.

2.3.3 Servizi di Intelligenza Artificiale e Machine Learning

AWS fornisce numerosi servizi di Intelligenza Artificiale e Machine Learning. In particolare, AWS mette a disposizione servizi basati su modelli pre-addestrati, ovvero modelli di Machine Learning precedentemente addestrati su specifici set di dati per svolgere determinati compiti. Inoltre, alcuni servizi di IA AWS consentono di accedere, attraverso specifiche API, a funzionalità basate su tecniche di Machine Learning e Deep Learning.

[10] Una *macchina virtuale* è un software in grado di eseguire tutte le funzioni di un computer fisico, inclusi le applicazioni in esecuzione e i sistemi operativi (definizione presa da Amazon).

[11] Con *calcolo serverless* si intende lo sviluppo di applicazioni con un'infrastruttura server sottostante gestita esternamente (definizione presa da Amazon).

[12] *Il bilanciamento del carico* è il processo di distribuzione omogenea delle risorse di calcolo e del carico di lavoro in un ambiente di cloud computing (definizione presa da Amazon).

[13] *Docker* fornisce un modo standard per creare pacchetti di codice di un'applicazione ed eseguirli in qualsiasi sistema (definizione presa da Amazon).

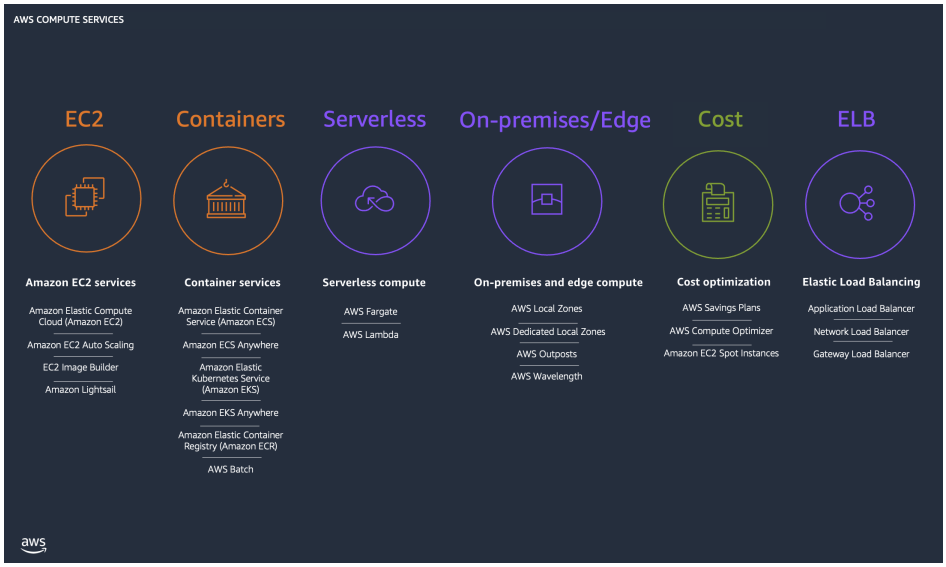


Figura 2.9: Servizi di calcolo AWS (fonte: Amazon).

Nella Figura 2.10 sono rappresentati alcuni dei principali servizi di Intelligenza Artificiale offerti da AWS.



Figura 2.10: Servizi di AI AWS (fonte: Itmagic).

I servizi di Intelligenza Artificiale offerti da AWS riguardano diversi ambiti, tra cui il linguaggio, l'estrazione e l'analisi dei dati, la visione artificiale, i servizi specializzati, il codice e DevOps e il settore sanitario. Di seguito vengono presentati alcuni dei principali servizi appartenenti a tali categorie, con particolare attenzione ad Amazon Comprehend e Amazon Rekognition, utilizzati nelle sperimentazioni descritte nei capitoli successivi.

Per quanto riguarda il linguaggio, i servizi possono essere distinti in base alla loro funzione, tra cui sintesi vocale, riconoscimento vocale, interazione tramite chatbot e traduzione. In particolare, AWS mette a disposizione Amazon Polly, Amazon Transcribe, Amazon Lex e Amazon Translate. Sebbene siano tutti legati all'elaborazione del linguaggio, tali servizi svolgono funzioni differenti:

- *Amazon Polly* è un servizio di IA che permette di trasformare un testo in contenuto vocale;
- *Amazon Transcribe* utilizza tecniche di Deep Learning per il riconoscimento vocale automatico (Automatic Speech Recognition, ASR), attraverso le quali il parlato viene convertito in testo;
- *Amazon Lex* consente di creare interfacce conversazionali basate sulla voce e sul testo;
- *Amazon Translate* utilizza tecniche di traduzione automatica neurale per tradurre contenuti in diverse lingue.

Nell'ambito dell'estrazione e dell'analisi dei dati, AWS mette a disposizione diversi servizi, tra cui Amazon Textract, per l'estrazione di testo e dati, Amazon Comprehend, per l'elaborazione del linguaggio naturale, e Amazon Augmented AI (A2I), per l'integrazione della revisione umana nei flussi di lavoro di Machine Learning. In particolare:

- *Amazon Textract* è un servizio utilizzato per estrarre testo, scrittura a mano e dati da documenti.
- *Amazon Comprehend* è un servizio di elaborazione del linguaggio naturale che utilizza il Machine Learning per ricavare informazioni dai testi.
- *Amazon Augmented AI (A2I)* facilita l'integrazione della revisione umana nei flussi di lavoro basati sul Machine Learning.

Amazon Comprehend è un servizio di elaborazione del linguaggio naturale (NLP) che utilizza il Machine Learning per ricavare informazioni dai testi. Il servizio consente, attraverso specifiche API, di svolgere diverse attività, tra cui la classificazione dei contenuti, il riconoscimento delle entità, l'estrazione di frasi chiave e l'analisi del sentiment. Grazie alle funzionalità di estrazione e classificazione, il servizio può elaborare un'ampia gamma di documenti ed estrarne informazioni. La sperimentazione si concentrerà sull'analisi del sentiment, attraverso la quale è possibile individuare il sentimento espresso all'interno di un testo, classificandolo, ad esempio, come positivo, negativo, neutro o misto.

Per quanto riguarda la visione artificiale, Amazon Rekognition consente di effettuare analisi automatiche di immagini e video attraverso tecniche di Deep Learning. Il servizio mette a disposizione diverse funzionalità, tra cui l'analisi dei volti, il confronto facciale e l'individuazione di contenuti inappropriati.

Una delle funzionalità di Amazon Rekognition utilizzate nelle sperimentazioni riguarda il confronto facciale. Attraverso specifiche API è possibile confrontare i volti presenti nelle immagini e determinarne il livello di somiglianza. Tra le API utilizzate vi sono CompareFaces, che consente di confrontare un volto presente in un'immagine di origine con i volti rilevati in un'immagine di destinazione, e le API di ricerca all'interno delle raccolte di volti, che consentono di confrontare un volto con quelli precedentemente memorizzati in una collection. In base al livello di somiglianza viene restituito un punteggio compreso tra 0 e 100, dove valori più elevati indicano una maggiore somiglianza tra i volti analizzati.

Un'altra funzionalità di Amazon Rekognition utilizzata nelle sperimentazioni è la Content Moderation, che utilizza il Machine Learning per supportare l'analisi automatica di immagini e video e individuare contenuti potenzialmente inappropriati o indesiderati attraverso specifiche API e criteri di moderazione.

Per quanto riguarda i servizi specializzati, AWS mette a disposizione soluzioni di Intelligenza Artificiale progettate per rispondere a specifiche esigenze applicative. Tra queste rientrano Amazon Kendra, un servizio di ricerca intelligente che consente di individuare

informazioni all'interno di diverse fonti e repository di contenuti, e Amazon Personalize, che utilizza tecniche di Machine Learning per creare sistemi di raccomandazione e fornire agli utenti contenuti e suggerimenti personalizzati.

Nell'ambito del codice e DevOps, AWS mette a disposizione servizi di Intelligenza Artificiale a supporto dello sviluppo software e della gestione delle applicazioni. Tra questi, Amazon DevOps Guru utilizza il Machine Learning per individuare comportamenti operativi anomali e supportare l'identificazione di possibili problemi; Amazon Q Developer è un assistente basato sull'IA generativa che supporta gli sviluppatori nelle attività di sviluppo software, mentre Amazon Bedrock mette a disposizione *modelli di fondazione*[14] e strumenti per lo sviluppo di applicazioni di IA generativa e agenti.

Infine, nel settore sanitario, AWS mette a disposizione AWS HealthLake, un servizio progettato per archiviare, trasformare e analizzare dati sanitari. Il servizio utilizza API basate sullo standard FHIR (Fast Healthcare Interoperability Resources) e consente alle organizzazioni sanitarie di gestire e analizzare informazioni relative a pazienti e popolazioni.

[14] *I modelli di fondazione* sono un tipo di Intelligenza Artificiale generativa (IA generativa). Essi generano output da uno o più input (prompt) sotto forma di istruzioni in linguaggio umano. I modelli si basano su reti neurali complesse, tra cui reti generative avversarie (GAN), trasformer ed encoder variazionali (definizione presa da Amazon).

CAPITOLO 3

Un case study relativo alla moderazione dei contenuti

In questo capitolo vengono descritte la progettazione e l'implementazione di un case study relativo alla moderazione automatica dei contenuti mediante l'utilizzo di Amazon Rekognition. In particolare, vengono analizzati contenuti caratterizzati da differenti livelli di rischio, al fine di valutare il comportamento del sistema nell'individuazione di elementi sensibili o potenzialmente dannosi. Successivamente, vengono approfondite le politiche di Content Moderation adottate dalle piattaforme social Meta e TikTok, evidenziando le principali modalità attraverso cui vengono gestiti i contenuti problematici e le situazioni caratterizzate da rischio implicito. Infine, viene effettuato un confronto tra i risultati ottenuti tramite Amazon Rekognition e gli approcci alla moderazione previsti dalle linee guida di Meta e TikTok, con particolare attenzione al ruolo del contesto e ai limiti dell'analisi basata esclusivamente sugli elementi visivi.

3.1 Obiettivi e progettazione del case study

Per l'implementazione della fase di analisi automatica è stato utilizzato *Amazon Rekognition*, un servizio di *Amazon Web Services* che consente di identificare automaticamente categorie e caratteristiche presenti nei contenuti multimediali; utilizzando questo servizio è possibile analizzare immagini e video per mezzo di tecniche di Intelligenza Artificiale. Il sistema è in grado di riconoscere oggetti, volti e scene all'interno di contenuti multimediali.

Tra le funzionalità messe a disposizione da Amazon Rekognition, sono state utilizzate quella relativa alla *Content Moderation*, che rappresenta il processo di controllo e gestione dei contenuti pubblicati dagli utenti su piattaforme online, come i social network e i siti web. L'obiettivo è garantire che i contenuti rispettino le linee guida della piattaforma, evitando la diffusione di materiale dannoso o inappropriato. La moderazione automatica utilizza algoritmi di Intelligenza Artificiale per analizzare i contenuti. Nel caso di immagini e video, vengono utilizzati sistemi basati su tecniche di computer vision per identificare elementi potenzialmente sensibili.

I contenuti sottoposti a moderazione possono appartenere a diverse categorie, tra cui contenuti violenti, espliciti e presenza di oggetti potenzialmente pericolosi. Ai fini della presente sperimentazione, l'analisi è stata concentrata principalmente sulla presenza di violenza, contenuti espliciti e oggetti sensibili, come le armi.

Come case study è stato scelto il contesto delle piattaforme social di *Meta* e *TikTok*, attraverso le quali gli utenti pubblicano contenuti multimediali sottoposti a processi di moderazione

finalizzati a verificarne la conformità alle linee guida delle piattaforme. L'obiettivo del case study è quello di analizzare il comportamento della funzionalità di Content Moderation di Amazon Rekognition nell'analisi di immagini caratterizzate da differenti tipologie di contenuto.

Lo studio consiste nel simulare il funzionamento dei sistemi di moderazione automatica adottati dalle principali piattaforme social. Il sistema riceve in input immagini pubblicate da un ipotetico utente e le analizza automaticamente per verificare che rispettino le linee guida.

Ai fini del case study, i risultati restituiti dal servizio vengono ricondotti a una classificazione semplificata in due categorie: *Safe* e *Unsafe*.

Il funzionamento del sistema può essere rappresentato mediante una *pipeline* di lavoro suddivisa nelle seguenti fasi:

- acquisizione dell'immagine caricata dall'utente;
- analisi automatica del contenuto;
- individuazione di eventuali elementi critici;
- classificazione finale del contenuto;
- decisione di accettazione o rifiuto del post.

3.2 Sperimentazione con Amazon Rekognition

Per la sperimentazione del case study sono stati presi in considerazione tre casi, due espliciti e uno implicito:

- un contenuto *Safe*, chiaramente privo di elementi problematici;
- un contenuto *Unsafe*, caratterizzato dalla presenza evidente di elementi sensibili;
- un contenuto *Borderline*, selezionato con l'obiettivo di verificare se il sistema sia in grado di distinguere una situazione realmente pericolosa da una situazione apparentemente rischiosa, ma svolta in condizioni di sicurezza.

3.2.1 Analisi di un contenuto Safe

Per la prima sperimentazione è stata utilizzata in input l'immagine di un paesaggio naturale, raffigurante un tramonto sul mare e priva di elementi riconducibili a contenuti sensibili o potenzialmente dannosi. Successivamente, sono stati analizzati gli output restituiti da Amazon Rekognition al fine di verificare la presenza di eventuali etichette di moderazione. Come risultato ci si aspetta che non venga rilevata alcuna etichetta di moderazione e che il contenuto venga classificato come *Safe*.

Nella Figura 3.1 viene riportata l'immagine utilizzata per la prima sperimentazione.

Il contenuto è stato sottoposto alle funzionalità di Content Moderation di Amazon Rekognition.

Nella Figura 3.2 viene riportato il risultato della prima sperimentazione.

Come mostrato, il sistema di Content Moderation non ha rilevato etichette di moderazione. Tale risultato è confermato dalla risposta restituita dal servizio, nella quale il campo *ModerationLabels* risulta vuoto.

Sulla base dell'assenza di etichette di moderazione, il contenuto viene considerato come *Safe*. Il risultato ottenuto è coerente con le aspettative, poiché nell'immagine analizzata non sono presenti elementi riconducibili alle categorie considerate come "*pericolose*" nella sperimentazione, quali contenuti violenti, espliciti o oggetti potenzialmente pericolosi.



Figura 3.1: Tramonto (fonte: Canva).

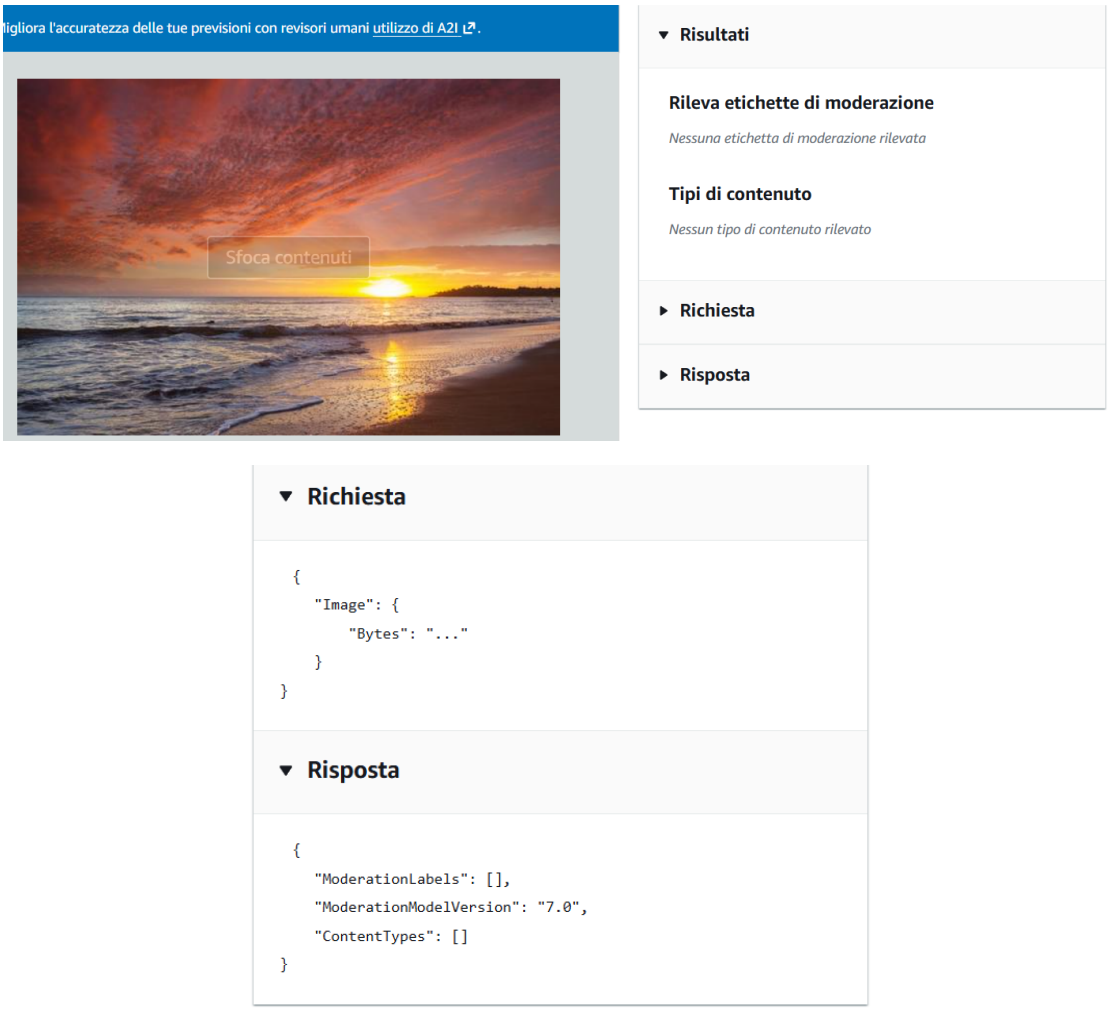


Figura 3.2: Risultato della prima sperimentazione (Svolta su: Amazon Web Services)

3.2.2 Analisi di un contenuto Unsafe

Per la seconda sperimentazione è stata utilizzata in input l'immagine raffigurante due militari in uno scenario di guerra, caratterizzata da una ricca presenza di contenuti sensibili o potenzialmente dannosi, come le armi. Successivamente, sono stati analizzati gli output restituiti da Amazon Rekognition al fine di verificare la presenza di etichette di moderazione. Come risultato, ci si aspetta un output caratterizzato dalla presenza di diverse etichette di moderazione e che il contenuto venga classificato come Unsafe.

Nella Figura 3.3 viene riportata l'immagine utilizzata per la seconda sperimentazione.



Figura 3.3: Militari (fonte: Rid).

Il contenuto è stato sottoposto alle funzionalità di Content Moderation di Amazon Rekognition.

Nella Figura 3.4 viene riportata la prima parte del risultato della seconda sperimentazione.

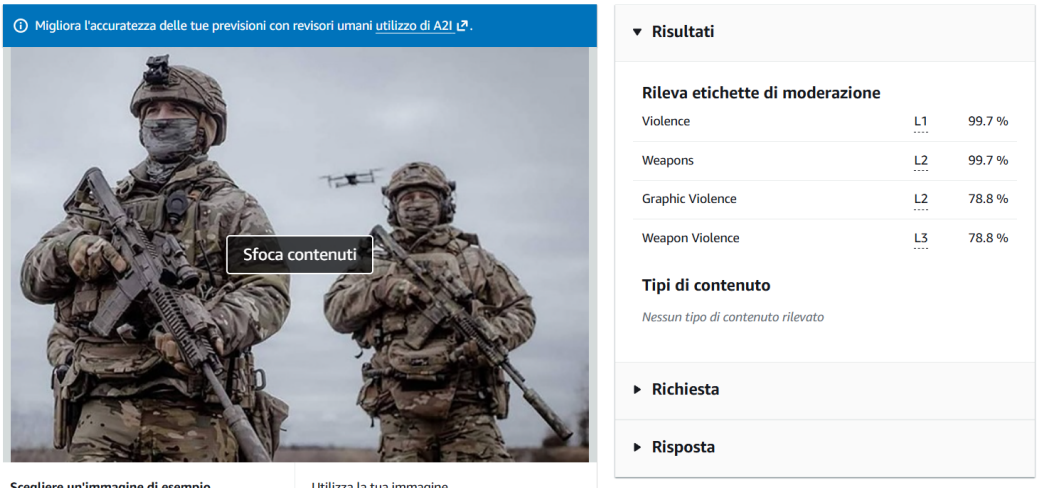


Figura 3.4: Prima parte del risultato della seconda sperimentazione (Svolta su: Amazon Web Services)

Nella Figura 3.5 viene riportata la seconda parte del risultato della seconda sperimentazione.

Come mostrato, il sistema di Content Moderation ha rilevato diverse etichette riconducibili a contenuti sensibili o potenzialmente dannosi. In particolare, sono state individuate le categorie Violence, Weapons e Graphic Violence, con valori di confidenza elevati, superiori al 78%.



Figura 3.5: Seconda parte del risultato della seconda sperimentazione (Svolta su: Amazon Web Services)

Sulla base delle etichette di moderazione rilevate, il contenuto viene classificato come *Unsafe*. Il risultato ottenuto è coerente con le aspettative, poiché l'immagine analizzata presenta elementi riconducibili a contenuti violenti e potenzialmente pericolosi, in particolare per la presenza delle armi.

3.2.3 Analisi di un contenuto Borderline

Per la terza sperimentazione è stata utilizzata in input un'immagine raffigurante un'attività di immersione in gabbia per l'osservazione degli squali. L'immagine non presenta contenuti esplicitamente dannosi; tuttavia, la situazione rappresentata può essere considerata potenzialmente pericolosa. Successivamente, sono stati analizzati gli output restituiti da Amazon Rekognition al fine di verificare la presenza di etichette di moderazione. Considerando il contenuto nel contesto di una piattaforma social, ci si potrebbe aspettare che questo venga pubblicato, ma che il rischio associato alla scena venga comunque preso in considerazione dal sistema di moderazione attraverso l'adozione di misure precauzionali, come l'applicazione di avvisi o eventuali limitazioni alla distribuzione.

Nella Figura 3.6 viene riportata l'immagine utilizzata per la terza sperimentazione.

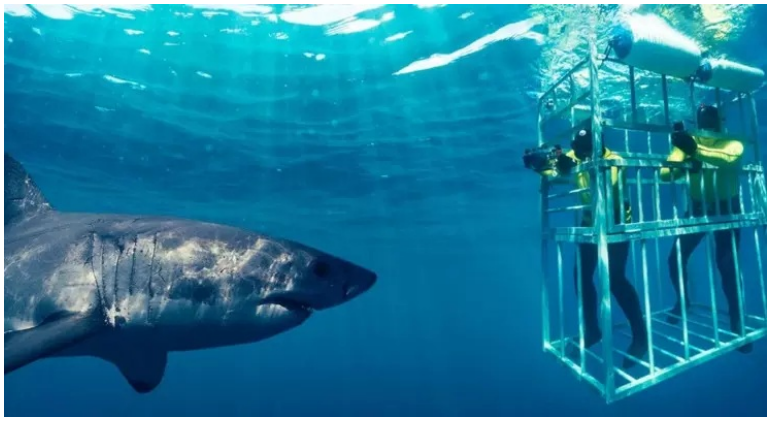


Figura 3.6: Squalo (fonte: Veratour).

Il contenuto è stato sottoposto alle funzionalità di Content Moderation di Amazon Rekognition.

Nella Figura 3.7 viene riportato il risultato della terza sperimentazione.

La scena rappresenta una situazione potenzialmente pericolosa, data la presenza di un animale predatore in prossimità dell'uomo; tuttavia, il rischio risulta mitigato dal fatto che i soggetti siano situati all'interno di una gabbia, progettata per garantire una separazione fisica e, quindi, una condizione di sicurezza controllata.

Dal punto di vista della Content Moderation, l'immagine non presenta elementi riconducibili a categorie critiche, quali *Violence*, *Weapons* o *Graphic Violence*, in quanto non sono osservabili né atti aggressivi né conseguenze lesive esplicite.

La scena può essere interpretata come un *caso borderline*, in cui è presente un rischio implicito, ma non una manifestazione diretta di violenza. Considerando il contenuto nel contesto di una piattaforma social, ci si potrebbe aspettare che la situazione di pericolo rappresentata venga comunque presa in considerazione dal sistema di moderazione, attraverso l'individuazione del rischio associato alla scena o l'applicazione di eventuali limitazioni al contenuto.

Nel caso di Amazon Rekognition, tuttavia, non viene rilevata alcuna etichetta di moderazione. Sulla base del risultato ottenuto, è possibile osservare come, nel caso analizzato, il

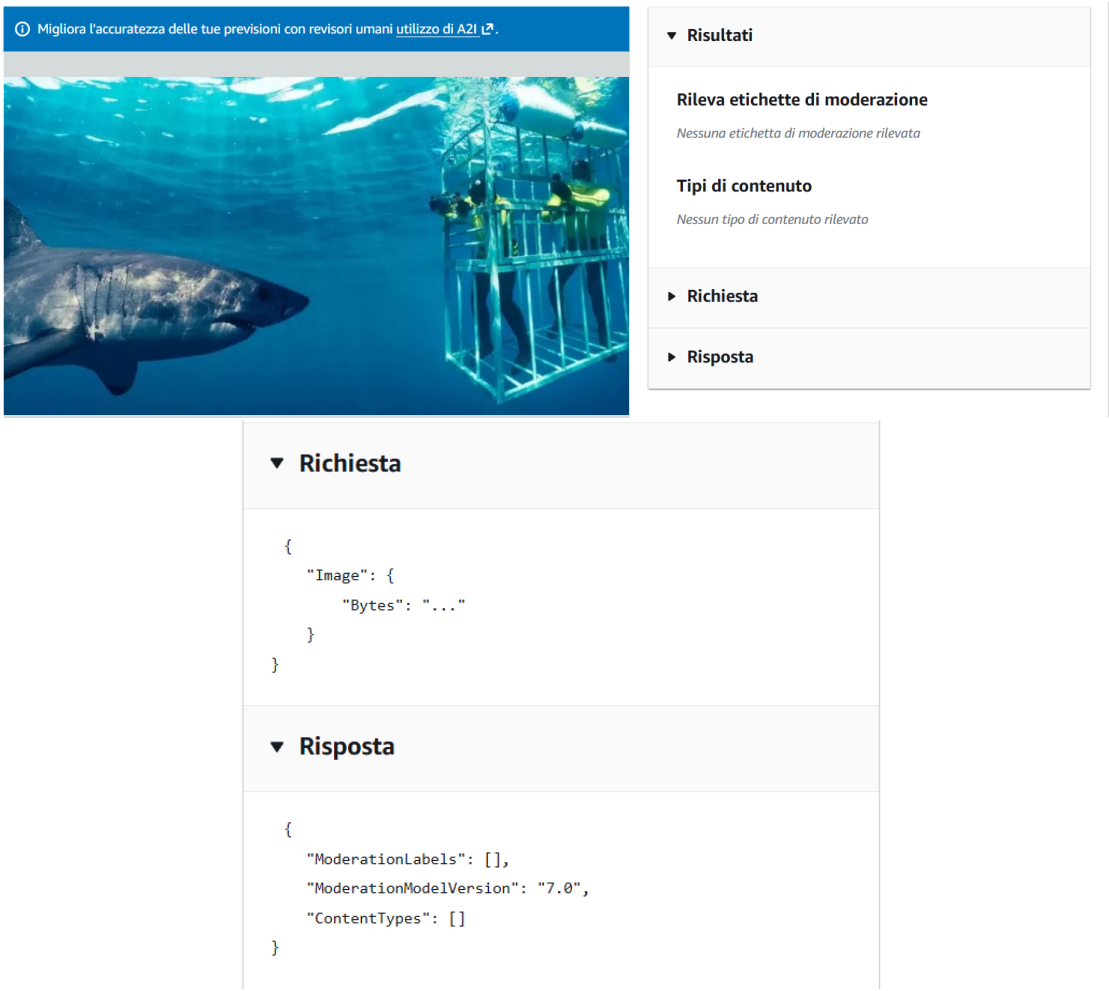


Figura 3.7: Risultato della terza sperimentazione (Svolta su: Amazon Web Services)

sistema di Content Moderation di Amazon Rekognition non rilevi la situazione potenzialmente pericolosa, poiché il rischio è implicito e non sono presenti elementi esplicitamente riconducibili alle categorie di moderazione individuate dal sistema.

A partire da tale risultato, è possibile svolgere un confronto tra il comportamento di Amazon Rekognition e i sistemi di moderazione adottati dalle principali piattaforme social, le cui linee guida possono considerare anche il contesto e il potenziale rischio associato ai contenuti.

3.3 Content Moderation nelle piattaforme social

Considerando i vari *Case Study* più strettamente nel contesto dei *Social Media*, piattaforme come Meta e TikTok adottano linee guida di moderazione più dettagliate e articolate. Tali linee guida non si limitano alla gestione di contenuti esplicitamente violenti o pericolosi, ma prevedono anche una valutazione del contesto e del potenziale rischio associato ai contenuti.

Di conseguenza, l'ultima situazione analizzata, pur non contenendo elementi di violenza esplicita, potrebbe essere considerata una rappresentazione di un'attività potenzialmente pericolosa o rischiosa e, pertanto, essere *soggetta a limitazioni* piuttosto che a una rimozione completa.

3.3.1 Politiche di moderazione di Meta

La piattaforma Meta rappresenta una delle più importanti aziende operanti nel settore dei social media. I social network che essa possiede, tra cui Facebook e Instagram, sono gestiti attraverso linee guida comuni. Dal 2016, infatti, queste piattaforme adottano una strategia basata sulla rimozione dei contenuti che violano le normative sulla riduzione della diffusione di contenuti problematici che non necessariamente vengono rimossi e sulla diffusione di informazioni aggiuntive che permettano agli utenti di comprendere meglio ciò che visualizzano. In alcuni casi, inoltre, determinati contenuti possono essere accompagnati da avvisi o essere accessibili soltanto a specifiche categorie di utenti, ad esempio ai maggiorenni.

Per garantire la sicurezza degli utenti, Meta adotta un sistema di Content Moderation basato su linee guida ufficiali, note come *Meta Community Standards*, che definiscono quali contenuti sono consentiti e quali devono essere rimossi o limitati.

I contenuti possono, quindi, essere distinti nelle seguenti categorie:

- contenuti non consentiti;
- contenuti consentiti che, però, richiedono informazioni aggiuntive per comprenderne il contesto e, sulla base di quest'ultimo, adottare i provvedimenti adatti.

Meta utilizza modelli di apprendimento automatico sviluppati dai propri team di Intelligenza Artificiale per analizzare differenti tipologie di contenuto, come immagini e testi. Tali modelli vengono successivamente impiegati dai team responsabili dell'integrità della piattaforma per individuare possibili violazioni delle normative e formulare previsioni sui contenuti pubblicati. Sulla base dei risultati ottenuti, il sistema può decidere di applicare differenti provvedimenti, tra cui la rimozione del contenuto, la penalizzazione, oppure l'invio a un team umano per ulteriori verifiche. I modelli vengono, inoltre, addestrati a riconoscere segnali specifici, come la presenza di nudità nelle immagini o elementi problematici nei testi. Il sistema di moderazione combina, quindi, tecniche di Intelligenza Artificiale con l'intervento di team addetti al controllo umano per ulteriori verifiche.

Rispetto all'approccio osservato nella sperimentazione condotta con Amazon Rekognition, le politiche di Meta prevedono una gestione più articolata dei contenuti, nella quale può assumere rilevanza anche il contesto in cui essi vengono presentati.

Nel caso analizzato in precedenza, relativo alla presenza di uno squalo in prossimità di due individui protetti da una gabbia, il contenuto non presenta violenza esplicita, ma rappresenta, comunque, una situazione potenzialmente pericolosa. Considerando le modalità di moderazione descritte da Meta, tale contenuto potrebbe richiedere una valutazione del contesto, senza comportare necessariamente una rimozione.

3.3.2 Politiche di moderazione di TikTok

TikTok è una piattaforma social basata sulla condivisione di contenuti video, caratterizzata da un'elevata diffusione tra utenti giovani e da una rapida circolazione dei contenuti. Le linee guida della community di TikTok risultano strutturate in diverse macro-categorie, ciascuna delle quali affronta specifici aspetti legati alla sicurezza e al benessere degli utenti.

Tra le principali categorie previste dalle linee guida figurano:

- principi della community;
- sicurezza e benessere dei più giovani;
- sicurezza e civiltà;
- salute mentale e comportamentale;

- temi sensibili e contenuti destinati a un pubblico maturo;
- integrità e autenticità;
- privacy e sicurezza.

Questa organizzazione evidenzia come TikTok adotti un approccio alla Content Moderation fortemente orientato alla tutela dell'utente, considerando non solo la presenza di contenuti esplicitamente dannosi, ma anche il loro potenziale impatto sul comportamento degli utenti e sul benessere della community.

Per raggiungere questo obiettivo, vengono adottate due azioni importanti:

- *Rimozione dei contenuti*: i contenuti che non rispettano le linee guida della community vengono rimossi dalla piattaforma.
- *Limitazione dell'accesso ai contenuti in base all'età*: alcuni contenuti, pur non essendo necessariamente rimossi, possono essere considerati non adatti agli utenti di età inferiore ai 18 anni e, di conseguenza, essere resi accessibili soltanto agli utenti maggiorenni.

Per garantire la sicurezza della community, TikTok adotta un sistema di Content Moderation basato sulle *TikTok Community Guidelines*. Il processo di moderazione combina tecnologie automatizzate con l'intervento di professionisti della sicurezza, al fine di identificare e gestire contenuti potenzialmente dannosi. Le tecnologie automatizzate analizzano i contenuti per individuare possibili violazioni delle politiche della piattaforma, eventuali restrizioni legate all'età o contenuti non idonei alla raccomandazione. La revisione umana continua, tuttavia, a svolgere un ruolo importante nel processo di moderazione.

TikTok affianca alla moderazione umana diverse tecnologie di Intelligenza Artificiale per l'analisi e il controllo dei contenuti, riportate di seguito:

- *Basate sulla visione*: i modelli di visione artificiale permettono di identificare elementi visivi che potrebbero violare le linee guida della community, come armi o simboli di odio.
- *Basate sull'audio*: i contenuti audio vengono analizzati attraverso sistemi in grado di individuare elementi simili o riconducibili a precedenti violazioni.
- *Basate sul testo*: i contenuti testuali, come commenti o hashtag, vengono analizzati attraverso parole chiave e tecniche di elaborazione del linguaggio naturale, che consentono di interpretare il contesto in cui determinate espressioni vengono utilizzate.
- *Basate sulla somiglianza*: i sistemi di rilevamento della somiglianza permettono di individuare contenuti identici o simili a quelli precedentemente riconosciuti come problematici, considerando anche alcune somiglianze contestuali.
- *Basate sull'attività*: l'analisi delle modalità di utilizzo degli account consente di individuare attività ingannevoli, come bot, spam o tentativi di aumentare artificialmente il coinvolgimento.
- *Basate su LLM*: i *Large Language Models (LLM)* vengono utilizzati per supportare la comprensione del linguaggio e lo svolgimento di attività di moderazione più complesse.
- *Basate su LLM multimodali*: i modelli multimodali permettono di analizzare diverse tipologie di contenuto. TikTok indica, ad esempio, il loro utilizzo nel supporto alla moderazione della disinformazione, attraverso l'individuazione nei video di specifiche affermazioni che possono essere successivamente valutate dai moderatori.

TikTok presta particolare attenzione al contesto del contenuto e al potenziale rischio per gli utenti, anche in assenza di violazioni esplicite.

Le principali categorie considerate nelle linee guida di TikTok includono:

- contenuti violenti;
- comportamenti pericolosi o rischiosi;
- sfide pericolose (*dangerous challenges*);
- contenuti che possono causare danni fisici o psicologici.

Nel caso analizzato in precedenza, relativo alla presenza di uno squalo in prossimità di due individui protetti da una gabbia, il contenuto non presenta violenza esplicita, ma rappresenta una situazione potenzialmente pericolosa.

Seguendo la logica delle linee guida di TikTok, un contenuto di questo tipo potrebbe essere valutato nell'ambito dei comportamenti rischiosi, in quanto mostra una situazione che, se imitata senza adeguate misure di sicurezza, potrebbe comportare un pericolo per l'utente. La piattaforma potrebbe, quindi, valutarne il contesto e applicare, qualora necessario, le misure di moderazione previste dalle proprie linee guida.

3.4 Confronto sperimentale tra Amazon Rekognition, Meta e TikTok

A seguito delle considerazioni precedenti, vengono analizzate diverse immagini al fine di confrontare l'output restituito da Amazon Rekognition con l'approccio alla moderazione previsto dalle linee guida di Meta e TikTok per contenuti analoghi.

Poiché non è possibile accedere direttamente ai sistemi interni di moderazione delle piattaforme Meta e TikTok, sono state utilizzate immagini provenienti da piattaforme come Pexels, selezionate per rappresentare contenuti tipicamente condivisibili sui social media. Amazon Rekognition viene utilizzato per analizzare direttamente le immagini e restituire eventuali etichette di moderazione; per Meta e TikTok, invece, il confronto viene effettuato sulla base delle rispettive linee guida ufficiali, considerando il contesto del contenuto e le eventuali misure di moderazione previste.

3.4.1 Contenuti caratterizzati da rischio implicito

Nel caso dei *contenuti a rischio implicito* sono state svolte due sperimentazioni: la prima riguardante la pratica del *Parkour* e la seconda associata all'*Urbex*.

La scena rappresentata nella prima immagine può essere associata alla pratica del parkour, attività che comporta l'esposizione a una situazione di rischio elevato. Pur in assenza di violenza esplicita, il contenuto rappresenta una condizione potenzialmente pericolosa e imitabile.

Nella Figura 3.8 viene riportata la prima immagine di confronto utilizzata per la sperimentazione.

In questo caso, *Amazon Rekognition* non rileva alcuna etichetta di moderazione, classificando l'immagine come contenuto sicuro. Il risultato evidenzia come, nella situazione analizzata, il sistema non rilevi il rischio implicito associato all'attività rappresentata, in assenza di elementi esplicitamente riconducibili alle proprie categorie di moderazione.

Nella Figura 3.9 viene riportato il risultato della prima sperimentazione di confronto.

Seguendo la logica delle *linee guida di Meta*, la scena potrebbe essere valutata considerando la presenza di un comportamento potenzialmente pericoloso e il rischio associato al contesto



Figura 3.8: Parkour (fonte: Kaique Rocha).

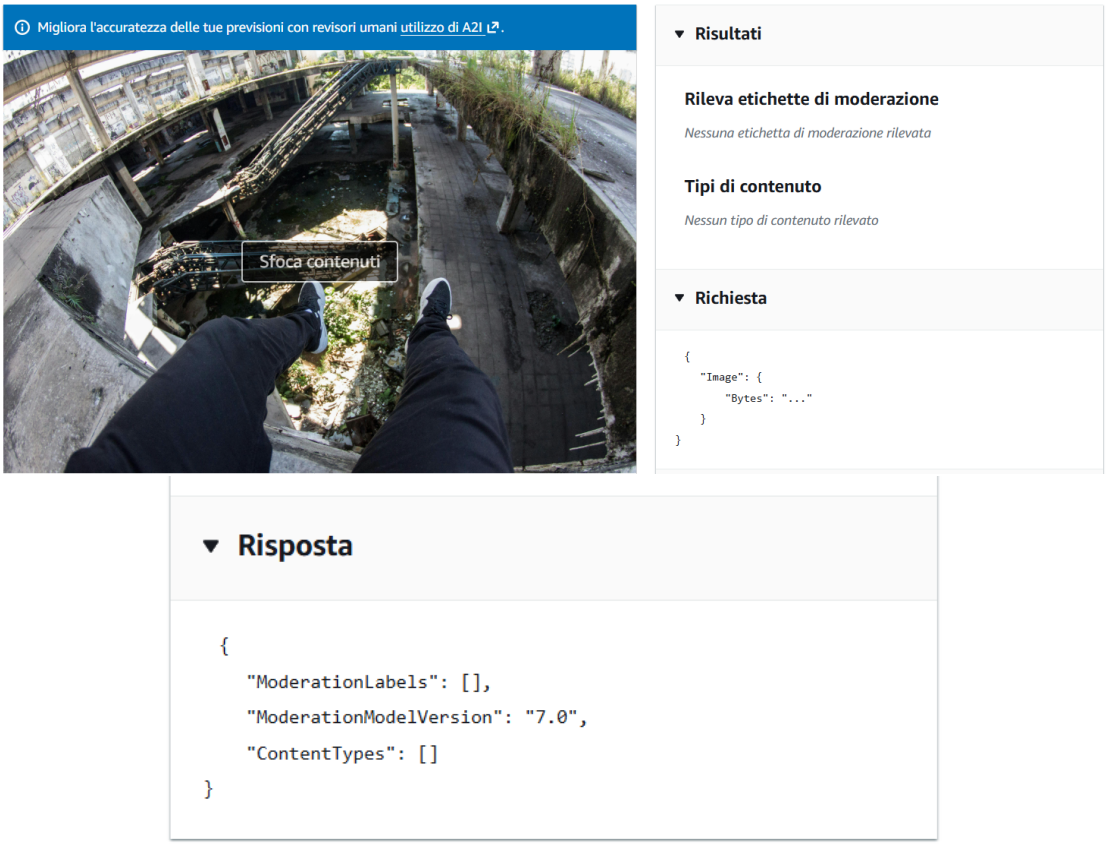


Figura 3.9: Risultato della prima sperimentazione di confronto (Svolta su: Amazon Web Services)

rappresentato. Qualora rientrasse nelle categorie previste dalle policy, potrebbero essere applicate le relative misure di moderazione.

Seguendo la logica delle *linee guida di TikTok*, la scena potrebbe essere valutata nell'ambito dei comportamenti rischiosi, considerando anche il possibile rischio di imitazione, soprattutto per gli utenti più giovani. Anche in questo caso, l'eventuale applicazione di misure di moderazione dipenderebbe dalla valutazione complessiva del contenuto.

La scena rappresentata nella seconda immagine può essere associata alla pratica dell'*Urbex*

(*Urban Exploration*), in quanto l'ambiente circostante appare come un edificio abbandonato e degradato. Tale attività può comportare l'esposizione a situazioni di rischio elevato. Pur in assenza di violenza esplicita, il contenuto rappresenta una condizione potenzialmente pericolosa e imitabile.

Nella Figura 3.10 viene riportata la seconda immagine di confronto utilizzata per la sperimentazione.



Figura 3.10: Urbex (fonte: Cottonbro Studio).

Anche in questo caso, *Amazon Rekognition* non rileva alcuna etichetta di moderazione, classificando l'immagine come contenuto sicuro. Il risultato evidenzia come, nella situazione analizzata, il sistema non rilevi il rischio implicito associato all'attività rappresentata, in assenza di elementi esplicitamente riconducibili alle proprie categorie di moderazione.

Nella Figura 3.11 viene riportato il risultato della seconda sperimentazione di confronto.

Seguendo la logica delle *linee guida di Meta*, la scena potrebbe essere valutata considerando i potenziali rischi associati all'attività rappresentata, come l'esposizione a pericoli strutturali. Qualora il contenuto rientrasse nelle categorie previste dalle policy, potrebbero essere applicate le relative misure di moderazione.

Seguendo la logica delle *linee guida di TikTok*, la scena potrebbe essere valutata nell'ambito dei comportamenti rischiosi, considerando il potenziale pericolo associato all'attività e il possibile rischio di imitazione, soprattutto per gli utenti più giovani. L'eventuale applicazione di misure di moderazione dipenderebbe dalla valutazione complessiva del contenuto.

Alla luce delle rispettive linee guida, Meta e TikTok possono adottare differenti misure di moderazione in funzione delle caratteristiche e del contesto del contenuto. Tali misure non implicano necessariamente la rimozione, ma possono comprendere anche interventi sulla distribuzione o sull'accessibilità del contenuto.

Di seguito vengono riportate le azione svolte da Meta e TikTok:

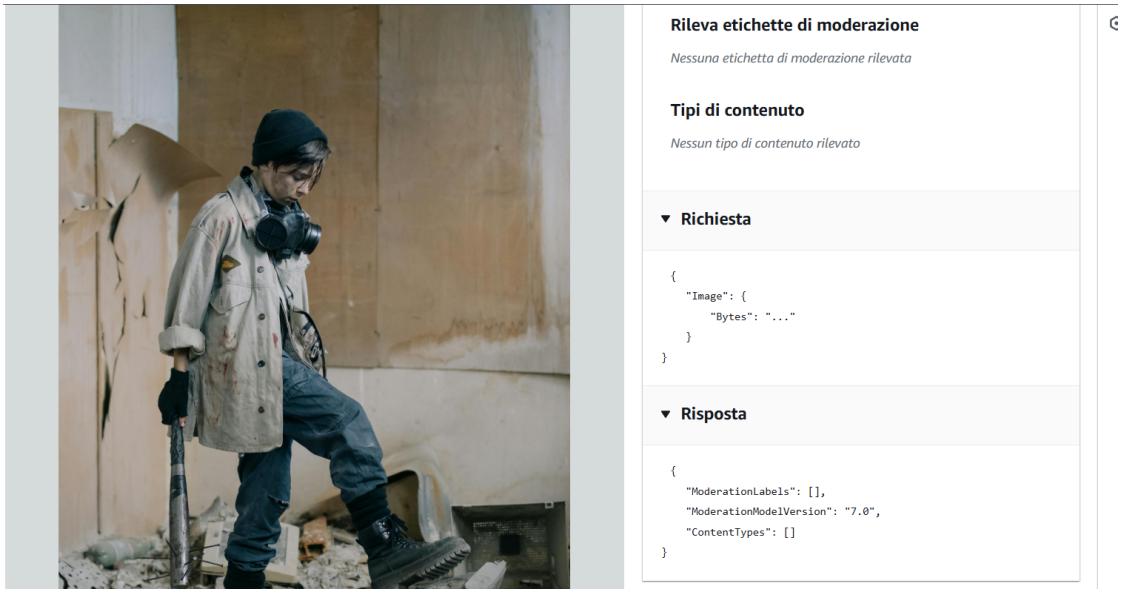


Figura 3.11: Risultato della seconda sperimentazione di confronto (Svolta su: Amazon Web Services)

- *Azioni di Meta*: Meta adotta un sistema di moderazione basato sulle proprie *Community Standards*, con un approccio orientato non esclusivamente alla rimozione del contenuto, ma anche alla riduzione della sua distribuzione e alla contestualizzazione informativa. Le principali azioni che Meta può applicare sono:

  1. *Riduzione della distribuzione algoritmica*: il contenuto può essere soggetto a una riduzione della distribuzione algoritmica, risultando meno visibile nel feed, nella sezione Esplora o nei sistemi di raccomandazione della piattaforma. Questa misura può essere applicata a contenuti considerati problematici, ma non necessariamente in violazione diretta delle policy.
  2. *Etichette informative*: in presenza di contenuti potenzialmente sensibili o che necessitano di una maggiore contestualizzazione, Meta può applicare avvisi o informazioni aggiuntive al fine di fornire maggiore contesto all'utente.
  3. *Rimozione del contenuto*: il contenuto viene rimosso nei casi in cui violi le *Community Standards*, ad esempio quando presenta o promuove contenuti non consentiti dalle politiche della piattaforma.

- *Azioni di TikTok*: TikTok adotta diverse misure di moderazione in base alle caratteristiche del contenuto e alla potenziale violazione delle proprie linee guida:

  1. *Esclusione dai sistemi di raccomandazione (ineligible for recommendation)*: il contenuto può essere classificato come non idoneo alla raccomandazione e, di conseguenza, escluso dai sistemi di raccomandazione della piattaforma, come la sezione *For You Page*[1]. Tale misura può essere applicata ai contenuti che non soddisfano gli standard di idoneità previsti dalla piattaforma.
  2. *Applicazione di etichette di avviso (warning label)*: alcuni contenuti, pur non essendo necessariamente rimossi, possono essere considerati non adatti agli utenti di età inferiore ai 18 anni e, di conseguenza, essere resi accessibili solamente agli utenti maggiorenni.

[1] La *For You Page (FYP)* è il feed home predefinito di TikTok: un flusso infinito di video selezionati per ogni spettatore da un sistema di raccomandazione, attinti in larghissima parte da account che lo spettatore non segue (definizione presa da Glossary).

3. *Rimozione del contenuto*: nei casi in cui il contenuto violi le *Community Guidelines*, TikTok può procedere alla sua rimozione. Ciò può riguardare, ad esempio, contenuti che mostrano o promuovono comportamenti estremamente pericolosi o sfide rischiose (*dangerous challenges*).

3.4.2 Contenuti sensibili e contestualizzati

Per la terza immagine oggetto di studio è stato creato un post fittizio, arricchito da una caption descrittiva e da una serie di hashtag coerenti con l'attività rappresentata.

Sulle piattaforme di Meta e TikTok, contenuti di questo tipo potrebbero essere valutati tenendo conto anche della presenza di elementi testuali, in linguaggio naturale, che permettono di contestualizzare la scena come attività sportiva o simulazione controllata. Tuttavia, tali contenuti potrebbero, comunque, essere soggetti a limitazioni nella distribuzione, venendo mostrati a un pubblico ristretto o esclusi dai sistemi di raccomandazione, in quanto potenzialmente sensibili o non adatti a una diffusione ampia.

Nell'immagine sono raffigurati due individui equipaggiati con abbigliamento tattico e mimetico, posizionati in un ambiente esterno caratterizzato da terreno sabbioso. Entrambi impugnano oggetti riconducibili visivamente ad armi da fuoco e assumono posture che potrebbero essere associate a un'attività di addestramento o simulazione operativa.

Nella Figura 3.12 viene riportata l'immagine utilizzata per la terza sperimentazione.



training.official Training day 💪
Focus, disciplina e lavoro di squadra fanno la differenza.
Simulazioni tattiche in ambiente controllato.
#training #teamwork #airsoft #simulation #discipline
#focus #outdoortraining #sport

Figura 3.12: Immagine di Softair rielaborata da (fonte: Pixabay).

L'analisi dell'immagine tramite i sistemi di Content Moderation di Amazon Rekognition evidenzia una classificazione del contenuto come violento.

Nella Figura 3.13 viene riportata la prima parte del risultato della sperimentazione contestualizzata.

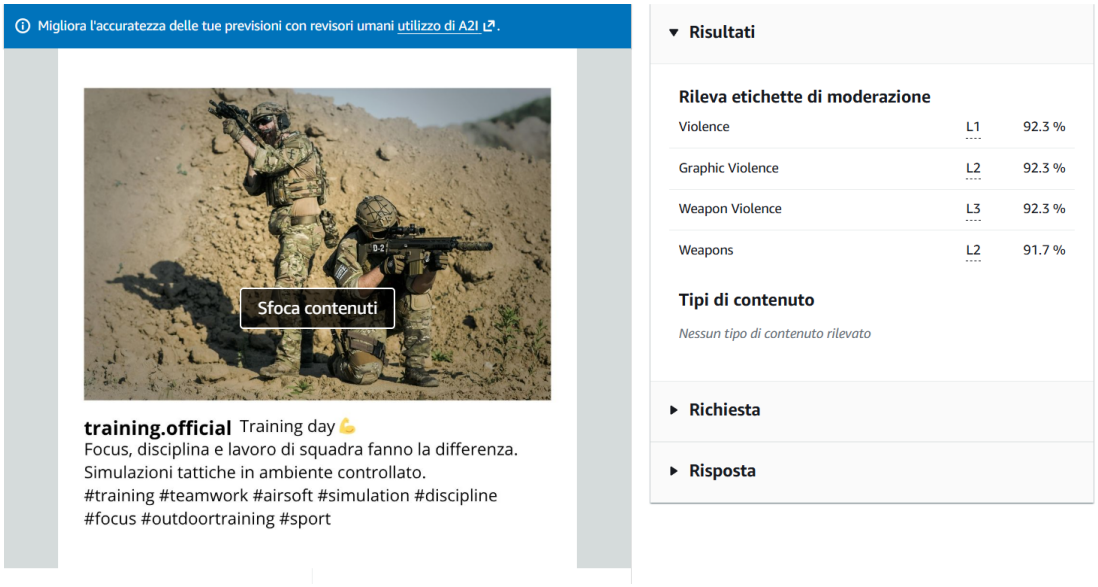


Figura 3.13: Prima parte del risultato della sperimentazione contestualizzata (Svolta su: Amazon Web Services)

Nella Figura 3.13 viene riportata la seconda parte del risultato della sperimentazione contestualizzata.

In particolare, vengono rilevate le seguenti etichette:

- *Weapon Violence*: con una confidenza pari al 92.3%;
- *Graphic Violence*: con una confidenza pari al 92.3%;
- *Violence*: con una confidenza pari al 92.3%;
- *Weapons*: con una confidenza pari al 91.7%;

Tali valori indicano come il sistema interpreti l'immagine come fortemente associata a contesti di violenza armata, senza considerare gli elementi testuali del post che permettono di contestualizzare la situazione rappresentata.

Mentre il sistema di moderazione automatica di Amazon Rekognition fornisce una classificazione basata principalmente sugli elementi visivi rilevati nell'immagine, come le armi e la postura dei soggetti, le piattaforme social possono integrare tali informazioni con il contesto semantico fornito dal contenuto.

Questo caso evidenzia una differenza significativa tra l'analisi automatica dell'immagine e una moderazione che considera anche il contesto: a fronte di un'elevata confidenza associata da Amazon Rekognition a categorie relative alla violenza e alle armi, il contenuto potrebbe essere valutato diversamente qualora ulteriori elementi contestuali ne chiarissero la natura, pur potendo rimanere soggetto a eventuali limitazioni previste dalle piattaforme.

3.4.3 Analisi dei risultati

Dall'analisi dei risultati emerge come Amazon Rekognition compia la classificazione dei contenuti principalmente sulla base degli elementi visivi rilevati nell'immagine, mostrando alcuni limiti nella valutazione di situazioni caratterizzate da rischio implicito o che richiedono una maggiore contestualizzazione. Le piattaforme social analizzate, invece, adottano un approccio alla Content Moderation più articolato, che può considerare anche il contesto associato al contenuto.



Figura 3.14: Seconda parte del risultato della contestualizzata (Svolta su: Amazon Web Services)

Nel caso specifico di TikTok, le linee guida mostrano una particolare attenzione non solo al contesto del contenuto, ma anche al potenziale effetto che questo potrebbe avere sugli utenti e ai comportamenti che potrebbe incentivare. Assume particolare importanza anche la valutazione del rischio di imitazione, soprattutto in presenza di contenuti che rappresentano comportamenti potenzialmente pericolosi.

CAPITOLO 4

Un case study relativo al riconoscimento visivo

In questo capitolo vengono descritte la progettazione e l'implementazione di un case study relativo al riconoscimento visivo mediante l'utilizzo di Amazon Rekognition. In particolare, vengono analizzati contenuti caratterizzati da differenti livelli di complessità, al fine di valutare il comportamento del sistema nel confronto e nell'analisi delle caratteristiche facciali degli individui.
Successivamente, viene svolta una sperimentazione sui volti dei soggetti analizzati, simulando alcune situazioni riconducibili a un contesto investigativo. In particolare, lo studio viene svolto su tre casistiche fondamentali: Aging Case, Basic Case e Twins Case.
L'obiettivo della sperimentazione è valutare le capacità del sistema e individuare eventuali criticità che possono emergere nell'utilizzo di tali strumenti in ambito investigativo.

4.1 Facial Analysis

La *Facial Analysis* è il processo di analisi automatizzata dei volti umani presenti in immagini o video, attraverso tecniche di computer vision e algoritmi di Intelligenza Artificiale. L'obiettivo è estrarre informazioni utili a partire dalle caratteristiche facciali, offrendo anche la possibilità di svolgere un confronto tra diversi volti.

Le principali funzionalità della *Facial Analysis* includono:

- il rilevamento dei volti (*face detection*);
- l'estrazione di caratteristiche facciali distintive;
- la stima di attributi quali età e genere;
- l'analisi delle espressioni facciali;
- il confronto tra volti (*face comparison*).

La *Facial Analysis* si focalizza sui soggetti presenti, risultando particolarmente rilevante in ambiti quali la sicurezza, la sorveglianza e il supporto alle attività investigative.

4.1.1 La Facial Analysis in ambito investigativo

Con l'avvento dell'Intelligenza Artificiale si sono verificati profondi cambiamenti nella vita delle persone. Tali trasformazioni non si limitano alla sfera quotidiana, ma si estendono anche al contesto lavorativo; tra gli ambiti maggiormente influenzati emerge quello investigativo, in cui le nuove tecnologie stanno ridefinendo strumenti e metodologie di analisi.

In questo contesto si inseriscono le tecniche di Facial Analysis, che permettono di analizzare i volti presenti in immagini e video al fine di estrarre informazioni utili per attività di identificazione e supporto alle indagini.

Il riconoscimento facciale rappresenta una delle principali applicazioni dell'Intelligenza Artificiale nell'ambito della sorveglianza e si basa sull'analisi delle *caratteristiche facciali* degli individui. In ambito investigativo, tale tecnologia consente di confrontare immagini di riferimento di un determinato soggetto con immagini acquisite da altre fonti, come ad esempio telecamere di sorveglianza, al fine di individuare possibili corrispondenze.

I sistemi di riconoscimento facciale possono operare secondo due modalità principali: *identificazione* e *verifica*. Nel primo caso viene effettuato un confronto *uno-a-molti* (1:N), in cui il volto acquisito viene confrontato con più volti presenti all'interno di un database, al fine di individuare una possibile corrispondenza. La verifica, invece, si basa su un confronto *uno-a-uno* (1:1), nel quale il volto acquisito viene confrontato con un singolo volto di riferimento per verificare l'identità dichiarata.

I moderni sistemi di riconoscimento facciale si basano sull'impiego di tecniche di Intelligenza Artificiale che consentono di elaborare le informazioni presenti nelle immagini, distinguendo i volti umani dagli altri elementi presenti nella scena e analizzando, successivamente, le caratteristiche facciali dei soggetti rilevati. Il processo di riconoscimento parte dall'acquisizione di un'immagine o di un video, all'interno del quale viene rilevata la presenza di uno o più volti. Successivamente, le caratteristiche facciali vengono individuate ed estratte per ottenere una *rappresentazione digitale* del volto. Quest'ultima può essere confrontata con quella di uno o più volti di riferimento, calcolandone matematicamente il livello di somiglianza. In base al risultato del confronto e alla soglia stabilita dal sistema è possibile determinare la presenza di una corrispondenza.

Il processo può essere sintetizzato attraverso la seguente *pipeline*:

Immagine → rilevamento del volto → estrazione delle caratteristiche → rappresentazione digitale → confronto → livello di similarità → possibile corrispondenza.

La Figura 4.1 mostra una rappresentazione schematica di tale pipeline.

È importante sottolineare la natura probabilistica del riconoscimento facciale. I risultati ottenuti non rappresentano identificazioni definitive, ma esprimono un livello di somiglianza tra i volti analizzati. Di conseguenza, il sistema può essere soggetto a errori di riconoscimento, quali *falsi positivi* e *falsi negativi*.

4.1.2 Amazon Rekognition e Facial Comparison

La *Facial Comparison* è uno dei vari servizi offerti da Amazon Rekognition e viene utilizzata per effettuare un confronto tra i volti contenuti in due immagini fornite dall'utente. La prima immagine viene definita *immagine sorgente*, mentre la seconda è detta *immagine di destinazione*. Il sistema ha lo scopo di confrontare i volti presenti nelle immagini al fine di individuare eventuali corrispondenze.

Questo servizio utilizza il parametro *SimilarityThreshold*, che definisce il livello di similarità necessario affinché una corrispondenza venga restituita dal sistema. Il confronto viene effettuato attraverso algoritmi di apprendimento automatico di natura probabilistica; di conseguenza, possono verificarsi errori, tra cui i *falsi negativi*, che si verificano quando un

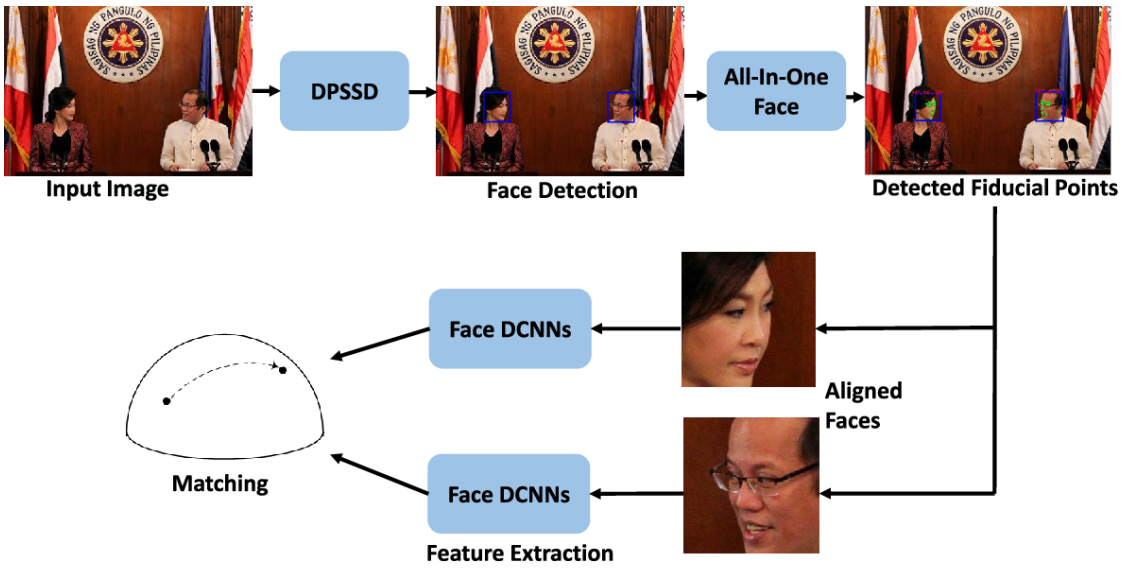


Figura 4.1: Rappresentazione della pipeline di Fiacial Analysis (fonte: Semanticscholar).

volto presente nell'immagine di destinazione ottiene erroneamente un basso punteggio di similarità rispetto al volto presente nell'immagine sorgente. Per rendere il confronto più affidabile e ridurre la probabilità di errori, è consigliabile confrontare il soggetto con più immagini di riferimento.

L'utilizzo di questa tecnologia può, però, comportare problematiche anche in ambito giuridico e di tutela della privacy.

4.2 Obiettivo e progettazione del case study

Ai fini della presente sperimentazione, l'analisi si è concentrata principalmente su tre casi: un *Aging Case*, un *Basic Case* e un *Twins Case*.

Come contesto di riferimento per il *case study* si è scelto quello investigativo, simulando alcune situazioni in cui potrebbe risultare necessario confrontare il volto di un soggetto ricercato con immagini provenienti da più fonti, quali, ad esempio, video di sorveglianza o foto pubblicate sui social. La sperimentazione considera:

- variazioni nell'aspetto dovute al passare degli anni;
- confronto tra individui differenti;
- soggetti legati da una forte somiglianza facciale.

L'aspettativa è che il sistema incontri maggiori difficoltà nel *Twins case*, caratterizzato dalla presenza di due soggetti con caratteristiche facciali estremamente simili. Tale scenario risulta particolarmente significativo per valutare le capacità e gli eventuali limiti della *Facial Comparison* in un contesto investigativo.

La sperimentazione proposta non riproduce integralmente le procedure adottate dalle forze dell'ordine, ma punta, tramite l'utilizzo di Amazon Rekognition, ad analizzare alcune problematiche che possono emergere dall'impiego del riconoscimento facciale in ambito investigativo.

Per l'implementazione della fase di analisi automatica si è scelto di utilizzare Amazon Rekognition, un servizio di Amazon Web Services che consente di analizzare immagini e video attraverso tecniche di Intelligenza Artificiale. Nell'ambito dell'analisi facciale, il

sistema consente di individuare la presenza di volti umani all'interno di un'immagine, distinguendoli dagli altri elementi nella scena, per procedere successivamente all'analisi delle loro caratteristiche.

Tra le funzionalità messe a disposizione da Amazon Rekognition si è scelto di utilizzare la *Facial Comparison* e la *Facial Analysis* dei singoli volti. L'obiettivo è analizzare in che modo questi strumenti possano supportare il processo di riconoscimento di un soggetto ricercato e osservare eventuali criticità che possano emergere dal loro utilizzo in ambito investigativo. I tre scenari sono stati, quindi, definiti secondo un livello crescente di complessità, così da osservare il comportamento del sistema.

4.3 Sperimentazione con Amazon Rekognition

Definiti gli obiettivi della sperimentazione, si procede all'analisi delle immagini selezionate nei tre casi precedentemente individuati: *Aging Case*, *Basic Case* e *Twins Case*, al fine di osservare e confrontare gli output restituiti dal sistema.

4.3.1 Aging Case

Le immagini utilizzate per il primo caso di studio sono state fornite direttamente dalla piattaforma AWS Rekognition a scopo dimostrativo. Il loro impiego è finalizzato esclusivamente all'analisi del comportamento del sistema in condizioni controllate.

Nella Figura 4.2 viene presentato l'*Aging Case*.

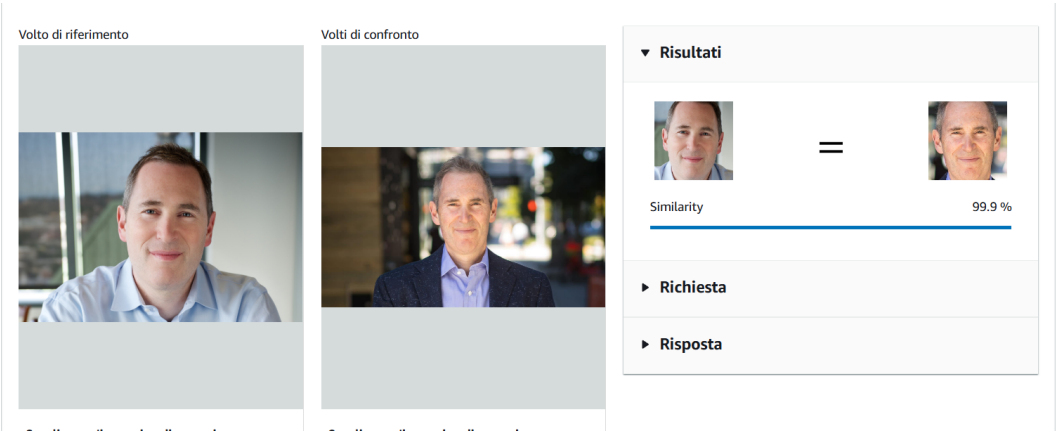


Figura 4.2: Aging Case: confronto tra immagini dello stesso soggetto in età differenti mediante Amazon Rekognition (immagini fornite da Amazon Rekognition).

Nell'*Aging Case* viene analizzato il riconoscimento di un individuo in presenza di variazioni estetiche dovute all'invecchiamento o alle condizioni di acquisizione dell'immagine.

Le due immagini utilizzate raffigurano lo stesso soggetto in momenti e contesti differenti, caratterizzati da variazioni nell'illuminazione, nell'espressione facciale, nell'ambiente circostante e nell'età dell'individuo. Tali differenze permettono di simulare una situazione riconducibile all'ambito investigativo, in cui le immagini disponibili possono provenire da fonti differenti, come fotografie, telecamere di sorveglianza o archivi pregressi. Nel caso di un soggetto ricercato per un periodo prolungato, ad esempio, le immagini più recenti possono essere acquisite a distanza di anni da quelle originariamente disponibili, determinando variazioni nell'aspetto dovute al processo di invecchiamento.

Il sistema di *Facial Comparison* ha restituito un valore di *similarità* pari al *99,9%*, indicando un'elevata corrispondenza tra i due volti analizzati.

Il risultato ottenuto mostra come, nel caso analizzato, il sistema sia stato in grado di riconoscere correttamente lo stesso individuo nonostante le variazioni presenti nelle due immagini.

Tuttavia, è importante sottolineare che un elevato valore di similarità non costituisce di per sé una garanzia assoluta dell'identità del soggetto, rendendo, comunque, necessaria la valutazione del risultato da parte dell'analista umano.

Nelle Figure 4.3 – 4.6 viene presentato il risultato dell'*Aging Case*.



Figura 4.3: Aging Case: prima parte del risultato.

La richiesta inviata al servizio specifica le due immagini da confrontare, denominate, rispettivamente, *SourceImage* e *TargetImage*. Inoltre, viene impostato il parametro *SimilarityThreshold*. La risposta restituita dal sistema contiene diverse informazioni utili ai fini dell'analisi. Oltre al valore di *Similarity* precedentemente analizzato, l'output presenta ulteriori dettagli tecnici relativi ai volti rilevati e al confronto effettuato. L'insieme di queste informazioni consente di analizzare in maniera più dettagliata il risultato del confronto e di valutarne la significatività rispetto alla soglia impostata.

4.3.2 Basic Case

Le immagini utilizzate nel secondo caso di studio provengono, rispettivamente, da Pexels e Tv Tropes e sono state impiegate a scopo dimostrativo. Il loro utilizzo è finalizzato esclusivamente all'analisi del comportamento del sistema in condizioni controllate.

Nella Figura 4.7 viene presentato il *Basic Case*. In esso, viene analizzata la capacità del sistema di distinguere due individui differenti. Le immagini utilizzate mostrano soggetti con caratteristiche facciali chiaramente distinguibili. Questo scenario rappresenta una condizione di riferimento utile per osservare il comportamento del sistema in un contesto semplice e controllato.

Il sistema di *Facial Comparison* ha restituito un valore di similarità quasi nulla pari a circa 0,6%, indicando un livello di somiglianza estremamente basso tra i due volti. Il risul-

▼ **Risposta**


```
{
  "SourceImageFace": {
      "BoundingBox": {
          "Width": 0.2888343930244446,
          "Height": 0.664422869682312,
          "Left": 0.38120219111442566,
          "Top": 0.15815633535385132
      },
      "Confidence": 99.99982452392578
  },
  "FaceMatches": [
      {
          "Similarity": 99.99077606201172,
          "Face": {
              "BoundingBox": {
                  "Width": 0.18647146224975586,
                  "Height": 0.4844239056110382,
                  "Left": 0.40962982177734375,
                  "Top": 0.16498763859272003
              },
```


Figura 4.4: Aging Case: seconda parte del risultato.


```
          "Landmarks": [
              {
                  "Type": "eyeLeft",
                  "X": 0.4640858769416809,
                  "Y": 0.36171993613243103
              },
              {
                  "Type": "eyeRight",
                  "X": 0.5430324673652649,
                  "Y": 0.3581247627735138
              },
              {
                  "Type": "mouthLeft",
                  "X": 0.4730469584465027,
                  "Y": 0.5184903740882874
              },
              {
                  "Type": "mouthRight",
                  "X": 0.5388632416725159,
                  "Y": 0.5155081152915955
              },
              {
                  "Type": "nose",
                  "X": 0.5056883692741394,
                  "Y": 0.4460171163082123
              }
          ],
```


Figura 4.5: Aging Case: terza parte del risultato.

```
                "Pose": {
                    "Roll": -2.5212061405181885,
                    "Yaw": 0.8507153391838074,
                    "Pitch": 3.568561553955078
                },
                "Quality": {
                    "Brightness": 93.53854370117188,
                    "Sharpness": 96.61495208740234
                }
            }
        }
    ],
    "UnmatchedFaces": []
}
```


Figura 4.6: Aging Case: quarta parte del risultato.

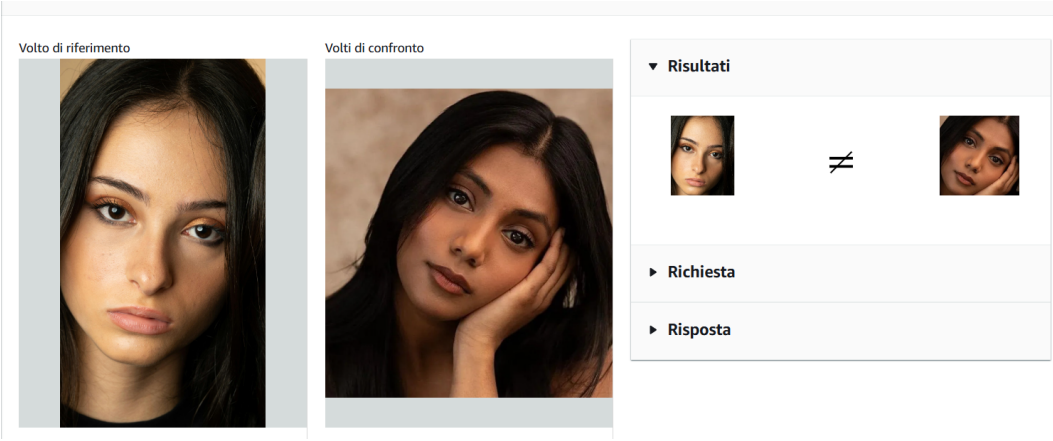


Figura 4.7: Basic Case (fonte: Pexels e Tvtropes)

tato ottenuto mostra come, nel caso analizzato, il sistema sia stato in grado di distinguere correttamente i due individui in presenza di caratteristiche facciali marcatamente differenti.

Nelle Figure 4.8 – 4.11 viene presentato il risultato del *Basic Case*.



Figura 4.8: Basic Case: prima parte del risultato.



Figura 4.9: Basic Case: seconda parte del risultato.

La richiesta inviata al servizio utilizza le due immagini da confrontare, indicate, rispettivamente, come *SourceImage* e *TargetImage*, e prevede l'impostazione del parametro *SimilarityThreshold*. La risposta restituita dal sistema presenta il risultato del confronto attraverso il valore di *Similarity*, che nel caso analizzato risulta pari a circa 0,6%.

Il *Basic Case* costituisce un caso di riferimento per la sperimentazione, in quanto permette di osservare il comportamento del sistema in una condizione caratterizzata da differenze facciali evidenti tra i soggetti. Il risultato potrà successivamente essere confrontato con quello ottenuto nel *Twins Case*, caratterizzato invece da un elevato grado di somiglianza tra gli individui.

4.3.3 Twins Case

Le immagini utilizzate nel terzo caso di studio sono state fornite da Pexels a scopo dimostrativo. Tali immagini sono utilizzate esclusivamente per analizzare il comportamento del sistema in condizioni controllate.

Nella Figura 4.12 viene presentato il *Twins Case*. In esso, viene analizzato il comportamento del sistema nel confronto tra due volti appartenenti a soggetti distinti ma caratterizzati da un'elevata somiglianza facciale.

Le immagini utilizzate mostrano due individui con caratteristiche morfologiche molto simili, in quanto gemelli.

Il sistema di *Facial Comparison* ha restituito un valore di similarità pari al *99,9%*, indicando un elevato livello di somiglianza tra i due volti analizzati, pur trattandosi di individui distinti.

Questo risultato evidenzia una criticità nel confronto tra individui caratterizzati da tratti facciali estremamente simili, poiché un valore di similarità così elevato potrebbe rendere più complessa la distinzione tra soggetti differenti.

```
"Landmarks": [
    {
        "Type": "eyeLeft",
        "X": 0.475136935710907,
        "Y": 0.3582390546798706
    },
    {
        "Type": "eyeRight",
        "X": 0.6651296615600586,
        "Y": 0.4758029878139496
    },
    {
        "Type": "mouthLeft",
        "X": 0.3541860580444336,
        "Y": 0.5604339241981506
    },
    {
        "Type": "mouthRight",
        "X": 0.5122057199478149,
        "Y": 0.6577498912811279
    },
    {
        "Type": "nose",
        "X": 0.5089825987815857,
        "Y": 0.521691083908081
    }
],
```


Figura 4.10: Basic Case: terza parte del risultato.

In un contesto investigativo, tale comportamento potrebbe aumentare il rischio di un *falso positivo* qualora il risultato automatico venisse considerato come unico strumento di identificazione. Tuttavia, è importante sottolineare che il valore di similarità non costituisce una prova certa dell'identità del soggetto, rendendo necessario il supporto dell'analisi umana nella validazione dei risultati.

Nelle Figure 4.13 – 4.16 viene presentato il risultato del *Twins Case*.

L'analista umano noterebbe che le due gemelle, pur presentando caratteristiche facciali estremamente simili, non sono completamente identiche. All'occhio umano è, infatti, possibile individuare alcuni elementi distintivi, tra cui la presenza di un *neo* in prossimità del lato sinistro del labbro in uno dei due soggetti. Nonostante tale differenza visibile, il sistema di *Facial Comparison* ha restituito un valore di *Similarity* pari a circa 99,9%, classificando i due volti come appartenenti allo stesso individuo, sebbene si tratti di due persone distinte.

Questo risultato evidenzia come la presenza di un elemento distintivo chiaramente osservabile, quale il neo, non sia stata sufficiente a evitare un valore di similarità estremamente elevato. L'osservatore umano può, invece, utilizzare tale caratteristica come elemento distintivo aggiuntivo per distinguere i due individui.

4.3.4 Facial Analysis dei soggetti del Twins Case

Nel contesto del *Twins Case*, viene effettuato un test di *Facial Analysis* sulle singole foto delle due gemelle, analizzando separatamente ciascun volto, al fine di verificare se il sistema

```
            "Pose": {
                "Roll": 34.01010513305664,
                "Yaw": 3.9924845695495605,
                "Pitch": 1.5788910388946533
            },
            "Quality": {
                "Brightness": 61.563446044921875,
                "Sharpness": 92.22801208496094
            }
        }
    }
],
"UnmatchedFaces": []
}
```


Figura 4.11: Basic Case: quarta parte del risultato.

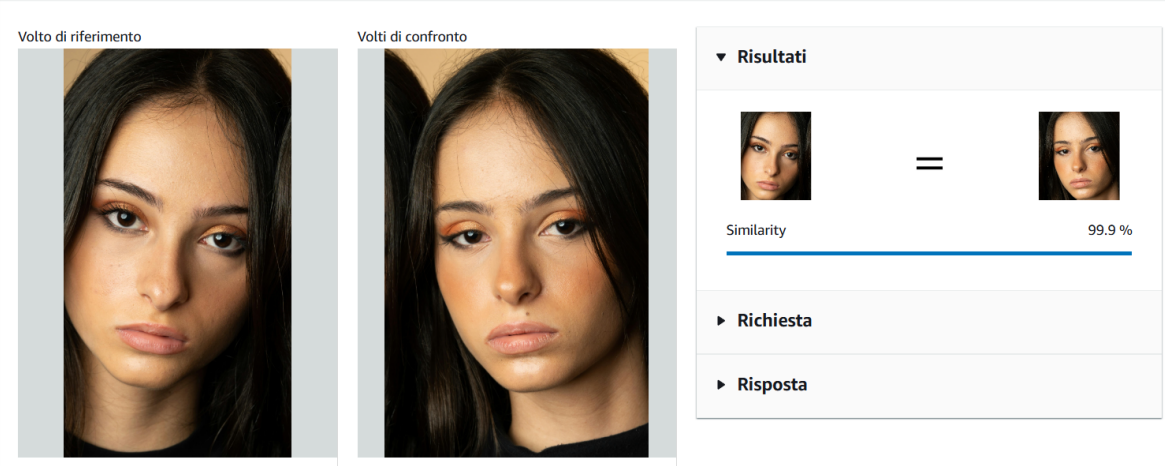


Figura 4.12: Twins Case (fonte: Pexels)

Richiesta

```
{
  "TargetImage": {
      "Bytes": {}
  },
  "SourceImage": {
      "Bytes": {}
  },
  "SimilarityThreshold": 0
}
```


Figura 4.13: Twins Case: prima parte del risultato.

evidenzi eventuali differenze nelle caratteristiche facciali rilevate che possano contribuire alla distinzione tra i due individui in assenza di un confronto diretto.

Nel caso di soggetti estremamente simili, come le gemelle considerate, l'obiettivo di questa parte della sperimentazione è verificare se le informazioni aggiuntive restituite dalla *Facial Analysis* evidenzino differenze tra i due soggetti che possano supportare l'analista



Figura 4.14: Twins Case: seconda parte del risultato.

umano nella loro distinzione.

Nella Figura 4.17 sono riportati l'immagine della prima gemella e i principali risultati ottenuti mediante la *Facial Analysis*

Il sistema di Facial Analysis ha rilevato correttamente la presenza di un volto umano con un livello di confidenza pari al 99,9%, identificando il soggetto come di genere femminile con un'età stimata compresa tra i 16 e i 22 anni.

Per quanto riguarda le caratteristiche facciali, il volto risulta non sorridente, con occhi aperti e bocca chiusa, mentre non viene rilevata la presenza di occhiali. Il sistema indica, inoltre, che il volto risulta completamente visibile e non presenta elementi che ne coprano significativamente i tratti.

Dal punto di vista emotivo, l'espressione viene classificata principalmente come *calm* (52.1%), evidenziando uno stato neutro del soggetto.

Nella Figura 4.18 viene riportata la richiesta generata dal sistema durante il processo di analisi facciale.

Nell'immagine viene rappresentata la richiesta inviata al servizio per l'analisi della prima gemella, specificando l'immagine attraverso il parametro *Image* e impostando il parametro *Attributes* su *ALL*.

Nelle Figure 4.19 – 4.20 viene riportata la risposta generata dal sistema durante il processo di analisi facciale.

Successivamente viene mostrata la risposta restituita dal sistema, contenente le informazioni relative al volto rilevato, tra cui la posizione del volto all'interno dell'immagine (*BoundingBox*), la stima dell'età e del genere e gli ulteriori attributi facciali analizzati.

Nella Figura 4.21 sono riportati l'immagine della seconda gemella e i principali risultati ottenuti mediante la *Facial Analysis*

```
"Landmarks": [
    {
        "Type": "eyeLeft",
        "X": 0.32681262493133545,
        "Y": 0.45542868971824646
    },
    {
        "Type": "eyeRight",
        "X": 0.6592627167701721,
        "Y": 0.43635445833206177
    },
    {
        "Type": "mouthLeft",
        "X": 0.3915637135505676,
        "Y": 0.6937611103057861
    },
    {
        "Type": "mouthRight",
        "X": 0.6681901216506958,
        "Y": 0.6783421635627747
    },
    {
        "Type": "nose",
        "X": 0.46399548649787903,
        "Y": 0.5760756134986877
    }
],
```


Figura 4.15: Twins Case: terza parte del risultato.

Anche in questo caso, il sistema di Facial Analysis rileva correttamente la presenza di un volto umano con un livello di confidenza pari al 99,9%, identificando il soggetto come di genere femminile e stimando un'età compresa tra i 16 e i 22 anni.

Rispetto al caso precedente, i principali attributi generali risultano sostanzialmente coerenti: il volto risulta completamente visibile e non presenta elementi che ne coprano significativamente i tratti, non sono rilevati occhiali e il soggetto mantiene un'espressione non sorridente, con bocca chiusa.

Una differenza rilevata riguarda la classificazione dell'espressione facciale: mentre nella prima gemella il sistema indica prevalentemente uno stato *calm*, nella seconda gemella viene rilevata principalmente l'emozione *sad*.

Tale informazione non può essere considerata un elemento utile per distinguere o identificare con certezza i due soggetti. In ambito investigativo, attributi di questo tipo possono essere utilizzati come informazioni accessorie, ma non possono essere considerati elementi necessari per determinare l'identità di un individuo.

Nel caso analizzato, la *Facial Analysis* fornisce elementi descrittivi aggiuntivi, senza tuttavia consentire una distinzione certa tra i due individui.

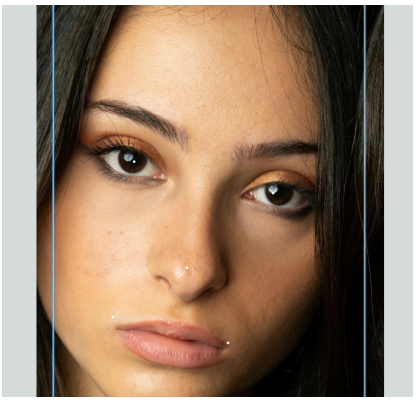
Nella Figura 4.22 viene riportata la richiesta generata dal sistema durante il processo di analisi facciale.

Nella Figura 4.23 viene riportata la risposta generata dal sistema durante il processo di analisi facciale.

La risposta restituita dal sistema presenta informazioni relative al volto analizzato, tra cui la posizione del volto all'interno dell'immagine (*BoundingBox*), la fascia d'età stimata

```
                ],
                "Pose": {
                    "Roll": -6.289372444152832,
                    "Yaw": -10.085320472717285,
                    "Pitch": 0.883030891418457
                },
                "Quality": {
                    "Brightness": 72.63345336914062,
                    "Sharpness": 98.08562469482422
                }
            }
        }
    ],
    "UnmatchedFaces": []
}
```


Figura 4.16: Twins Case: quarta parte del risultato.



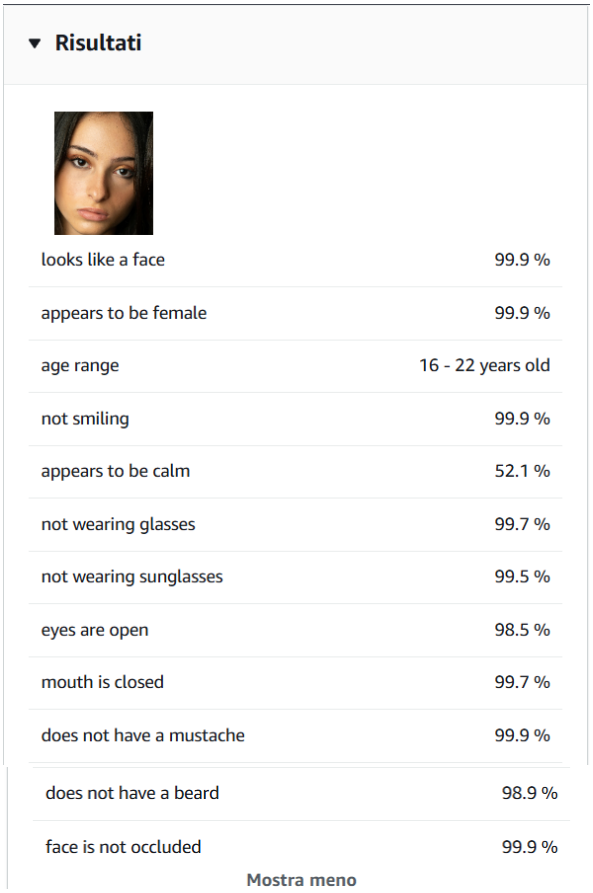


Figura 4.17: Analisi facciale della prima gemella

(*AgeRange*) e ulteriori attributi descrittivi. Tali informazioni risultano complessivamente coerenti con quanto osservato nell'analisi della prima gemella, ad eccezione della differente classificazione dell'espressione facciale.

Tuttavia, la sperimentazione non ha evidenziato gli elementi distintivi attesi. Infatti, è importante sottolineare che, anche attraverso l'analisi dei singoli volti, il *neo*, individuato mediante l'osservazione umana, non emerge fra gli attributi restituiti dal sistema. Tale



Figura 4.18: Richiesta relativa alla prima gemella.



Figura 4.19: Risposta relativa alla prima gemella: prima parte

```
"Gender": {
    "Value": "Female",
    "Confidence": 99.93692779541016
},
"Beard": {
    "Value": false,
    "Confidence": 98.9089584350586
},
"Mustache": {
    "Value": false,
    "Confidence": 99.90691375732422
},
"EyesOpen": {
    "Value": true,
    "Confidence": 98.57299041748047
},
"MouthOpen": {
    "Value": false,
    "Confidence": 99.70561981201172
},
"Emotions": [
    {
        "Type": "CALM",
        "Confidence": 52.193363189697266
    },
    {
        "Type": "SAD",
        "Confidence": 28.111570358276367
    },
```


Figura 4.20: Risposta relativa alla prima gemella: seconda parte

caratteristica rappresenta uno degli elementi distintivi più evidenti tra i volti delle due ragazze e può essere utilizzata per distinguere visivamente le gemelle.

La sua assenza tra le informazioni restituite evidenzia come la *Facial Analysis* non sia sufficiente a distinguere con certezza soggetti caratterizzati da un'elevata somiglianza facciale. In un contesto investigativo, ciò rafforza la necessità di integrare l'analisi automatica con la valutazione dell'analista umano.

4.3.5 Analisi dei risultati

I tre scenari analizzati hanno consentito di valutare il sistema in condizioni differenti:

- riconoscimento dello stesso individuo in immagini acquisite in momenti diversi;
- confronto tra soggetti chiaramente differenti;
- infine, il caso più complesso, rappresentato da individui caratterizzati da un'elevata somiglianza facciale.

Nell'*Aging Case*, il sistema è riuscito a riconoscere lo stesso individuo nonostante le variazioni nell'aspetto legate all'invecchiamento e alle differenti condizioni di acquisizione delle immagini. Nel *Basic Case* il sistema ha prodotto un valore di similarità quasi nullo, restituendo un risultato coerente con il fatto che le immagini raffigurassero due individui differenti. I primi due esperimenti mostrano, quindi, un comportamento coerente con le caratteristiche dei soggetti rappresentati nelle immagini.

La criticità più significativa emerge dal *Twins Case*. Nonostante le immagini raffigurino due individui distinti, il sistema di *Facial Comparison* ha mostrato una forte difficoltà nel



▼ **Risultati**



looks like a face	99.9 %
appears to be female	99.7 %
age range	16 - 22 years old
not smiling	99.9 %
appears to be sad	98 %
not wearing glasses	99.5 %
not wearing sunglasses	99 %
eyes are open	83.8 %
mouth is closed	99.6 %
does not have a mustache	99.8 %
does not have a beard	98.7 %
face is not occluded	99.9 %

Mostra meno

Figura 4.21: Analisi facciale della seconda gemella

▼ **Richiesta**

```
{
  "Image": {
      "Bytes": "..."
  },
  "Attributes": [
      "ALL"
  ]
}
```

Figura 4.22: Richiesta relativa alla seconda gemella.

distinguere i due soggetti. L'osservazione umana ha, invece, permesso di individuare un elemento distintivo evidente, rappresentato dalla presenza di un *neo*.

L'analisi separata dei due volti mediante *Facial Analysis* ha fornito ulteriori attributi descrittivi, evidenziando anche alcune differenze nella classificazione dell'espressione facciale; tuttavia, il neo individuato attraverso l'osservazione umana non emerge tra gli attributi restituiti dal sistema. Le informazioni ottenute attraverso la Facial Analysis risultano utili come elementi descrittivi aggiuntivi, ma non sufficienti, da sole, a determinare con certezza la distinzione tra i due soggetti in un contesto investigativo.

Nel complesso, la sperimentazione evidenzia come le tecniche di Facial Analysis e Facial Comparison possano costituire strumenti di supporto alle attività investigative, fornendo



Figura 4.23: Risposta seconda gemella.

informazioni utili all'analisi dei soggetti presenti nelle immagini.

CAPITOLO 5

Un case study relativo all'analisi dei sentimenti

In questo capitolo vengono descritte la progettazione e l'implementazione di un case study relativo all'analisi dei sentimenti mediante l'utilizzo di Amazon SageMaker. In particolare, vengono analizzati testi letterari e poetici caratterizzati da differenti significati e componenti emotive, al fine di valutare il comportamento del sistema nel confronto e nell'analisi delle emozioni espresse. La sperimentazione prevede il confronto tra un modello di sentiment analysis binaria e un modello di emotion classification multi-classe, applicato a testi in lingua inglese. I testi selezionati presentano elementi allegorici, simbolici e significati impliciti, che permettono di osservare il comportamento dei modelli nell'analisi di contenuti emotivamente complessi. I risultati ottenuti dai due modelli vengono confrontati fra di loro e, successivamente, verificati rispetto all'interpretazione complessiva dei testi, al fine di individuare eventuali differenze tra la classificazione prodotta dall'Intelligenza Artificiale e il significato delle opere analizzate.

5.1 Amazon SageMaker e Natural Language Processing

Amazon SageMaker è una piattaforma di Machine Learning sviluppata da Amazon Web Services che consente di creare, addestrare ed eseguire modelli di Intelligenza Artificiale attraverso strumenti creati per lo sviluppo e l'analisi dei dati.

Tra gli ambiti applicativi supportati da SageMaker rientra il *Natural Language Processing (NLP)*, settore dell'Intelligenza Artificiale che si occupa dell'elaborazione e dell'interpretazione del linguaggio umano.

Amazon SageMaker è stato utilizzato per analizzare testi poetici e letterari in lingua inglese attraverso modelli NLP preaddestrati, con l'obiettivo di evidenziare il comportamento dell'Intelligenza Artificiale discriminativa nell'interpretazione di testi caratterizzati da un linguaggio allegorico, simbolico ed emotivamente complesso.

5.2 Sentiment Analysis ed Emotion Classification

La sentiment analysis è una tecnica di elaborazione del linguaggio naturale che permette di classificare automaticamente un testo in base al tono emotivo espresso.

Le tecniche di Natural Language Processing analizzano le parole presenti nel testo e attribuiscono un punteggio di sentiment che fornisce una misura della componente emotiva

espressa, positiva o negativa. Alcuni modelli sono in grado di rilevare le emozioni espresse in un testo, come rabbia, gioia o tristezza.

Tuttavia, l'Intelligenza Artificiale non sempre riesce a rilevare il tono con cui i concetti vengono espressi, mostrando alcuni limiti nel riconoscimento di elementi quali, ad esempio, il *sarcasmo*, la *negazione* e la *multipolarità*.

Il sarcasmo può essere interpretato come positivo a causa dei termini utilizzati; le negazioni possono non essere interpretate correttamente, soprattutto quando si sviluppano su due frasi; infine, la multipolarità può rendere più complessa l'interpretazione quando all'interno della stessa frase sono presenti sentiment differenti.

Il modello di *emotion classification multi-classe* permette di identificare le emozioni presenti nel documento fornito in input. Nel corso della sperimentazione, il modello utilizzato ha classificato le emozioni nelle seguenti categorie: paura, sorpresa, gioia, rabbia, neutrale, tristezza e disgusto. Tale approccio si differenzia dalla *sentiment analysis binaria* utilizzata nella sperimentazione, che restituisce una classificazione generale del testo come positiva o negativa.

5.3 Obiettivo e progettazione del case study

Per l'implementazione della fase di analisi automatica si è scelto di utilizzare Amazon SageMaker, un servizio di Amazon Web Services che consente di utilizzare modelli di Natural Language Processing per svolgere l'analisi di testi letterari.

Nell'ambito della sperimentazione sono stati utilizzati due modelli diversi: un modello di *sentiment analysis binaria*, che restituisce una classificazione del testo come *POSITIVE* o *NEGATIVE* associata a un determinato punteggio, e un modello di *emotion classification multi-classe*, che consente di individuare differenti componenti emotive presenti nel testo, insieme ai relativi punteggi.

L'obiettivo della sperimentazione è analizzare il comportamento dei modelli di Natural Language Processing nell'interpretazione di testi letterari e poetici caratterizzati da elementi allegorici, simbolici ed emotivamente complessi.

Nella sperimentazione è stato svolto un confronto fra i risultati ottenuti attraverso il modello di *sentiment analysis binaria* e il modello di *emotion classification multi-classe*.

A tale scopo sono stati selezionati diversi testi letterari e poetici in lingua inglese, caratterizzati dalla presenza di elementi quali metafore, simbolismo, significati impliciti e differenti componenti emotive. I risultati ottenuti dai modelli sono stati successivamente confrontati con l'interpretazione del testo dal punto di vista umano, evidenziando eventuali differenze tra la classificazione prodotta dall'Intelligenza Artificiale e il significato effettivo dell'opera.

5.4 Implementazione e sperimentazione

Per valutare il comportamento dei modelli Natural Language Processing sono stati selezionati estratti appartenenti a:

- poesie;
- testi allegorici;
- opere narrative.

Tra gli autori analizzati figurano:

- Emily Dickinson;

- Luis Sepúlveda;
- Edgar Allan Poe;
- William Wordsworth.

L'analisi ha evidenziato numerosi casi in cui l'Intelligenza Artificiale interpreta *letteralmente* metafore e simboli, semplifica emozioni complesse, non coglie il contesto narrativo complessivo oppure attribuisce emozioni differenti rispetto all'interpretazione umana. Questi risultati mostrano come la letteratura rappresenti un ambito particolarmente complesso per i sistemi di Natural Language Processing.

5.4.1 Emily Dickinson: The Battle-Field

Per il primo caso di studio è stata selezionata la poesia *The Battle-Field* di Emily Dickinson. Il testo presenta immagini inizialmente delicate e armoniose, come stelle, petali e vento estivo, che contrastano con il reale significato della poesia, legato alla morte e alla scomparsa dei soldati sul campo di battaglia.

L'obiettivo dell'analisi è verificare se i modelli Natural Language Processing siano in grado di cogliere il significato simbolico e malinconico del componimento oppure se vengano influenzati prevalentemente dai termini positivi presenti nel testo.

Nella Figura 5.1 viene riportata la poesia originale e la sua versione tradotta.

They dropped like Flakes -	Caddero come Fiocchi -
They dropped like Stars -	Caddero come Stelle -
Like Petals from a Rose -	Come Petali da una Rosa -
When suddenly across the June	Quando d'improvviso in Giugno
A Wind with fingers - goes -	Un Vento con le sue dita - passa -
They perished in the seamless Grass -	Perirono nell'Erba uniforme -
No eye could find the place -	Nessun occhio ne troverebbe il luogo -
But God can summon every face	Ma Dio può convocare ogni volto
On his Repealless - List.	Sulla sua Irrevocabile - Lista.

Figura 5.1: Estratto della poesia *The Battle-Field* di Emily Dickinson (Screen preso da Emilydickinson).

Per la prima parte della sperimentazione vengono analizzati i risultati ottenuti dal modello di *sentiment analysis binaria*. Nella Figura 5.2 viene riportato il risultato restituito da tale modello.

In questo caso, il modello ha restituito una classificazione *NEGATIVE* con un livello di confidenza pari a circa 0,77. Il risultato appare coerente con un'interpretazione generale del testo, caratterizzato dalla presenza di immagini che riconducono alla concezione di perdita, morte, dissoluzione e malinconia.

Per arrivare a questa classificazione, il modello potrebbe essere stato influenzato dall'utilizzo di espressioni dal significato fortemente negativo, come *perished* e *no eye*. Tuttavia, sebbene il livello di confidenza associato alla classe *NEGATIVE* sia elevato, il valore 0,77 mostra come la classificazione non sia completamente netta. Tale risultato potrebbe essere ricondotto alla complessità del testo, nel quale alle immagini negative si contrappone, soprattutto nel finale della poesia, una componente spirituale che introduce uno spiraglio di speranza.

```
text = """
They dropped like Flakes -
They dropped like Stars -
Like Petals from a Rose -
When suddenly across the June
A Wind with fingers - goes -

They perished in the seamless Grass -
No eye could find the place -
But God can summon every face
On his Repealless - List.
"""

result = classifier(text)

print(result)
```

```
[{'label': 'NEGATIVE', 'score': 0.7747423648834229}]
```

Figura 5.2: Sentiment analysis binaria *The Battle-Field* di Emily Dickinson.

La spiritualità assume un ruolo di fondamentale importanza nella parte conclusiva, in cui viene presentato il passaggio dalla dimensione terrena a quella ultraterrena. Nonostante molti soldati siano dispersi e i loro corpi non riceveranno mai sepoltura, Dio saprà ritrovarli e richiamarli, introducendo nella poesia una connotazione religiosa.

Trattandosi di un'analisi binaria, in grado di classificare il testo esclusivamente come *NEGATIVE* o *POSITIVE*, il modello consente di individuare l'orientamento generale del sentiment, senza fornire informazioni sulle diverse emozioni espresse o sulla complessità del significato del testo.

La classificazione ottenuta può, quindi, essere considerata coerente con un'interpretazione generale della poesia.

Successivamente è stato analizzato lo stesso estratto tramite un modello di *emotion classification multi-classe*, in grado di suddividere il contenuto testuale in differenti categorie emotive, tra cui: *anger*, *fear*, *sadness*, *joy*, *neutral* e *surprise*.

Nella Figura 5.3 viene riportato il risultato dello studio della emotion classification multi-classe della poesia.

```
text = """
They dropped like Flakes -
They dropped like Stars -
Like Petals from a Rose -
When suddenly across the June
A Wind with fingers - goes -

They perished in the seamless Grass -
No eye could find the place -
But God can summon every face
On his Repealless - List.
"""

result = emotion_classifier(text)

print(result)
```

```
[[{'label': 'anger', 'score': 0.07451540976762772}, {'label': 'disgust', 'score': 0.023805435746908188}, {'label': 'fear', 'score': 0.33048415184020996}, {'label': 'joy', 'score': 0.013669000007212162}, {'label': 'neutral', 'score': 0.12630777060985565}, {'label': 'sadness', 'score': 0.13418267667293549}, {'label': 'surprise', 'score': 0.2970355749130249}]]
```

Figura 5.3: Emotion classification multi-classe*The Battle-Field* di Emily Dickinson.

A differenza del modello binario precedente, il classificatore multi-classe ha restituito risultati più articolati. L'emozione predominante individuata dal modello è risultata essere *fear*, con uno score di circa 0,33, accompagnata da una componente significativa di *surprise*, con uno score di circa 0,29, da una presenza minore di *sadness*, con uno score di circa 0,13, e infine *neutral*, con uno score di circa 0,12.

Il modello sembra percepire nel testo una sensazione di inquietudine e smarrimento, data dall'elevata presenza di elementi quali *fear* e *sadness*, probabilmente influenzata da termini come *perished*, *no eye* e dall'idea della scomparsa improvvisa dei soggetti descritti nella poesia.

Sebbene il modello riesca a cogliere parte dell'atmosfera malinconica del componimento, non sembra comprendere pienamente il significato allegorico delle immagini utilizzate da Emily Dickinson. Elementi poetici come stelle, petali e vento potrebbero essere interpretati principalmente come termini descrittivi e non come simboli della fragilità della vita e della morte dei soldati sul campo di battaglia.

Il modello sembra aver colto, almeno in parte, la conclusione spirituale della poesia, nella quale la perdita fisica viene superata attraverso l'idea di una vita eterna e di un richiamo divino. La presenza significativa di *surprise* potrebbe essere ricondotta al cambiamento introdotto nella parte conclusiva.

Il confronto tra i due modelli evidenzia, quindi, come l'Intelligenza Artificiale discriminativa riesca ad associare emozioni a specifici pattern linguistici, ma presenti notevoli difficoltà nella comprensione profonda del linguaggio poetico, simbolico e metaforico.

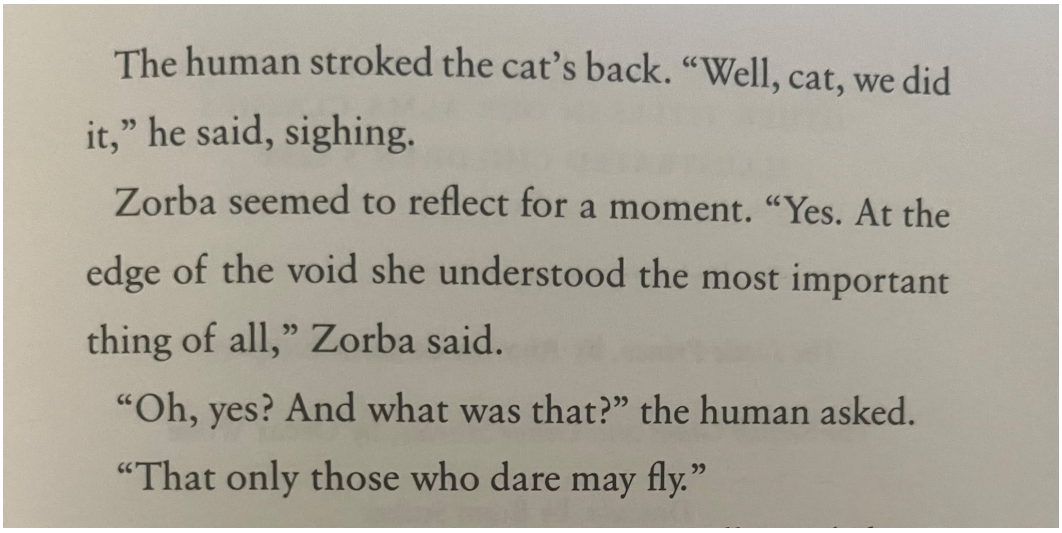
Nella Tabella 5.1 vengono confrontati i due modelli.

	Sentiment analysis binaria	**Emotion classification multi-classe**
Classificazione	Individuazione dell'orientamento generale del testo	Individuazione di più componenti emotive
Output ottenuto	NEGATIVE 0,77	Fear 0,33; Surprise 0,29; Sadness 0,13; Neutral 0,12
Interpretazione	Riconoscimento dell'orientamento derivato dall'analisi di termini associati alla perdita e alla morte	Riconoscimento di un'atmosfera di inquietudine e smarrimento
Aspetti rilevati	Il livello di confidenza del testo pari a 0,77, non completamente netto, potrebbe essere dovuto alla presenza di alcuni elementi contrastanti	Il riconoscimento di diversi tipi di emozioni nel testo ne sottolinea la complessità emotiva
Limiti rilevati	Non distingue le singole emozioni presenti nel testo e non sembra coglierne pienamente il significato metaforico	Distingue diverse componenti emotive, ma non sembra riconoscere pienamente il significato simbolico e allegorico del testo poetico

Tabella 5.1: Confronto tra la sentiment analysis binaria e l'emotion classification multi-classe di The Battle-Field

5.4.2 Luis Sepúlveda: The Story of a Seagull and the Cat Who Taught Her to Fly

Per il secondo caso di studio è stato selezionato un estratto tratto dal romanzo *The Story of a Seagull and the Cat Who Taught Her to Fly* di Luis Sepúlveda. Nella Figura 5.4 viene riportato un estratto di tale romanzo.



The human stroked the cat's back. "Well, cat, we did it," he said, sighing.

Zorba seemed to reflect for a moment. "Yes. At the edge of the void she understood the most important thing of all," Zorba said.

"Oh, yes? And what was that?" the human asked.

"That only those who dare may fly."

Figura 5.4: Estratto del romanzo *The Story of a Seagull and the Cat Who Taught Her to Fly* di Luis Sepúlveda

Il testo rappresenta uno dei momenti conclusivi dell'opera, nel quale la giovane gabbianella ha finalmente imparato a volare e Zorba (*il gatto*) la vede allontanarsi. L'estratto presenta un linguaggio apparentemente semplice e positivo, ma racchiude significati simbolici profondi legati alla crescita personale e alla libertà.

Nella Figura 5.5 viene riportato il risultato dello studio della sentiment analysis binaria dell'estratto.

```
text = """
At the edge of the void she understood the most important thing of all.

That only those who dare may fly.
"""

result = classifier(text)

print(result)
```

```
[{'label': 'NEGATIVE', 'score': 0.9798765182495117}]
```

Figura 5.5: Sentiment analysis binaria *The Story of a Seagull and the Cat Who Taught Her to Fly* di Luis Sepúlveda.

Tramite un'analisi umana, il testo verrebbe classificato come positivo, in quanto, al suo interno il messaggio è legato al coraggio, alla crescita personale, alla libertà e al superamento delle proprie paure. L'allegoria del volo rappresenta, infatti, la capacità di affrontare il rischio e superare i propri limiti interiori.

Tuttavia, il modello di sentiment analysis binaria ha classificato il testo come *NEGATIVE*, con un livello di confidenza pari a circa 0,97, quindi con una concezione estremamente negativa del testo analizzato. Il risultato evidenzia un limite significativo dell'Intelligenza Artificiale discriminativa applicata ai testi allegorici. Il modello potrebbe essere stato influenzato dalla presenza di singole parole potenzialmente associate a sentiment negativi, come *void* o *dare*, senza comprendere il significato simbolico molto positivo che, in realtà, queste esprimono nel testo.

In particolare, il sistema ignora:

- il significato allegorico del volo;
- il messaggio simbolico del superamento delle paure;
- il contesto emotivo positivo presente nella conclusione del brano.

Questo comportamento evidenzia come i sistemi tradizionali di sentiment analysis possano presentare notevoli difficoltà nell'interpretazione di:

- metafore;
- linguaggio figurato;
- significati impliciti;
- contenuti emotivamente complessi.

L'analisi conferma, quindi, che il riconoscimento del lessico non coincide necessariamente con una reale comprensione semantica ed emotiva del testo letterario.

Successivamente, lo stesso estratto è stato analizzato tramite un modello di *emotion classification multi-classe*. Nella Figura 5.6 viene riportato il risultato restituito da tale modello.

```
text = """
At the edge of the void she understood the most important thing of all.

That only those who dare may fly.
"""

result = emotion_classifier(text)

print(result)
```

```
[[{'label': 'anger', 'score': 0.2874053418636322}, {'label': 'disgust', 'score': 0.014599808491766453}, {'label': 'fear', 'score': 0.6726841330528259},
{'label': 'joy', 'score': 0.002079102909192443}, {'label': 'neutral', 'score': 0.013566464185714722}, {'label': 'sadness', 'score': 0.00518067460507154
5}, {'label': 'surprise', 'score': 0.004484476987272501}]]
```

Figura 5.6: Emotion classification multi-classe *The Story of a Seagull and the Cat Who Taught Her to Fly* di Luis Sepúlveda.

A differenza del modello di *sentiment analysis binaria*, il classificatore multi-classe ha restituito un risultato più articolato. L'emozione predominante è risultata essere *fear*, con uno score pari a circa 0,67, seguita da una componente significativa di *anger*, con uno score pari a circa 0,28.

Il risultato appare particolarmente interessante poiché, dal punto di vista umano, il testo viene generalmente interpretato come positivo e motivazionale. Sebbene la paura possa essere un'emozione associata alle nuove esperienze, e possa, quindi, rappresentare un'emozione in parte coerente con il significato del testo, lo score ottenuto risulta particolarmente elevato. Diversamente, la presenza significativa di *anger* appare difficilmente riconducibile al significato complessivo del componimento.

Il modello potrebbe essere stato influenzato dalla presenza di termini come *void* e *dare*, associandoli a emozioni negative e interpretando il testo come una situazione caratterizzata da paura e tensione.

Il confronto tra i due modelli mostra come l'Intelligenza Artificiale discriminativa riesca a riconoscere pattern emotivi superficiali, ma presenti notevoli difficoltà nella comprensione semantica profonda dei testi allegorici e letterari.

Nella Tabella 5.3 vengono confrontati i due modelli.

	Sentiment analysis binaria	**Emotion classification multi-classe**
Classificazione	Individuazione dell'orientamento generale del testo	Individuazione di più componenti emotive
Output ottenuto	NEGATIVE 0,97	Fear 0,67; Anger 0,28
Interpretazione	Riconoscimento di termini associati a vuoto e rischio	Riconoscimento di un'atmosfera di paura e tensione
Aspetti rilevati	L'elevato livello di confidenza potrebbe essere legato alla presenza di termini come void e dare, associati a sentiment negativi	Il riconoscimento di fear risulta in parte coerente con il testo, mentre la componente significativa di anger appare difficilmente riconducibile al suo significato complessivo
Limiti rilevati	Non comprende allegoria e messaggio simbolico	Associa le emozioni al linguaggio ma non ne coglie il significato profondo

Tabella 5.2: Confronto tra la sentiment analysis binaria e l'emotion classification multi-classe The Story of a Seagull and the Cat Who Taught Her to Fly

5.4.3 Edgar Allan Poe: Annabel Lee

Per il terzo caso di studio è stato selezionato un estratto della poesia *Annabel Lee* di Edgar Allan Poe mostrato in Figura 5.7.

It was many and many a year ago,	*Or son molti e molti anni,*
In a kingdom by the sea,	*In un regno in riva al mare,*
That a maiden there lived whom you may know	*Là viveva una fanciulla che col nome puoi chiamare*
By the name of Annabel Lee;	*Di Annabel Lee;*
And this maiden she lived with no other thought	*E questa fanciulla viveva e soltanto pensava*
Than to love and be loved by me.	*Ad amare e ad essere amata da me.*

Figura 5.7: Estratto della poesia *Annabel Lee* di Edgar Allan Poe (Testo preso da Poetryfoundation. Traduzione presa da Oubliettemagazine).

Il testo presenta inizialmente immagini romantiche e apparentemente serene, come il regno sul mare e l'amore tra i due protagonisti. L'estratto scelto per la sperimentazione risulta particolarmente complesso da analizzare, poiché combina elementi che potrebbero portare a interpretazioni differenti. Da un lato sono presenti immagini dolci, legate all'amore tra i due protagonisti; il riferimento a un tempo ormai lontano e l'utilizzo del passato introducono una componente nostalgica. Inoltre, il ritmo della poesia richiama quello di una narrazione fiabesca che contribuisce a rendere meno immediata la percezione della componente malinconica del testo.

L'obiettivo dell'analisi è verificare se i modelli Natural Language Processing siano in grado di distinguere tra il tono apparentemente positivo del linguaggio e il significato malinconico del componimento.

Inizialmente sono stati analizzati i risultati ottenuti tramite un modello di *sentiment analysis binaria*.

Nella Figura 5.8 viene riportato il risultato dello studio della sentiment analysis binaria della poesia.

```
text = """
It was many and many a year ago,
In a kingdom by the sea,
That a maiden there lived whom you may know
By the name of Annabel Lee;

And this maiden she lived with no other thought
Than to love and be loved by me.
"""

result = classifier(text)

print(result)
```

```
[{'label': 'POSITIVE', 'score': 0.9976068735122681}]
```

Figura 5.8: Sentiment analysis binaria *Annabel Lee* di Edgar Allan Poe.

Il modello ha classificato il testo come *POSITIVE*, con un livello di confidenza pari a circa 0,99. Il risultato dimostra come, nel caso analizzato, la sentiment analysis restituisca una classificazione coerente con il tono apparentemente positivo dell'estratto, senza, però, comprendere la componente malinconica che caratterizza il componimento.

Il modello potrebbe essere stato influenzato prevalentemente dalla presenza di termini associati all'amore e all'affetto, come *love* e *maiden*, nonchè dalla struttura romantica del linguaggio. Nell'estratto analizzato il tema della morte non viene toccato esplicitamente, ma emerge considerando il contesto complessivo della poesia.

Questo comportamento evidenzia uno dei limiti che può emergere nell'applicazione della sentiment analysis binaria a testi letterari caratterizzati da ambiguità nella descrizione delle differenti sfumature emotive. La classificazione in sole due categorie può, infatti, non essere sufficiente a rappresentare pienamente la complessità emotiva del testo.

Per questo motivo, si è deciso di analizzare lo stesso estratto mediante un modello di *emotion classification multi-classe*, in grado di classificare il contenuto testuale in differenti categorie emotive.

Nella Figura 5.9 viene riportato il risultato dello studio dell'emotion classification multi-classe dell'estratto.

```
text = """
At the edge of the void she understood the most important thing of all.

That only those who dare may fly.
"""

result = emotion_classifier(text)

print(result)
```

```
[[{'label': 'anger', 'score': 0.2874053418636322}, {'label': 'disgust', 'score': 0.014599808491766453}, {'label': 'fear', 'score': 0.6726841330528259},
{'label': 'joy', 'score': 0.002079102909192443}, {'label': 'neutral', 'score': 0.013566464185714722}, {'label': 'sadness', 'score': 0.00518067460507154
5}, {'label': 'surprise', 'score': 0.004484476987272501}]]
```

Figura 5.9: Emotion classification multi-classe *Annabel Lee* di Edgar Allan Poe.

A differenza del modello binario, il classificatore multi-classe ha restituito risultati più articolati. L'emozione predominante identificata dal modello è risultata essere *neutral*, con uno score pari a circa 0,64, accompagnata da componenti minori di *surprise* e *joy*, con score, rispettivamente, pari a circa 0,13 e 0,11.

Il risultato evidenzia come il modello di *emotion classification multi-classe* riesca a percepire una certa ambiguità emotiva del testo. Il linguaggio fiabesco del componimento potrebbe influenzare il comportamento del modello, che tende a interpretare il testo come emotivamente neutro o leggermente positivo. Tuttavia, i livelli associati alle emozioni positive risultano piuttosto ridotti; ciò lascia intendere che il modello di Natural Language Processing, sulla base del solo estratto analizzato, non riesca a individuare con chiarezza la direzione emotiva del componimento poetico. L'assenza del contesto complessivo potrebbe rendere l'interpretazione meno immediata anche per un essere umano, il quale potrebbe, tuttavia, cogliere la componente nostalgica dell'estratto attraverso riferimenti a un tempo lontano, e l'utilizzo del passato, associandovi una sfumatura di tristezza.

Nel caso analizzato emerge una difficoltà del modello nell'interpretazione delle sfumature emotive del linguaggio poetico. Sebbene il *classificatore multi-classe* restituisca una rappresentazione più articolata rispetto alla *sentiment analysis binaria*, le emozioni individuate non sembrano riconoscere pienamente la componente nostalgica dell'estratto.

Nella Tabella 5.3 vengono confrontati i due modelli.

	Sentiment analysis binaria	**Emotion classification multi-classe**
Classificazione	Individuazione dell'orientamento generale del testo	Individuazione di più componenti emotive
Output ottenuto	POSITIVE 0,99	Neutral 0,64; Surprise 0,13; Joy 0,11
Interpretazione	Riconoscimento del tono, all'apparenza positivo, del testo poetico	Riconoscimento di ambiguità emotiva, con prevalenza della componente neutra
Aspetti rilevati	L'elevato livello di confidenza potrebbe essere dovuto ai termini legati all'amore e all'affettto e al linguaggio fiabesco del componimento	Il riconoscimento di diversi tipi di emozioni nel testo restituisce una rappresentazione più articolata dell'estratto
Limiti rilevati	Non coglie la componente malinconica del componimento	Non individua con chiarezza la direzione emotiva del componimento

Tabella 5.3: Confronto tra la sentiment analysis binaria e l'emotion classification multi-classe Annabel Lee

5.4.4 William Wordsworth: My Heart Leaps Up

Per il quarto caso di studio è stato selezionato un estratto della poesia *My Heart Leaps Up* di William Wordsworth mostrato in Figura 5.10.

Il testo presenta immagini estremamente positive e armoniose, legate alla natura, alla meraviglia e all'infanzia. L'arcobaleno può essere interpretato come simbolo di continuità emotiva e spirituale tra le diverse fasi della vita umana.

Tuttavia, dietro queste immagini gioiose, il componimento racchiude riflessioni profonde sul trascorrere del tempo, sull'invecchiamento e sulla paura di perdere la capacità di provare stupore di fronte al mondo naturale.

L'obiettivo dell'analisi è verificare se i modelli Natural Language Processing siano in grado di cogliere la componente filosofica ed esistenziale della poesia.

My heart leaps up when I behold
A rainbow in the sky:

So was it when my life began;
So is it now I am a man;
So be it when I shall grow old,
Or let me die!

The Child is father of the Man;
And I could wish my days to be
Bound each to each by natural piety.

Il mio cuore sobbalza quando vedo
Un arcobaleno nel cielo:

Così fu quando la mia vita cominciò;
Così è adesso che sono un uomo;
Così sia quando sarò diventato vecchio,
O fatemi morire!

Il Bambino è padre dell'uomo;
E io potrei desiderare che i miei giorni fossero
Legati l'uno all'altro dalla devozione per la natura.

Figura 5.10: Estratto della poesia *Annabel Lee* di Edgar Allan Poe (Testo preso da Poets. Traduzione presa da Isoladellapoesia).

Inizialmente sono stati analizzati i risultati ottenuti tramite un modello di *sentiment analysis binaria*. Nella Figura 5.11 viene riportato il risultato di tale modello.

```
text = """
My heart leaps up when I behold
A rainbow in the sky:

So was it when my life began;
So is it now I am a man;
So be it when I shall grow old,
Or let me die!

The Child is father of the Man;
And I could wish my days to be
Bound each to each by natural piety.
"""

result = classifier(text)

print(result)
```

```
[{'label': 'POSITIVE', 'score': 0.9981753826141357}]
```

Figura 5.11: Sentiment analysis binaria *My Heart Leaps Up* di William Wordsworth.

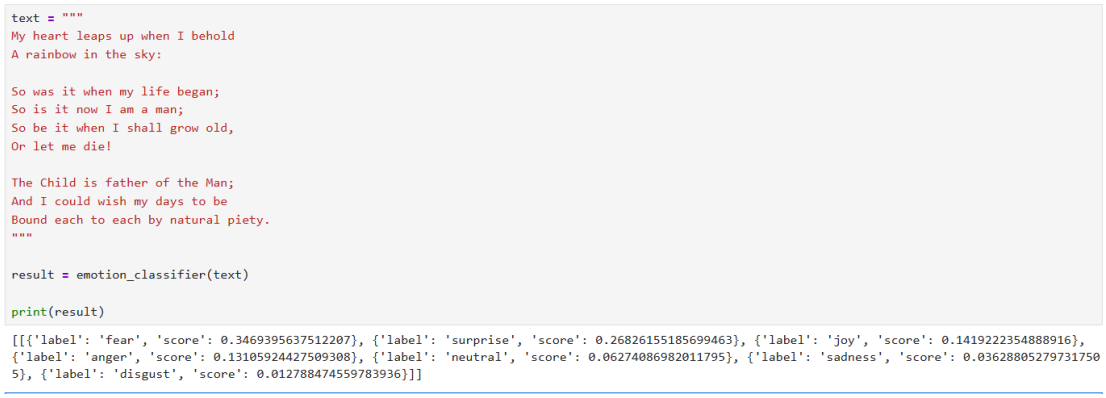
Il modello di Natural Language Processing utilizzato ha classificato il testo come *POSITIVE*, con un livello di confidenza estremamente elevato, pari a circa 0,99. Il risultato del modello di sentiment analysis binaria risulta coerente con il significato apparentemente positivo del testo.

Il componimento trasmette inizialmente un forte senso di meraviglia, armonia e legame con la natura. L'immagine dell'arcobaleno e l'espressione *"My heart leaps up"* suggeriscono gioia, stupore e contemplazione positiva della realtà naturale.

Tuttavia, il componimento presenta anche una significativa componente esistenziale e malinconica, non rilevata dal sistema. Il riferimento al trascorrere del tempo, alla vecchiaia, e soprattutto il verso *"Or let me die!"* introducono riflessioni profonde sulla mortalità e sul timore di perdere la capacità di provare meraviglia di fronte al mondo naturale.

Il risultato ottenuto evidenzia come, nel caso analizzato, la classificazione binaria riesca a individuare correttamente l'orientamento prevalentemente positivo del testo, senza tuttavia rappresentare le ulteriori sfumature emotive legate alla componente esistenziale del componimento.

Successivamente, lo stesso estratto è stato analizzato tramite un modello di *emotion classification multi-classe*, capace di classificare il contenuto testuale in differenti categorie emotive. Nella Figura 5.12 viene riportato il risultato di tale modello.



```
text = """
My heart leaps up when I behold
A rainbow in the sky:

So was it when my life began;
So is it now I am a man;
So be it when I shall grow old,
Or let me die!

The Child is father of the Man;
And I could wish my days to be
Bound each to each by natural piety.
"""

result = emotion_classifier(text)

print(result)
```

```
[[{'label': 'fear', 'score': 0.3469395637512207}, {'label': 'surprise', 'score': 0.2682615518569946 3}, {'label': 'joy', 'score': 0.1419222354888916},
{'label': 'anger', 'score': 0.13105924427509308}, {'label': 'neutral', 'score': 0.06274086982011795}, {'label': 'sadness', 'score': 0.03628805279731750
5}, {'label': 'disgust', 'score': 0.012788474559783936}]]
```

Figura 5.12: Emotion classification multi-classe *My Heart Leaps Up* di William Wordsworth.

A differenza del modello binario, il classificatore multi-classe ha restituito una distribuzione emotiva più articolata. L'emozione predominante identificata dal modello è risultata essere *fear*, con uno score pari a circa 0,35, seguita da componenti significative di *surprise*, *joy* e *anger*, con score, rispettivamente, pari a circa 0,27, 0,14 e 0,13.

Il modello sembra percepire la presenza di emozioni contrastanti. In particolare, l'espressione *"Or let me die!"* potrebbe aver contribuito alla presenza di emozioni negative, tra cui *fear* e *anger*. Allo stesso tempo, elementi naturali e positivi, come *rainbow*, e l'espressione *"My heart leaps up"* potrebbero aver contribuito alla presenza della componente *joy*.

L'analisi evidenzia come, nel caso considerato, il modello di Natural Language Processing riesca a individuare differenti componenti emotive presenti nel testo. Tuttavia, la distribuzione ottenuta non consente di affermare che il sistema abbia effettivamente compreso il significato simbolico ed esistenziale del componimento.

In questo caso, la differenza fra i due modelli risulta particolarmente evidente: mentre la *sentiment analysis binaria* restituisce soltanto la connotazione positiva della poesia, l'*emotion classification multi-classe* riesce a rilevare anche la componente di inquietudine presente nel componimento.

Nella Tabella 5.4 vengono confrontati i due modelli.

5.4.5 Analisi dei risultati

Dall'analisi dei risultati emerge come i modelli di Natural Language Processing utilizzati tramite Amazon SageMaker riescano a individuare differenti elementi emotivi presenti nei testi, mostrando, tuttavia, alcuni limiti nell'interpretazione delle sfumature emotive e del significato complessivo dei contenuti analizzati.

I diversi scenari sono stati selezionati con l'obiettivo di studiare il comportamento dei modelli di sentiment analysis nell'analisi di differenti tipologie di testo. In particolare, la

	Sentiment analysis binaria	**Emotion classification multi-classe**
Classificazione	Individuazione dell'orientamento generale del testo	Individuazione di più componenti emotive
Output ottenuto	POSITIVE 0,99	Fear 0,35; Surprise 0,27; Joy 0,14; Anger 0,13
Interpretazione	Riconoscimento della connotazione positiva della poesia	Riconoscimento della presenza di diverse emozioni
Aspetti rilevati	L'elevato livello di confidenza risulta coerente con la presenza di elementi positivi legati alla natura e alla meraviglia	La presenza di fear, surprise, joy e anger restituisce una rappresentazione emotiva più articolata
Limiti rilevati	Non rappresenta le ulteriori sfumature presenti nel componimento	Il modello rileva differenti componenti emotive, ma non si è in grado di stabilire se il modello abbia compreso a pieno il significato della poesia

Tabella 5.4: Confronto tra la sentiment analysis binaria e l'emotion classification multi-classe My Heart Leaps Up

sperimentazione è stata svolta confrontando i risultati ottenuti tramite un modello di *sentiment analysis binaria* e un modello di *emotion classification multi-classe*.

Nella *sentiment analysis binaria* nei casi di *The Battle-Field* e *My Heart Leaps Up*, la classificazione restituita dal modello risulta essere coerente con l'orientamento del componimento, seppur in maniera riduttiva. Negli altri casi, invece, la classificazione nelle sole categorie di *POSITIVE* e *NEGATIVE* non è stata in grado di rappresentarne la complessità emotiva. I risultati evidenziano maggiori difficoltà nell'interpretazione dei testi caratterizzati da simbolismo, ambiguità emotiva e significati impliciti. Un esempio è rappresentato da *Annabel Lee*: l'estratto analizzato, non contenendo riferimenti espliciti alla morte, porta la sentiment analysis binaria a riconoscerne prevalentemente gli elementi positivi presenti nel testo, classificandolo come *POSITIVE*.

La sperimentazione evidenzia come il modello di *emotion classification multi-classe* abbia restituito una rappresentazione più articolata dei contenuti analizzati, evidenziando la presenza simultanea di diverse componenti emotive. Tuttavia, la presenza di un maggior numero di categorie emotive non implica necessariamente una comprensione più profonda del simbolismo e del significato letterario del testo, come osservato nell'analisi dell'estratto di Luis Sepúlveda.

CAPITOLO 6

Discussione

Questo capitolo propone una discussione dei case study trattati in precedenza, offrendo una visione più concreta di come gli errori nei sistemi di Intelligenza Artificiale possano avere conseguenze reali sulla vita delle persone. Si parte dall'analisi di alcuni casi legati alla Content Moderation nei Social Media, in cui gli errori nei sistemi possono portare al mancato riconoscimento di contenuti problematici. Successivamente, vengono affrontate le problematiche giuridiche connesse all'utilizzo della Facial Analysis nell'ambito delle indagini. Infine, viene analizzato il problema della comprensione e del riconoscimento delle emozioni da parte dell'Intelligenza Artificiale.

6.1 Discussione sul Case Study 1: i limiti nella Content Moderation nei Social Media

I risultati riportati nel terzo capitolo mostrano come Amazon Web Services si basi prevalentemente sull'analisi visiva, presentando alcuni limiti nell'interpretazione di contenuti ambigui. Al contrario, piattaforme come Meta e TikTok adottano approcci *multimodali* [1] che integrano differenti tipologie di informazioni, come elementi visivi, testuali e audio.

Sebbene tali sistemi di Intelligenza Artificiale risultino essere più avanzati, ciò non implica che siano privi di un margine di errore. Nel corso degli anni, numerosi casi di cronaca hanno evidenziato criticità nei processi di moderazione dei contenuti in rete. In particolare, si riscontrano fenomeni di *falsi positivi*, in cui contenuti leciti vengono erroneamente censurati, e di *falsi negativi*, in cui contenuti potenzialmente pericolosi non vengono tempestivamente individuati o rimossi.

In questa sezione verranno descritti alcuni casi di cronaca che dimostrano come i sistemi di *Content Moderation* possano non riconoscere correttamente determinate problematiche, evidenziando i limiti intrinseci di tali tecnologie.

Tali problematiche non iniziano con le piattaforme Meta e TikTok, ma hanno radici pregresse. Già intorno al 2016 divenne nota la cosiddetta *Blue Whale Challenge*. Il caso venne associato principalmente a forum online, gruppi social e piattaforme digitali precedenti agli attuali sistemi di moderazione avanzata.

[1] L'*Intelligenza Artificiale multimodale* fa riferimento a modelli di *Machine Learning* in grado di elaborare e integrare informazioni provenienti da più modalità o tipologie di dati, combinando diversi input al fine di ottenere una comprensione più completa e generare output più accurati (definizione ricavata da IBM).

La challenge consisteva, secondo quanto riportato dai media, in una serie di prove progressive assegnate agli utenti da presunti *curatori*, che includevano attività psicologicamente dannose, isolamento sociale, autolesionismo e, nei casi più estremi, comportamenti suicidari. Il fenomeno suscitò un forte dibattito internazionale riguardo ai rischi legati alla diffusione incontrollata di contenuti online e alla difficoltà delle piattaforme digitali nell'individuare tempestivamente contenuti pericolosi o manipolativi.

6.1.1 Il caso Meta

Le piattaforme Meta, come Facebook e Instagram, hanno mostrato negli anni alcune criticità nei sistemi di moderazione automatica dei contenuti. Nonostante l'utilizzo di sistemi avanzati di Intelligenza Artificiale e moderazione multimodale, diversi casi di cronaca hanno evidenziato la presenza di errori sia nell'identificazione sia nella gestione di contenuti problematici.

Uno dei principali limiti riguarda la difficoltà nel distinguere correttamente contenuti realmente pericolosi da contenuti leciti. In alcuni casi, i post dannosi non vengono rimossi tempestivamente, causando *falsi negativi*, mentre in altri contenuti leciti vengono erroneamente limitati o censurati causando *falsi positivi*.

Meta stessa ha ammesso la presenza di errori nei propri sistemi di moderazione automatica, riconoscendo difficoltà nella gestione di contenuti complessi o ambigui, soprattutto nei casi in cui il significato dipende fortemente dal contesto del contenuto pubblicato. In un articolo di *TechTalking* del 2024 viene riportata l'ammissione di Nick Cleg, presidente degli Affari globali di Meta, il quale riconobbe un elevato tasso di errore nella moderazione dei contenuti, con la conseguente rimozione di post innocui.

Invece, in un articolo del *Corriere della Sera* del 2025 viene riportato un discorso di Mark Zuckerberg nel quale viene annunciata una riduzione nella complessità dei sistemi di moderazione a causa dei numerosi errori riscontrati nella rimozione dei contenuti.

Tra le varie problematiche riscontrate nel riconoscimento di falsi positivi e falsi negativi sono presenti diversi esempi:

- La *generazione di immagini sessualizzate con l'utilizzo di altre Intelligenze Artificiali*. Nel corso degli anni sono state effettuate diverse segnalazioni relative a immagini sessualizzate generate tramite Intelligenza Artificiale e non rimosse dalla *Content Moderation* di Meta.

  Un caso importante risale al 2025, quando su Instagram venne pubblicato un video apparentemente generato tramite Intelligenza Artificiale che impersonava una donna. Il video venne individuato dai sistemi automatici di Meta come potenzialmente dannoso e con un'elevata probabilità di diventare virale, ma non venne considerato prioritario e, quindi, non sottoposto a revisione umana. Successivamente, il video venne segnalato da due utenti, ma non venne, comunque, eliminato dalla piattaforma.

  Uno di questi utenti presentò ricorso a Meta, ma il contenuto non venne sottoposto a revisione. L'utente decise, quindi, di rivolgersi all'*Oversight Board* [2], dichiarando che la donna rappresentata nel video era una sua amica, che il contenuto era stato generato senza il suo consenso e che la donna aveva già chiuso il proprio account.

  Tuttavia, quando il caso venne portato all'attenzione di Meta dall'*Oversight Board*, gli esperti stabilirono che, secondo le normative della piattaforma, non fosse necessaria la rimozione completa del contenuto ma fosse sufficiente limitarne la visione agli utenti adulti. L'*Oversight Board* non condivise tale decisione, ritenendo che il contenuto

[2] L'*Oversight Board* è stato creato per aiutare Meta ad affrontare alcune tra le questioni più difficili in merito al tema della libertà di espressione online; esso decide cosa rimuovere, cosa lasciare pubblicato e perché (definizione ricavata da Transparency).

violasse le politiche relative alla condivisione di immagini intime non consensuali, e ribaltò la decisione di Meta.

La diffusione di contenuti sessualizzati generati tramite Intelligenza Artificiale rappresenta una problematica reale. Alcuni ricercatori nell'ambito degli abusi sessuali online sottolineano come i danni provocati dalla diffusione di immagini intime *deepfake* possano essere tanto gravi quanto quelli associati alla diffusione senza consenso di immagini sessuali autentiche.

- Un'ulteriore problematica riguarda la diffusione di *deepfake* e di contenuti manipolati in ambito politico. Nel corso degli anni sono stati diffusi dei video modificati sia attraverso l'utilizzo dell'Intelligenza Artificiale sia attraverso strumenti di editing, con il rischio di ingannare gli utenti. Un esempio riguarda un video del presidente statunitense Joe Biden, modificato attraverso tecniche di editing in modo da far sembrare che stesse toccando in maniera inappropriata la propria nipote. In questo caso, il video non era stato modificato attraverso l'utilizzo dell'Intelligenza Artificiale, ma tramite dei tagli e dei loop di alcune parti del filmato originale. Questo caso evidenzia un'ulteriore difficoltà nella moderazione di contenuti manipolati.

- Un ulteriore problema presente nelle piattaforme Meta riguarda la *disparità linguistica*. Le risorse destinate alla moderazione dei contenuti non risultano distribuite uniformemente tra le diverse lingue. Tale disparità potrebbe favorire una maggiore diffusione di informazioni false o fuorvianti nelle lingue che dispongono di minori risorse rispetto all'inglese. Nell'ambito della disinformazione riguardante le notizie relative al Covid-19 è stato evidenziato come la maggior parte delle lingue diverse dall'inglese disponga di una copertura di *fact-checking* significativamente inferiore. Inoltre, sono state sollevate criticità riguardo alla capacità degli strumenti automatici di Meta di interpretare correttamente il contesto, soprattutto nei contenuti in lingue diverse dall'inglese. Questa disparità può, quindi, determinare differenze nell'efficacia dei sistemi di moderazione automatica e aumentare il rischio di errori nell'identificazione dei contenuti.

6.1.2 Il caso TikTok

Tra i casi più discussi di errori nella *Content Moderation* troviamo le cosiddette *dangerous challenges* diffuse sulla piattaforma TikTok. In particolare, la *Blackout Challenge* ha attirato l'attenzione mediatica internazionale dopo essere stata collegata alla morte di alcuni minori. Questi casi hanno sollevato interrogativi sull'efficacia dei sistemi adottati dalla piattaforma nell'individuazione e nella limitazione tempestiva di contenuti potenzialmente dannosi, evidenziando le difficoltà nel riconoscere situazioni pericolose in cui il rischio è legato ai comportamenti che potrebbero essere incentivati dal contenuto, piuttosto che alla presenza di violenza esplicita.

La Blackout Challenge consiste in una pratica basata sul soffocamento volontario fino allo svenimento. Il gioco può essere praticato in solitudine, tramite, ad esempio, l'utilizzo di corde, oppure con l'aiuto di un'altra persona. Lo scopo è quello di sperimentare gli effetti provocati dalla mancanza di ossigeno fino alla perdita di coscienza, con l'intento di comprendere cosa si prova in prossimità della morte. La sfida si conclude con lo svenimento e vince chi riprende conoscenza per primo.

Fenomeni simili sono stati osservati anche nel *Fainting Game* e nel *Choking Game*, pratiche basate sul soffocamento volontario temporaneo allo scopo di provocare perdita di coscienza o sensazioni di euforia.

Questi casi evidenziano come il pericolo associato ai contenuti online non dipenda esclusivamente dalla presenza di elementi visivamente violenti, ma anche dal contesto,

dal significato implicito del contenuto e dai comportamenti che esso può incentivare negli utenti.

Ulteriori criticità riguardanti la *Content Moderation* di TikTok emergono nell'ambito della disinformazione sanitaria e psicologica. In un'indagine pubblicata dal quotidiano *The Guardian* sono stati analizzati cento tra i video più popolari relativi alla salute mentale presenti su TikTok. Dall'analisi effettuata con il supporto di esperti è emerso che 52 dei 100 video esaminati contenevano forme di disinformazione.

Questo evidenzia le difficoltà che possono emergere nella moderazione dei contenuti, soprattutto quando il rischio non deriva dagli elementi visivi ma dal significato delle informazioni trasmesse, aumentando la possibilità di diffusione di informazioni errate o fuorvianti.

Le problematiche legate alla moderazione dei contenuti hanno portato anche a interventi delle autorità competenti. In Italia, nel 2024, TikTok è stato sanzionato dall'Antitrust per 10 milioni di euro per controlli ritenuti inadeguati nella tutela degli utenti minori e vulnerabili rispetto alla diffusione di contenuti potenzialmente pericolosi. Tra questi viene richiamata la cosiddetta *cicatrice francese*, una challenge che consisteva nel procurarsi segni sul volto attraverso il pizzicamento della pelle. Tali episodi evidenziano i limiti che possono emergere nei sistemi di controllo e moderazione dei contenuti pubblicati online.

In alcuni casi TikTok applica sistemi di avviso preventivo (*warning labels*) per contenuti considerati sensibili o potenzialmente disturbanti. In questi casi, il contenuto non viene necessariamente rimosso, ma la piattaforma limita l'accesso tramite schermate di avviso che informano l'utente della natura del contenuto. Tale approccio evidenzia come le moderne piattaforme social non si limitino alla semplice rimozione dei contenuti, ma adottino ulteriori strategie intermedie basate sulla contestualizzazione e sulla limitazione della distribuzione dei contenuti.

Nella Figura 6.1 viene riportato un esempio di *warning label* di TikTok.

Questi casi dimostrano come la Content Moderation non possa basarsi esclusivamente su classificazioni automatiche rigide, ma richieda una valutazione contestuale più complessa, spesso supportata da revisione umana e da sistemi multimodali avanzati.

6.1.3 Conclusioni

I casi di cronaca descritti dimostrano come la Content Moderation presenti ancora dei limiti significativi. Le problematiche riscontrate con TikTok mostrano come un video non verificato correttamente possa avere conseguenze tragiche. Invece, tramite i casi di Meta si è visto come l'utilizzo dell'Intelligenza Artificiale o di altre pratiche per modificare i video possano danneggiare l'immagine di una persona. Sebbene l'analisi dei contenuti sia una componente fondamentale per avere un ambiente social sicuro, le attuali tecnologie, basate sulle linee guida, non risultano ancora in grado di valutare sempre correttamente il contesto effettivo in cui un contenuto viene pubblicato.

6.2 Discussione sul Case Study 2: riconoscimento facciale nelle indagini

I risultati riportati nel quarto capitolo mostrano come Amazon Web Services si basi prevalentemente sull'analisi visiva, presentando alcuni limiti nell'interpretazione di somiglianze ambigue. Tuttavia, nelle indagini reali il riconoscimento facciale può essere applicato a immagini provenienti da fonti differenti e caratterizzate da condizioni di acquisizione eterogenee.

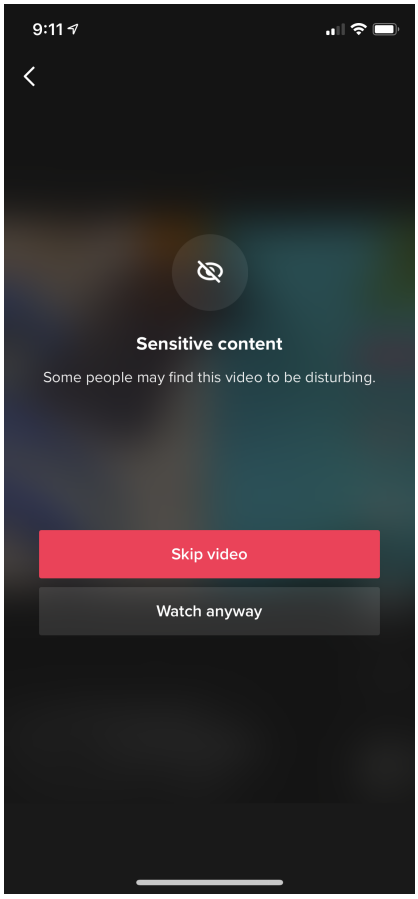


Figura 6.1: Esempio di Warning Label su TikTok. (fonte: TikTok)

Le informazioni utilizzate per effettuare i diversi tipi di confronto possono provenire da database governativi contenenti dati biometrici di individui precedentemente identificati, come le foto segnaletiche. Le fonti utilizzate non si limitano ai database delle forze dell'ordine, ma possono comprendere archivi con fotografie associate a patenti di guida e passaporti. Sono possibili, inoltre, le identificazioni tramite immagini acquisite attraverso sistemi di videosorveglianza e risorse disponibili pubblicamente online, come le immagini pubblicate sui social media. Un esempio significativo di utilizzo di queste ultime è rappresentato da *Clearview AI*, un sistema di riconoscimento facciale basato su un vasto database di immagini raccolte da fonti accessibili online.

Tuttavia, nelle indagini le immagini disponibili potrebbero non presentare necessariamente le stesse caratteristiche di quelle utilizzate in condizioni controllate. Esse possono essere acquisite in situazioni non ottimali oppure provenire da sistemi di videosorveglianza nei quali il soggetto viene ripreso all'interno di una sequenza di immagini in movimento. In tali circostanze, il riconoscimento facciale può risultare più complesso a causa di variazioni nella posa, nella scala, nell'illuminazione, nella sfocatura e nell'espressione del volto.

Alle criticità legate alle condizioni delle immagini si affiancano quelle derivanti dalle modalità con cui tali tecnologie vengono impiegate nell'ambito delle indagini. Il riconoscimento facciale può, infatti, essere utilizzato con finalità differenti, rendendo necessario distinguere tra un impiego *preventivo* e uno *repressivo*.

Ai fini *preventivi*, il riconoscimento facciale può essere impiegato come strumento di sorveglianza all'interno di luoghi pubblici, senza necessariamente essere collegato a un evento criminoso già verificatosi. Un esempio è rappresentato da una sperimentazione condotta tra il 2017 e il 2018 presso la stazione ferroviaria di Berlino Südkreuz, dove la tecnologia

di riconoscimento facciale venne utilizzata nell'ambito di un sistema di videosorveglianza finalizzato al riconoscimento dei volti delle persone presenti nell'area.

Al contrario, l'utilizzo per finalità *repressive* avviene in un momento successivo rispetto a quando viene commesso un reato ed è finalizzato all'identificazione delle persone coinvolte. In questo contesto, il riconoscimento può avvenire *ex post*, attraverso l'analisi di immagini precedentemente acquisite, oppure *in tempo reale*, al fine di localizzare e identificare persone ricercate. In quest'ultimo caso, i dati biometrici estratti da immagini acquisite in diretta possono essere confrontati immediatamente con quelli contenuti all'interno di un database.

L'impiego dei dati biometrici, le diverse fonti da cui possono essere ricavate le immagini e la possibilità di effettuare il riconoscimento in tempo reale sollevano importanti interrogativi in materia di tutela della privacy.

Nella prossima sottosezione verranno analizzate le principali problematiche che hanno alimentato il dibattito sull'utilizzo di tali tecnologie nell'ambito delle indagini.

6.2.1 Problematiche nella tutela della privacy

Il processo di riconoscimento facciale può essere suddiviso in tre momenti principali: raccolta dei dati, analisi e, infine, utilizzo delle informazioni ottenute.

Durante la prima fase vengono raccolti dati sotto forma di immagini facciali. Successivamente, durante la fase di analisi, le caratteristiche del volto vengono estratte e trasformate in modelli biometrici che vengono confrontati con quelli presenti all'interno di un database. Infine, le informazioni generate attraverso il riconoscimento possono essere utilizzate per l'attività investigativa, ad esempio attraverso l'avvio di altre misure investigative o la condivisione dei dati con altre autorità.

Dal punto di vista procedurale, l'impiego delle tecnologie di riconoscimento facciale da parte delle autorità comporta il trattamento di dati personali nelle diverse fasi del processo. Di conseguenza, ciascuna di queste fasi può presentare implicazioni relative alla protezione dei dati e alla tutela della privacy.

Un caso significativo relativo alle problematiche sulla tutela della privacy riguarda *Clearview AI*. Il Garante per la protezione dei dati personali ha imposto alla società una sanzione di 20 milioni di euro in seguito all'accertamento di diverse violazioni della normativa sulla protezione dei dati personali. In particolare, tra le problematiche riscontrate emergeva il monitoraggio biometrico anche di persone presenti sul territorio italiano.

La società ha dichiarato di possedere un database composto da oltre 10 miliardi di immagini di volti, raccolte attraverso tecniche di *web scraping*[3] da fonti pubblicamente accessibili, tra cui, ad esempio, i social media. A partire da tali immagini, il sistema è in grado di estrarre dati biometrici e creare profili, ai quali possono essere associate ulteriori informazioni, come la geolocalizzazione della fotografia e la pagina web dalla quale questa è stata acquisita.

La criticità, quindi, non riguardava soltanto il successivo utilizzo delle immagini per il riconoscimento facciale, ma anche le modalità con cui i dati personali e biometrici venivano raccolti e trattati.

Le problematiche relative alla tutela della privacy possono emergere anche durante l'utilizzo concreto dei sistemi di riconoscimento facciale da parte delle forze dell'ordine. Un esempio è rappresentato dal caso di *Ed Bridges*, riguardante l'impiego di sistemi automatizzati di riconoscimento facciale da parte della *South Wales Police*. Quest'ultima aveva utilizzato sistemi di riconoscimento facciale automatizzato in diverse occasioni. In particolare, Ed Bridges contestò due impieghi della tecnologia avvenuti il 21 dicembre 2017 e il 27 marzo 2018. La Corte sottolineò la mancanza di criteri sufficientemente chiari per stabilire quali

[3] *il web scraping* è una particolare tecnica di web mining, cioè di raccolta e organizzazione di informazioni presenti sul web (definizione presa da Agendadigitale).

persone potessero essere inserite nelle *watchlist*[4] e in quali luoghi potesse essere utilizzato il sistema, lasciando un'eccessiva discrezionalità alle forze di polizia. La Corte d'Appello ritenne, pertanto, che, nelle circostanze esaminate, l'utilizzo del sistema non fosse conforme alla legge ai fini dell'Articolo 8 della Convenzione Europea dei diritti dell'uomo, relativo al diritto al rispetto della vita privata.

Le problematiche emerse nei casi riportati mostrano come l'impiego del riconoscimento facciale nelle attività investigative debba essere valutato anche in relazione alla tutela dei diritti fondamentali. A livello europeo, tale tutela trova espressione nella *Carta dei diritti fondamentali dell'Unione Europea*.

Di particolare importanza sono l'Articolo 7 e l'Articolo 8. Il primo riconosce il diritto di ogni persona al rispetto della propria vita privata e familiare, mentre il secondo riconosce il diritto alla protezione dei dati di carattere personale. Quest'ultimo stabilisce, in particolare, che tali dati debbano essere trattati secondo il *principio di lealtà*, per finalità stabilite e sulla base del consenso della persona interessata o di un fondamento legittimo previsto dalla legge. Inoltre, ogni persona ha il diritto di accedere ai dati raccolti che la riguardano e di ottenerne la rettifica. Il rispetto di tali disposizioni è sottoposto al controllo di un'autorità indipendente.

L'utilizzo del riconoscimento facciale non pone, quindi, problematiche esclusivamente in relazione all'efficacia della tecnologia, ma anche rispetto alle modalità con cui i dati biometrici vengono raccolti, analizzati e successivamente utilizzati.

6.2.2 Conclusioni

Sebbene la Facial Analysis possa rappresentare un aiuto importante per velocizzare le indagini, le problematiche legate alla protezione della privacy del singolo cittadino risultano troppo rilevanti per essere trascurate. Inoltre, come dimostrato nei casi analizzati, l'affidabilità di questi sistemi può essere influenzata dalla qualità e dalle caratteristiche delle immagini che vengono adoperate nell'analisi. È, quindi, necessario riuscire a trovare un equilibrio tra i vantaggi offerti da queste tecnologie, che potrebbero velocizzare le indagini e, con il supporto umano, rendere il procedimento più accurato, e la tutela dei dati personali e dei diritti dell'individuo.

6.3 Discussione sul Case Study 3: sentiment analysis nei testi letterari

I risultati riportati nel quinto capitolo mostrano come Amazon Web Services presenti alcuni limiti nell'interpretazione di componimenti caratterizzati da allegorie complesse e significati impliciti.

La Sentiment Analysis trova applicazione in numerosi contesti differenti da quello letterario. Tra questi rientrano, ad esempio, l'analisi dei feedback e delle recensioni dei clienti, il monitoraggio dei social media, come X, con particolare interesse verso i commenti pubblicati online, e l'automatizzazione delle analisi di mercato. Gli input testuali vengono raccolti e analizzati al fine di individuare e classificare il sentiment espresso dagli utenti.

Nella Figura 6.2 viene rappresentato un esempio del processo di Sentiment Analysis.

Nel 2018 la Sentiment Analysis venne utilizzata da Nike, azienda operante nel settore dell'abbigliamento e degli articoli sportivi, nel contesto dei social media. In particolare, venne impiegata per analizzare la reazione del pubblico alla scelta dell'azienda di supportare Colin Kaepernick, giocatore di football americano della NFL che si inginocchiò durante

[4]Una *watchlist* è un insieme di immagini di riferimento note rispetto alle quali viene effettuata la ricerca di un'immagine di prova (definizione presa da South Wales Police).

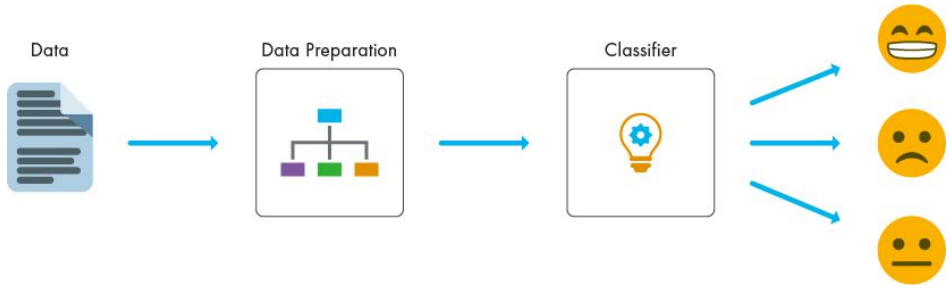


Figura 6.2: Esempio del processo di Sentiment Analysis. (fonte: MathWorks)

l'inno statunitense. Questo gesto rappresentava una forma di protesta contro il razzismo e le discriminazioni nei confronti degli afroamericani negli Stati Uniti. Sebbene esso fosse stato criticato da molte persone, altre mostrarono reazioni positive. La scelta di Nike risultava rischiosa e l'azienda monitorò, quindi, la reazione dell'opinione pubblica.

Per svolgere questa analisi venne utilizzato *Sentieo* [5], tramite il quale Nike riuscì a monitorare l'andamento dell'opinione pubblica in quel periodo. Tra gli altri strumenti utilizzati per l'analisi e il monitoraggio dei social media rientrano Talkwalker e Critical Mention.

Inizialmente l'opinione pubblica non risultò essere positiva e molti cercarono di boicottare Nike; con il passare del tempo, il sentiment si spostò verso reazioni positive.

Tuttavia, l'applicazione della Sentiment Analysis a contesti caratterizzati da una maggiore complessità linguistica può evidenziare alcune criticità, come emerso anche nella sperimentazione condotta nel quinto capitolo.

6.3.1 I limiti della Sentiment Analysis

La Sentiment Analysis non risulta essere infallibile. Infatti, sebbene possa essere utile per individuare il sentiment generale di un testo, non sempre riesce a riconoscerne le diverse sfumature. Queste difficoltà sono amplificate dalla presenza di ironia, sarcasmo, espressioni idiomatiche e negazioni, nonché dalla dipendenza del significato dal contesto.

Il contesto risulta essere fondamentale per comprendere se una frase esprima un sentiment positivo o negativo. Certi elementi, scorporati dal loro contesto, infatti, potrebbero risultare difficili da classificare; un esempio evidente è emerso dall'analisi di *Annabel Lee*.

I modelli di Sentiment Analysis riscontrano difficoltà nel riconoscere sarcasmo e ironia. Nel linguaggio parlato, il sarcasmo può essere espresso anche attraverso l'intonazione della frase, elemento assente nei testi scritti. Nello scritto, il sarcasmo può emergere attraverso la scelta di parole e il contesto in cui queste vengono utilizzate, rendendo complessa per il modello l'individuazione di tale sentimento.

Anche il linguaggio idiomatico può confondere i modelli di NLP, i quali possono interpretare le espressioni secondo il loro significato letterale, non cogliendone le diverse sfumature.

Una difficoltà simile può essere riscontrata in presenza di linguaggio figurato e metaforico. Questa conclusione è risultata evidente anche nell'interpretazione dell'espressione *"At the edge of the void"*, presente in *The Seagull and the Cat Who Taught Her to Fly*.

Un'ulteriore problematica è rappresentata dalla multipolarità, ovvero dalla presenza, all'interno dello stesso testo, di sentiment differenti. Un esempio può essere rappresentato dalle recensioni di prodotti, all'interno delle quali si descrivono lati positivi e negativi del prodotto acquistato.

[5] *Sentieo* è una suite di software per la ricerca sugli investimenti che utilizza l'Intelligenza Artificiale per scansionare documenti finanziari, analizzare set di dati alternativi e creare visualizzazioni (definizione presa da TechCrunch).

Per cercare di risolvere queste problematiche sono stati sviluppati modelli di Sentiment Analysis più sofisticati che integrano informazioni contestuali provenienti da testo, immagini e contenuti audio. Questi modelli sono stati in grado di migliorare significativamente l'accuratezza nella classificazione del sentiment. Ulteriori sviluppi potrebbero, quindi, ampliare le possibilità di applicazione della Sentiment Analysis in settori quali l'economia, la sanità e la Business Intelligence.

La Sentiment Analysis può, quindi, riconoscere e classificare il sentiment presente nei testi, ma per quanto riguarda le opere letterarie questa classificazione può non coincidere con la comprensione del significato emotivo dell'opera.

6.3.2 Riconoscere, comprendere o provare emozioni?

Fra i diversi interrogativi emersi con lo sviluppo dell'Intelligenza Artificiale, particolare attenzione è stata rivolta alla capacità dei sistemi artificiali di riconoscere e interpretare le emozioni umane. A partire da questa capacità è possibile porsi un ulteriore interrogativo: l'Intelligenza Artificiale potrà mai essere in grado di provare emozioni?

Sebbene un sistema di Intelligenza Artificiale possa essere progettato per individuare determinati stati emotivi e generare risposte che appaiono empatiche o emotivamente appropriate, come nel caso del chatbot ELIZA, ciò non significa che sia effettivamente in grado di provare tali emozioni.

Provare un'emozione, infatti, non significa soltanto riprodurre una risposta o imitare un comportamento, ma implica un'esperienza soggettiva legata a un determinato stato emotivo.

In questo contesto nacque il concetto di *computazione affettiva* (o *affective computing*), introdotto da Rosalind W. Picard e approfondito nel suo libro *Affective Computing*. Tale concetto si basa sull'idea di sviluppare sistemi capaci di riconoscere, interpretare ed elaborare informazioni legate alle emozioni.

Si è iniziato, quindi, ad applicare la computazione affettiva all'Intelligenza Artificiale con l'obiettivo di sviluppare sistemi capaci non soltanto di riconoscere e classificare gli stati emotivi umani, ma anche di integrare componenti di natura emotiva nei propri processi di elaborazione e decisione. In questo contesto si inserisce anche il concetto di *emotional computing*, inteso come un processo nel quale i meccanismi di elaborazione e deduzione propri dei sistemi artificiali vengono integrati con componenti di natura emotiva. L'obiettivo dell'*emotional computing* è, infatti, quello di riprodurre, almeno in parte, alcuni meccanismi del pensiero emotivo umano. In questa prospettiva, la componente emotiva non si limita al riconoscimento degli stati d'animo, ma viene considerata come un elemento capace di influenzare il ragionamento e le decisioni della macchina.

L'*affective computing* coinvolge l'analisi del sentiment e il riconoscimento delle emozioni. Per rappresentare le emozioni e i sentimenti possono essere utilizzati due modelli: il *modello di emozioni discrete o categoriali* e il *modello di emozioni dimensionali*.

Il modello delle emozioni discrete considera le emozioni come categorie distinte, mentre il modello dimensionale le rappresenta attraverso dimensioni continue.

Il riconoscimento delle emozioni può essere effettuato attraverso diverse tipologie di dati: informazioni fisiche, come dati testuali, audio e visivi, oppure segnali fisiologici. Lo studio si può basare, ad esempio, su informazioni testuali, come nel caso della sperimentazione svolta nel capitolo precedente, o sul *data mining* [6] delle opinioni, come le recensioni sui prodotti.

Gli esseri umani esprimono le proprie emozioni tramite diversi segnali, come espressioni facciali, voce e linguaggio, che rappresentano i dati fisici. Tuttavia, questo tipo di dato non risulta sempre affidabile in quanto l'essere umano può nascondere le proprie emozio-

[6] Il *data mining* è l'uso del machine learning e dell'analisi statistica per scoprire modelli e altre informazioni preziose da grandi set di dati. (definizione presa da IBM).

ni. I segnali fisiologici possono fornire informazioni più affidabili sullo stato emotivo, ma presentano una maggiore difficoltà di acquisizione.

Per ottenere maggiori informazioni sullo stato emotivo è, inoltre, possibile combinare informazioni fisiche e segnali fisiologici, attraverso un approccio multimodale.

Riconoscere le emozioni non equivale, però, a comprenderle o, addirittura, a provarle. Negli ultimi anni si è sviluppato un ampio dibattito sulla possibilità che l'Intelligenza Artificiale possa arrivare a possedere una forma di coscienza e, di conseguenza, a provare emozioni.

Molto importante risulta essere la *Theory of Mind*, concetto introdotto da Premack e Woodruff nel 1978. La *Theory of Mind* si riferisce alla capacità dell'uomo di attribuire stati mentali, come credenze, emozioni e intenzioni, a sé stessi e agli altri. Gli stati mentali non sono direttamente osservabili, ma vengono attribuiti sulla base del comportamento.

Questa tendenza ad attribuire stati mentali è particolarmente rilevante nell'interazione tra esseri umani e sistemi artificiali. Nel corso degli anni sono stati osservati numerosi casi di utenti che hanno attribuito caratteristiche tipicamente umane all'Intelligenza Artificiale. Un primo esempio è rappresentato da ELIZA, già analizzato nel primo capitolo. Successivamente, con lo sviluppo di assistenti vocali, robot sociali e chatbot sempre più avanzati, l'interazione con questi sistemi è diventata progressivamente più simile a quella che avviene tra esseri umani.

Un caso particolarmente discusso si verificò nel 2022: un ingegnere di Google, Blake Lemoine, affermò che *Language Model for Dialogue Applications* (LaMDA), un sistema avanzato di Intelligenza Artificiale, potesse essere senziente e possedere sentimenti propri, basandosi sulle conversazioni avute con LaMDA e sulla sua capacità di esprimersi riguardo a sentimenti, paure e desideri.

Tale interpretazione non fu, tuttavia, condivisa da Google e suscitò numerose critiche. Molti ricercatori non concordarono sul fatto che i sistemi di Machine Learning potessero effettivamente provare emozioni, sostenendo, piuttosto, che fossero in grado di imitare alcuni aspetti delle emozioni umane e che ciò potesse dare l'impressione di trovarsi di fronte a un essere senziente in grado di sentire.

Quindi, sebbene LaMDA potesse produrre risposte linguistiche apparentemente guidate dalle emozioni, dando all'interlocutore l'impressione di trovarsi di fronte a un soggetto capace di sentire, tale impressione non costituisce una dimostrazione del fatto che il sistema possieda la capacità di vivere un'esperienza emotiva.

Il caso LaMDA evidenzia, quindi, la difficoltà nel distinguere tra la capacità di un sistema artificiale di simulare o riprodurre linguisticamente un'emozione e la possibilità che esso la provi effettivamente.

6.3.3 Conclusioni

Gli studi citati mostrano come la Sentiment Analysis possa rappresentare uno strumento utile per riconoscere e classificare il sentiment espresso in un testo, ma presenti comunque molte difficoltà quando viene applicata a contenuti caratterizzati da significati impliciti, allegorie e una maggiore complessità interpretativa. Tuttavia, il riconoscimento di uno stato emotivo non coincide necessariamente con la sua comprensione o con la capacità di provare emozioni o, addirittura, avere una coscienza. Questo evidenzia come sia importante distinguere tra ciò che un sistema di Intelligenza Artificiale è in grado di riconoscere e classificare e ciò che può effettivamente comprendere o provare.

Conclusioni

Arrivati al termine del presente lavoro, si può affermare che gli obiettivi proposti nell'Introduzione sono stati complessivamente raggiunti. Inizialmente, per fornire un inquadramento del tema trattato, è stata proposta una descrizione panoramica sull'Intelligenza Artificiale, presentando la definizione, l'evoluzione storica e i principali modelli di apprendimento. Successivamente, è stata presentata la piattaforma cloud Amazon Web Services (AWS), introducendo i principi fondamentali del cloud computing e i principali modelli di servizio e distribuzione. Sono state analizzate l'infrastruttura globale e l'architettura della piattaforma, con particolare attenzione alle modalità di accesso e interazione con i servizi AWS. In questo contesto, particolare attenzione è stata dedicata al servizio Amazon Rekognition, attraverso il quale sono state utilizzate la Content Moderation e la Facial Analysis, e ad Amazon SageMaker, attraverso il quale sono state svolte sperimentazioni per l'analisi binaria del sentiment e l'Emotion Classification multi-classe.

L'analisi tramite Content Moderation ha permesso di osservare il funzionamento dei sistemi di controllo dei contenuti, per poi approfondire il confronto con le linee guida dei Social Media. La Facial Analysis ha evidenziato alcuni limiti nei casi più complessi, mostrando invece risultati più accurati negli altri casi. Lo studio della Sentiment Analysis binaria e dell'Emotion Classification multi-classe ha consentito di osservare il comportamento dell'Intelligenza Artificiale nell'analisi del linguaggio naturale, con particolare attenzione alla comprensione delle emozioni espresse nei testi poetici.

I risultati delle sperimentazioni hanno dimostrato come, nonostante tali strumenti siano utili e abbiano notevoli potenzialità, essi presentino ancora ampi margini di miglioramento. Le prestazioni possono risultare più o meno precise in base alle caratteristiche dei dati utilizzati. Tutte le sperimentazioni hanno, infatti, evidenziato alcuni errori. Sia a livello sperimentale sia a livello pratico, nella Content Moderation sono emerse difficoltà nel riconoscimento delle immagini più ambigue.

Nel contesto della Facial Analysis, il case study relativo alle gemelle non ha prodotto risultati particolarmente accurati, mentre nell'analisi dei testi sia la Sentiment Analysis binaria sia l'Emotion Classification multi-classe non sempre sono state in grado di interpretare correttamente il sentiment e le emozioni espresse nei testi. Questo evidenzia come, se affiancate a un operatore umano, tali tecnologie possano effettivamente rappresentare un supporto nello svolgimento di determinate attività, mentre, se utilizzate autonomamente, presentino ancora limiti da non sottovalutare.

L'Intelligenza Artificiale è ancora un campo in pieno sviluppo, caratterizzato da continui miglioramenti e da una rapida evoluzione delle tecnologie. Non dovrebbe essere considerata

in contrapposizione all'essere umano, ma come uno strumento che, se utilizzato correttamente, può supportare e velocizzare il lavoro umano. Attualmente sono numerose le posizioni riguardo a come e quanto rapidamente debba procedere lo sviluppo dell'Intelligenza Artificiale e alla capacità delle attuali tecnologie di gestire tale evoluzione in modo sicuro. All'interno di questo dibattito rientrano anche le discussioni relative alla possibilità che sistemi di Intelligenza Artificiale sempre più avanzati possano raggiungere capacità maggiormente autonome. La crescente autonomia di tali sistemi pone, quindi, il problema del loro controllo. All'aumentare delle capacità dell'Intelligenza Artificiale devono parallelamente svilupparsi strumenti in grado di garantirne un utilizzo sicuro e mantenere un'adeguata supervisione umana. In questo contesto va riconosciuta anche l'importanza dello sviluppo della cybersecurity, non solo nella protezione dell'AI ma anche nella tutela di dati, dei sistemi e delle infrastrutture da eventuali azioni indesiderate compiute autonomamente dall'AI.

Lo sviluppo di queste tecnologie potrebbe contribuire a ridurre alcuni dei problemi riscontrati nella presente tesi, migliorando la capacità dei sistemi di interpretare situazioni complesse e ambigue. Continueranno, tuttavia, a emergere interrogativi non soltanto sulle loro capacità tecniche, ma anche sul rapporto tra queste tecnologie e l'essere umano. Tra questi interrogativi rientra quello relativo alla possibilità che sistemi sempre più avanzati possano un giorno arrivare a comprendere le emozioni umane o, spingendosi ancora oltre, persino a provarle.

Il futuro dell'Intelligenza Artificiale, quindi, dipenderà non soltanto dalle capacità che queste tecnologie riusciranno a raggiungere, ma anche dal modo in cui l'essere umano sceglierà di svilupparle e utilizzarle.

Bibliografia

AMAZON WEB SERVICES (2021), «Using AWS in the Context of Common Privacy and Data Protection Considerations», Rap. tecn., Amazon Web Services, updated September 28, 2021.

BANDARU, A. (2020), «Amazon Web Services», Bournemouth University, department of Information Technology.

BARR, A. e FEIGENBAUM, E. A., curatori (1982), *The Handbook of Artificial Intelligence*, vol. 2, William Kaufmann, Los Altos, CA.

BOMMASANI, R., HUDSON, D. A., ADELI, E. e OTHERS (2021), «On the Opportunities and Risks of Foundation Models», *arXiv preprint arXiv:2108.07258*.

DEWAN, M. A. A., GRANGER, E., MARCIALIS, G. L., SABOURIN, R. e ROLI, F. (2016), «Adaptive Appearance Model Tracking for Still-to-Video Face Recognition», *Pattern Recognition*, vol. 49, p. 129–151.

GOODFELLOW, I., BENGIO, Y. e COURVILLE, A. (2016), *Deep Learning*, MIT Press.

GUINGRICH, R. E. e GRAZIANO, M. S. A. (2024), «Ascribing Consciousness to Artificial Intelligence: Human-AI Interaction and Its Carry-Over Effects on Human-Human Interaction», *Frontiers in Psychology*, vol. 15, p. 1322 781.

JAKHRANI, M. H. e SHAIKH, M. H. (2024), «A Critical Analysis of Imagery in My Heart Leaps Up: Illuminating Wordsworth's Nature Poetry in the Context of Ecocriticism», *LET: Linguistics, Literature and English Teaching Journal*, vol. 14 (1), p. 143–160.

MCCULLOCH, W. S. e PITTS, W. (1943), «A Logical Calculus of the Ideas Immanent in Nervous Activity», *The Bulletin of Mathematical Biophysics*, vol. 5 (4), p. 115–133.

MELL, P. e GRANCE, T. (2011), «The NIST Definition of Cloud Computing», Rap. Tecn. NIST Special Publication 800-145, National Institute of Standards and Technology.

MINSKY, M. e PAPERT, S. (1969), *Perceptrons: An Introduction to Computational Geometry*, MIT Press.

MITCHELL, T. M. (1997), *Machine Learning*, McGraw-Hill.

MOHAMMAD, S. M. (2022), «Ethics Sheet for Automatic Emotion Recognition and Sentiment Analysis», *Computational Linguistics*, vol. 48 (2), p. 239–278.

NILSSON, N. J. (2009), *The Quest for Artificial Intelligence*, Cambridge University Press.

RUMELHART, D. E., HINTON, G. E. e WILLIAMS, R. J. (1986), «Learning Representations by Back-Propagating Errors», *Nature*, vol. 323 (6088), p. 533–536.

RUSSELL, S. e NORVIG, P. (2021), *Artificial Intelligence: A Modern Approach*, Pearson, 4 ed.

SAMUEL, A. L. (1959), «Some Studies in Machine Learning Using the Game of Checkers», *IBM Journal of Research and Development*, vol. 3 (3), p. 210–229.

SEPÚLVEDA, L. (2016), *The Story of a Seagull and the Cat Who Taught Her to Fly*, Alma Books, Richmond, Surrey, translated by Margaret Sayers Peden.

SHI, F., YANG, J., LIFELO, Z., DING, J. e NING, H. (2026), «AGI-Powered Affective Computing in Cyber-Physical-Social-Thinking Space: Framework, Techniques, Applications and Future Directions», *Artificial Intelligence Review*, vol. 59, p. 215.

SHUKLA, P., KUMAR, R., DWIVEDI, V. K. e SINGH, A. K. (2026), «Aspect Based Sentiment Analysis: A Systematic Review, Taxonomy, Applications, and Future Research Directions», *Computer Science Review*, vol. 61, p. 100 924.

SIMMLER, M. e CANOVA, G. (2025), «Facial Recognition Technology in Law Enforcement: Regulating Data Analysis of Another Kind», *Computer Law & Security Review*, vol. 56.

TURING, A. M. (1936), «On Computable Numbers, with an Application to the Entscheidungsproblem», *Proceedings of the London Mathematical Society*, vol. 42 (2), p. 230–265.

TURING, A. M. (1950), «Computing Machinery and Intelligence», *Mind*, vol. 59 (236), p. 433–460.

WANG, Y., SONG, W., TAO, W., LIOTTA, A., YANG, D., LI, X., GAO, S., YANG, J., GE, W., ZHANG, W. e ZHANG, W. (2022), «A Systematic Review on Affective Computing: Emotion Models, Databases, and Recent Advances», *Information Fusion*, vol. 83–84, p. 19–52.

WEIZENBAUM, J. (1966), «ELIZA: A Computer Program for the Study of Natural Language Communication Between Man and Machine», *Communications of the ACM*, vol. 9 (1), p. 36–45.

ZHOU, Z. (2021), «Emotional Thinking as the Foundation of Consciousness in Artificial Intelligence», *Cultures of Science*, vol. 4 (3), p. 112–123.

Sitografia

- Agenda Digitale – https://www.agendadigitale.eu
- AI4Business – https://www.ai4business.it
- Amazon Web Services (AWS) – https://aws.amazon.com
- Andrea Minini – https://www.andreaminini.com
- Artea – https://www.artea.com
- AWS Documentation – https://docs.aws.amazon.com
- Berkeley EECS – https://people.eecs.berkeley.edu
- BioDiritto – https://www.biodiritto.org
- Biomedical CuE – https://www.biomedicalcue.it
- Britannica – https://www.britannica.com
- Built In – https://builtin.com
- Changes Unipol – https://changes.unipol.it
- Cittadinanza Digitale – https://www.cittadinanzadigitale.eu
- Cloudflare – https://www.cloudflare.com
- Conseguenze – https://www.conseguenze.org
- Contentsquare – https://contentsquare.com
- Copyleaks – https://copyleaks.com
- Corriere della Sera – https://www.corriere.it
- Culture Digitali – https://culturedigitali.eu
- Deep Learning Italia – https://deeplearningitalia.com
- Diario di un Analista – https://www.diariodiunanalista.it
- Emily Dickinson – https://www.emilydickinson.it

- EOLO – `https://www.eolo.it`
- EUR-Lex – `https://eur-lex.europa.eu`
- FlowHunt – `https://www.flowhunt.io`
- Garante per la protezione dei dati personali – `https://www.garanteprivacy.it`
- Google Cloud – `https://cloud.google.com`
- HPE – `https://www.hpe.com`
- HTML.it – `https://www.html.it`
- Hugging Face – `https://huggingface.co`
- HumAI – `https://www.humai.it`
- IBM Documentation – `https://www.ibm.com/docs`
- IBM Think – `https://www.ibm.com/think`
- INGV Ambiente – `https://ingvambiente.com`
- Internet Encyclopedia of Philosophy – `https://iep.utm.edu`
- Ionos Digital Guide – `https://www.ionos.it/digitalguide`
- Istituto Calvino – `https://www.istitutocalvino.edu.it`
- Jotform AI – `https://www.jotform.com/it/ai`
- Kinsta – `https://kinsta.com`
- Lenovo Knowledge Base – `https://www.lenovo.com`
- LUM – `https://www.lum.it`
- Mailchimp Resources – `https://mailchimp.com`
- MathWorks – `https://it.mathworks.com`
- Matteo Gorini – `https://matteogorini.it`
- MediaWorld – `https://www.mediaworld.it`
- Medium – `https://medium.com`
- Meta Transparency Center – `https://transparency.meta.com`
- National Geographic Italia – `http://nationalgeographic.it`
- NeuronUP – `https://neuronup.com`
- Neuroscienze.net – `https://www.neuroscienze.net`
- Noovle – `https://www.noovle.com`
- Okpedia – `https://www.okpedia.it`
- OpenLearn – `https://www.open.edu/openlearn`

- Oscar Prata – `https://www.oscarprata.com`
- Oubliette Magazine – `https://oubliettemagazine.com`
- Oversight Board – `https://www.oversightboard.com`
- Pictet – `https://am.pictet.com`
- phoenixNAP – `https://phoenixnap.it`
- Poetry Foundation – `https://www.poetryfoundation.org`
- PoetryVerse – `https://www.poetryverse.com`
- Poets.org – `https://poets.org`
- Prezi – `https://prezi.com`
- Progetto HMR – `https://www.progettohmr.it`
- Quix Consulting – `https://www.quixconsulting.com`
- Red Hat – `https://www.redhat.com`
- ResearchGate – `https://www.researchgate.net`
- Rivista AI – `https://www.rivista.ai`
- Rosa del Deserto – `https://rosadeldeserto.weebly.com`
- SAP – `https://www.sap.com`
- Skilla – `https://www.skilla.com`
- Sky Sport – `https://sport.sky.it`
- Società Servizi Italia – `https://www.sservizi.it`
- Storia delle Idee – `https://www.storiadelleidee.it`
- Studenti.it – `https://www.studenti.it`
- Studocu – `https://www.studocu.com`
- TechTalking – `https://techtalking.it`
- TechTarget – `https://www.techtarget.com`
- TikTok – `https://www.tiktok.com`
- Treccani – `https://www.treccani.it`
- Veridas – `https://veridas.com`
- Vincos – `https://vincos.it`
- Wired Italia – `https://www.wired.it`
- YouSciences – `https://www.yousciences.it`
- ZeroUno – `https://www.zerounoweb.it`

Ringraziamenti

Oggi si chiude un percorso estremamente importante, che mi lascerà tantissimi ricordi. Un percorso fatto di alti e bassi, di esami superati e non superati. Sebbene alla me che aveva appena iniziato tutto questo sembrasse impossibile, e quella corona tanto attesa sembrasse non arrivare mai, oggi questo traguardo non avrebbe lo stesso significato se non fosse per le persone che, in questi anni, mi sono state accanto.

Desidero ringraziare il Professor Ursino per la grande disponibilità e gentilezza dimostrate durante tutto il periodo di tirocinio e di tesi. Grazie per aver seguito con attenzione e interesse ogni mio progresso. La ringrazio, infine, per i consigli e per il tempo che mi ha dedicato, accompagnandomi alla conclusione di questo percorso.

Vorrei ringraziare i miei genitori per essermi stati accanto in questi anni e per non avermi mai fatto mancare il loro sostegno. Grazie per aver condiviso con me ogni momento e per aver festeggiato ogni traguardo come se fosse anche il vostro. Grazie per aver sempre creduto in me.

A mia sorella Sara, che, seppur lontana, mi ha sempre fatto sentire la sua presenza e ha fatto il tifo per me a ogni esame. Grazie per avermi distratta dallo studio ogni volta che tornavi a casa, con la tua costante voglia di fare qualcosa insieme e, soprattutto, di andare a fare shopping. Anche a distanza hai vissuto con me questo percorso che, senza di te, non sarebbe stato altrettanto divertente.

Ringrazio Carlo per avermi fatto conoscere più da vicino il mondo dell'informatica e per avermi aiutata a capire quale strada volessi intraprendere una volta concluso questo percorso. A te devo tanti consigli, il sostegno e tutte le serate a giocare a Solo insieme a Sara.

A Claudia, la mia compagna di progetti, senza la quale questi ultimi due anni non sarebbero stati gli stessi. Ora non ci rimane che aprire la nostra azienda di ricerca di bug sconosciuti.

Ringrazio Rachele e Valentina, il duo più comico che conosca: insieme sapete sempre come farmi ridere. Senza di voi, le uscite tra un esame e l'altro non sarebbero state così divertenti.

A Luca e Filippo, la coppia di amici che mi ha fatto scoprire che, effettivamente, due opposti possono funzionare bene insieme. Con voi ho iniziato questo percorso e posso dire che non avrei potuto desiderare compagni di viaggio migliori. Grazie per le giornate in SUP, per le cene in spiaggia e, soprattutto, per aver sopportato la mia ansia da esame.

Ringrazio Giulia e Alessio, la coppia più strana e bella che abbia mai visto, per essere stati sempre presenti e per aver creduto in me a ogni esame. Sono felice di aver condiviso con voi tanti momenti in questi anni.

A Federico, Filippo, Alessio e Mattia, per esserci sempre stati anche nella distanza. A voi devo tanti pomeriggi e serate trascorsi insieme, tra i racconti di viaggi di Federico e Filippo, quelli delle gare di Mattia, che giuro verrò a vedere più spesso, e le lezioni di storia di Alessio.

A Iacopo ed Elena, perché, nonostante le nostre uscite sporadiche, ho sempre sentito il vostro sostegno. Con voi le uscite non sono mai scontate e so che, ogni volta che ci vediamo, proveremo sempre qualcosa di nuovo. Che la nostra amicizia, iniziata alle superiori, possa durare per sempre.

Un ringraziamento speciale va a Baloo, il mio gattino, per essere stato presente a ogni sessione di studio e per avermi ascoltata ripetere ogni esame. Soprattutto per l'accurata meticolosità con cui sceglievi gli appunti sui quali addormentarti, che puntualmente erano proprio quelli che dovevo studiare.

E a Frida, la mia cagnolina, per tutte le passeggiate fatte insieme, per le corse in spiaggia e per aver trascorso quest'ultimo anno come mia piccola assistente nello studio. In ogni stanza in cui studierò, so che ci sarai anche tu.

A Minou, il mio angioletto, perchè anche oltre l'arcobaleno possa festeggiare questo traguardo insieme me.

E, infine, un grazie a me, perché, se sono qui, un po' di merito è anche mio.