

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 realizzazione di casi di studio nel contesto
dell'Intelligenza Artificiale: Sentiment Analysis, Image Recognition
e Speech-To-Text**

**Design and implementation of case studies in the context of
Artificial Intelligence: Sentiment Analysis, Image Recognition, and
Speech-To-Text**

Relatore

Prof. Domenico Ursino

Candidato

Cristian Signoretti

ANNO ACCADEMICO 2025-2026

*Spesso l'uomo diventa ciò che crede di essere.
Se continuo a dirmi che non riesco a fare una certa cosa, è possibile che finisca per diventare davvero
incapace di farla. Al contrario, se ho la convinzione di poterla fare, sicuramente acquisterò la capacità
di farla, anche se all'inizio potrei non averla.*

Mahatma Gandhi

Sommario

L'Intelligenza Artificiale, negli ultimi tempi, sta stravolgendo il modo di lavorare delle persone, determinando una progressiva automazione delle attività all'interno del contesto lavorativo. La presente tesi esplora tale scenario ponendosi due principali obiettivi. In primo luogo, vuole fornire gli elementi teorici necessari per comprendere le origini, l'evoluzione e le logiche di funzionamento di questa tecnologia. In secondo luogo, mira a valutare in modo critico l'efficacia di tre soluzioni "pronte all'uso" offerte in modalità *SaaS* dalla piattaforma cloud pubblica Amazon Web Services (AWS), ossia Amazon Comprehend, Amazon Rekognition e Amazon Transcribe. Tali servizi sono stati scelti poiché ricoprono tre dimensioni differenti, ovvero, l'elaborazione del linguaggio naturale, la percezione visiva e la trascrizione del parlato. In questo contesto, l'attività si è concentrata sull'esecuzione e sulla progettazione di tre case study con l'obiettivo di mostrare i confini operativi, i punti di forza e i punti di rottura degli algoritmi alla luce di ciascun servizio.

Keyword: Intelligenza Artificiale (IA), Amazon Web Services (AWS), Software-as-a-Service (SaaS), Elaborazione del Linguaggio Naturale (NLP), Computer Vision, Sentiment Analysis, Speech-to-Text

Introduzione	1
1 Introduzione all'Intelligenza Artificiale	3
1.1 Definizione e concetti fondamentali dell'Intelligenza Artificiale	3
1.1.1 Che cos'è l'Intelligenza Artificiale	4
1.1.2 Modelli di Intelligenza Artificiale più diffusi	4
1.1.3 Obiettivi dell'Intelligenza Artificiale	5
1.2 Tipi di approccio all'Intelligenza Artificiale	6
1.2.1 L'approccio di Turing	6
1.2.2 Approccio umano	7
1.2.3 Approccio razionale	7
1.2.4 Differenze tra i due approcci	7
1.3 Fondamenti interdisciplinari dell'Intelligenza Artificiale	7
1.3.1 Matematica e statistica	8
1.3.2 Informatica e ingegneria del software	9
1.3.3 Neuroscienze e modelli cognitivi	9
1.3.4 Linguistica e elaborazione del linguaggio	11
1.4 Tecniche e approcci dell'Intelligenza Artificiale	11
1.4.1 Introduzione al <i>Machine Learning</i>	12
1.4.2 Tipologie di apprendimento	13
1.4.3 Introduzione al Deep Learning	14
1.5 Applicazioni e impatti dell'Intelligenza Artificiale	15
1.5.1 Applicazioni nel <i>Natural Language Processing (NLP)</i>	15
1.5.2 Applicazioni nella Computer Vision (Image Recognition)	16
1.5.3 Applicazioni nello Speech Processing (Speech-To-Text)	17
1.5.4 Impatti, rischi, implicazioni etiche e prospettive future	18
1.6 L'evoluzione dell'Intelligenza Artificiale nel <i>Cloud</i>	19
2 Amazon Web Services	20
2.1 Genesi ed evoluzione del modello <i>Cloud</i> di Amazon	20
2.2 Architettura e infrastruttura globale	21
2.2.1 Concetti di regioni e zone di disponibilità	22
2.2.2 Sicurezza e conformità dei dati nel <i>Cloud</i>	22
2.3 Modelli di servizio di AWS	24
2.4 L'ecosistema per l'Intelligenza Artificiale e il <i>Machine Learning</i>	25

2.4.1	Gestione della potenza computazionale tramite istanze scalabili e acceleratori hardware	25
2.4.2	La piattaforma <i>Amazon SageMaker</i> per lo sviluppo di modelli	26
2.4.3	Intelligenza Artificiale generativa ready-to-use: <i>Amazon Bedrock</i>	27
2.4.4	<i>SageMaker</i> vs <i>Bedrock</i>	28
2.5	Servizi cognitivi e strumenti di Intelligenza Artificiale pronti all'uso	29
2.5.1	Analisi del testo e dei sentimenti: <i>Amazon Comprehend</i>	29
2.5.2	Visione artificiale e analisi d'immagine: <i>Amazon Rekognition</i>	30
2.5.3	Trascrizione vocale: <i>Amazon Transcribe</i>	32
2.6	Vantaggi strategici e limiti nell'adozione di soluzioni AWS	33
2.6.1	Vantaggi Strategici ed Efficientamento Architetture	33
2.6.2	Limiti Tecnici, Vincoli e Fattori di Rischio	33
2.6.3	Sintesi dei compromessi	34
3	Un case study nel contesto della Sentiment Analysis	35
3.1	Sentiment Analysis	35
3.1.1	Definizione di Sentiment Analysis	36
3.1.2	Aspetti tecnici sulla Sentiment Analysis	36
3.1.3	Applicazioni attuali della Sentiment Analysis	37
3.2	Contesto e metodologia per attuare la Sentiment Analysis	37
3.2.1	Implementazione con <i>Amazon Comprehend</i>	38
3.3	Case study: analisi di recensioni di prodotti	40
3.3.1	Dataset e preparazione	41
3.3.2	Data Cleaning e Pre-processing	42
3.3.3	Il problema della Privacy: Integrazione della PII Detection	42
3.3.4	Stress Test di <i>Amazon Comprehend</i> sulle entità PII	43
3.3.5	Metriche di accuratezza del mascheramento PII	45
3.3.6	Risultati e metriche della Sentiment Analysis	46
3.3.7	Analisi qualitativa	48
3.4	Punti di forza e limitazioni riscontrati	48
3.4.1	Punti di forza riscontrati	48
3.4.2	Limitazioni riscontrate	49
4	Un case study nel contesto dell'Image Recognition	50
4.1	Image Recognition	50
4.1.1	Definizione di Image Recognition	51
4.1.2	Aspetti tecnici sull'Image Recognition	51
4.1.3	Applicazioni attuali dell'Image Recognition	54
4.2	Contesto e metodologia per attuare l'Image Recognition	56
4.2.1	Implementazione con <i>Amazon Rekognition</i>	57
4.3	Case study: Analisi sperimentale dei servizi di <i>Amazon Rekognition</i>	59
4.3.1	Rilevamento e classificazione degli oggetti (Labels Detection)	60
4.3.2	Analisi facciale e riconoscimento delle emozioni (Face Detection and Analysis)	65
4.3.3	Riconoscimento di personaggi famosi (Celebrity Recognition)	69
4.3.4	Videosorveglianza smart: Analisi temporale e tracciamento di intrusi	75
4.3.5	Analisi qualitativa	78
4.4	Punti di forza e limitazioni riscontrate	79
4.4.1	Punti di forza riscontrati	79
4.4.2	Limitazioni riscontrate	80

5	Un case study nel contesto dello Speech-To-Text	81
5.1	Speech-To-Text	81
5.1.1	Definizione di Speech-To-Text	82
5.1.2	Aspetti tecnici dello Speech-To-Text	82
5.1.3	Applicazioni attuali dello Speech-To-Text	83
5.2	Contesto e metodologia per attuare lo Speech-To-Text	84
5.2.1	Implementazione con Amazon Transcribe	84
5.3	Case study: Analisi sperimentale dei servizi di Amazon Transcribe	85
5.3.1	Stress test dialettico con Amazon Transcribe	86
5.3.2	Analisi delle chiamate Amazon Transcribe	92
5.3.3	Analisi qualitativa	97
5.4	Punti di forza e limitazioni riscontrate	98
5.4.1	Punti di forza riscontrati	98
5.4.2	Limitazioni riscontrate	99
6	Discussione	100
6.1	Confronto dei servizi analizzati tramite i risultati dei case study	100
6.2	Implicazioni concrete nel mondo aziendale	101
6.3	Strategie per mitigare i rischi e massimizzare i benefici	103
6.4	Prospettive future della cooperazione tra uomo e Intelligenza Artificiale . . .	105
	Conclusioni	108
	Bibliografia	110
	Sitografia	112
	Ringraziamenti	114

Elenco delle figure

1.1	Linea temporale dell'evoluzione dell'Intelligenza Artificiale dal 1943 ai giorni nostri (fonte: easy-peasy.ai)	5
1.2	Definizione di gradiente di una funzione $f(x)$ in N variabili (fonte: ai4business.it)	8
1.3	Calcolo del valore atteso di una variabile nel discreto, data la sua funzione di probabilità $P(X = x)$ (fonte: ai4business.it)	8
1.4	Calcolo del valore atteso di una variabile nel continuo, data la sua funzione di probabilità $p(x)$ (fonte: ai4business.it)	9
1.5	Teorema di Bayes nella sua forma più semplice (fonte: ai4business.it)	9
1.6	Neurone naturale (fonte: skillbook.it)	10
1.7	Confronto tra neurone e schema neurale per gli algoritmi (fonte: skillbook.it)	10
1.8	Rappresentazione vettoriale di come sono distribuite le parole all'interno di un documento con evidenziate le relazioni tra i termini (fonte: wikipedia.it)	11
1.9	Rappresentazione delle relazioni tra Intelligenza Artificiale, <i>Machine Learning</i> e Deep Learning (fonte: oracle.it)	12
1.10	Rappresentazione del ciclo di vita di un sistema di <i>Machine Learning</i> (fonte: medium.com)	13
1.11	Rappresentazione dell'interazione con "Amazon Echo" tramite la tecnologia <i>Speech-To-Text</i> (fonte: enerzai.com)	18
2.1	Schema temporale dell'evoluzione di AWS fino ad oggi (fonte: touchlane.com)	21
2.2	Panoramica dell'architettura delle regioni e delle zone di disponibilità di AWS (fonte: docs.aws.amazon.com)	22
2.3	Illustrazione della responsabilità "DEL" Cloud (responsabilità di AWS) e "NEL" Cloud (responsabilità del cliente) (fonte: aws.amazon.com)	23
2.4	Illustrazione dei tre modelli di servizio di AWS con annessa breve descrizione dell'utilizzo (fonte: kinsta.com)	24
2.5	Ciclo di vita del <i>Machine Learning</i> con Amazon SageMaker (fonte: aws.amazon.com)	27
2.6	Illustrazione del flusso generale di Amazon Comprehend (fonte: aws.amazon.com)	30
2.7	Uno dei possibili flussi che può essere ottenuto tramite l'utilizzo di Amazon Rekognition (fonte: aws.plainenglish.io)	32
2.8	Illustrazione dei due possibili flussi di <i>Amazon Transcribe</i> : trascrizione in batch e trascrizione in streaming (fonte: aws.amazon.com)	33
3.1	Rappresentazione delle applicazioni attuali della Sentiment Analysis (fonte: www.delve.ai)	38

3.2	Rappresentazione della configurazione di Comprehend	39
3.3	Rappresentazione dei risultati di Comprehend via interfaccia utente selezionando "PII"	40
3.4	Rappresentazione dei risultati di Comprehend via API selezionando "PII" . .	41
3.5	Rappresentazione dei risultati di Comprehend via interfaccia utente selezionando "Sentiment"	41
3.6	Rappresentazione dei risultati di Comprehend via interfaccia API selezionando "Sentiment"	42
3.7	Illustrazione dei passaggi eseguiti per effettuare la Sentiment Analysis	43
4.1	Illustrazione del principio di funzionamento dell' <i>Image Recognition</i> (fonte: fastweb.it)	51
4.2	Esempio di <i>Object Detection</i> applicato al monitoraggio del traffico urbano in tempo reale (fonte: ibm.com)	53
4.3	Esempio di elaborazione in spazio vettoriale tramite la proiezione Bird's-Eye View nel sistema Tesla FSD. L'algoritmo di Computer Vision unifica i flussi video bidimensionali per generare una mappa tridimensionale predittiva dell'ambiente circostante (fonte: autoevolution.com)	54
4.4	Implementazione di un sistema automatizzato di <i>Computer Vision</i> per il monitoraggio e la gestione dei flussi pedonali in contesti urbani. L'architettura integra algoritmi di <i>object detection</i> e <i>multi-target tracking</i> per stimare in tempo reale densità, velocità e traiettorie dei soggetti. (fonte: mdpi.com)	55
4.5	Esempio di localizzazione di un tumore polmonare mediante CMYOLOv5 . .	56
4.6	Illustrazione della configurazione di Rekognition	57
4.7	Illustrazione dei passaggi da seguire per utilizzare il microservizio di etichette personalizzate di Rekognition (fonte: aws.amazon.com)	58
4.8	Illustrazione dei passaggi da seguire per utilizzare il microservizio di analisi di massa di Rekognition (fonte: aws.amazon.com)	58
4.9	Illustrazione del comportamento di Rekognition relativamente a degli scaffali di un supermercato (parte 1)	61
4.10	Illustrazione del comportamento di Amazon Rekognition relativamente a degli scaffali di un supermercato (parte 2)	62
4.11	Illustrazione del comportamento di Amazon Rekognition relativamente a degli scaffali di un supermercato (parte 3)	62
4.12	Illustrazione del comportamento di Amazon Rekognition relativamente a un aeroporto (caso diurno) (fonte: maltairport.com)	63
4.13	Illustrazione del comportamento di Amazon Rekognition relativamente a un aeroporto in condizioni di scarsa visibilità (caso notturno) (fonte: maltairport.com)	64
4.14	Illustrazione del comportamento di Rekognition relativamente a un aeroporto nel dettaglio (dettaglio ravvicinato)	65
4.15	Illustrazione del comportamento di Rekognition tramite la <i>Face analysis</i> relativamente a un individuo alla guida (posizione frontale) (fonte: DMD - Driver Monitoring Dataset)	66
4.16	Rappresentazione dei risultati dell'array delle emozioni rispetto al comportamento di un guidatore (inquadratura frontale), valutati tramite <i>Face analysis</i> di Amazon Rekognition	67
4.17	Illustrazione del comportamento di Amazon Rekognition tramite <i>Face analysis</i> + relativamente a un individuo alla guida (posizione laterale) (fonte: DMD - Driver Monitoring Dataset)	68

4.18	Rappresentazione dei risultati dell'array delle emozioni rispetto al comportamento di un guidatore (inquadratura laterale), valutati tramite <i>Face Analysis</i> di Amazon Rekognition	69
4.19	Illustrazione del comportamento di Amazon Rekognition tramite <i>Face analysis</i> relativamente a un individuo alla guida (sonnolenza) (fonte: DMD - Driver Monitoring Dataset)	69
4.20	Rappresentazione dei risultati dell'array delle emozioni rispetto al comportamento di un guidatore (faccia occlusa), valutati tramite <i>Face Analysis</i> di Amazon Rekognition	70
4.21	Illustrazione del comportamento di Rekognition tramite la <i>Celebrity Recognition</i> relativamente alla popstar Ariana Grande	71
4.22	Illustrazione delle emozioni rilevate da <i>Celebrity Recognition</i> nell'immagine precedente	72
4.23	Illustrazione del comportamento di Amazon Rekognition tramite <i>Celebrity Recognition</i> relativamente agli influencer Chiara Ferragni e Khaby Lame	73
4.24	Illustrazione delle emozioni rilevate da <i>Celebrity Recognition</i> nell'immagine precedente	73
4.25	Illustrazione di un gruppo di attori internazionali e nazionali sottoposti all'analisi di <i>Celebrity Recognition</i>	74
4.26	Illustrazione del comportamento del microservizio <i>Celebrity Analysis</i> in merito a un gruppo di attori nazionali e a un gruppo di attori internazionali	75
4.27	Illustrazione del dettaglio della comparsa delle specifiche entità nel video	76
4.28	Illustrazione dell'interfaccia grafica del microservizio <i>Analisi dei video archiviati</i> di Amazon Rekognition	77
4.29	Illustrazione delle rilevazioni del microservizio <i>Analisi dei video archiviati</i> di Amazon Rekognition in merito al primo frame video fornito ad esso (fonte: youtube.com - canale: Bor Security)	77
4.30	Illustrazione delle rilevazioni del microservizio <i>Analisi dei video archiviati</i> di Amazon Rekognition in merito al secondo frame video fornito ad esso (fonte: youtube.com - canale: Bor Security)	78
4.31	Illustrazione delle rilevazioni del microservizio <i>Analisi dei video archiviati</i> di Amazon Rekognition in merito al terzo frame video fornito ad esso (fonte: youtube.com - canale: Bor Security)	78
5.1	Illustrazione del principio generale di funzionamento dello <i>Speech-To-Text</i> (fonte: medium.com)	82
5.2	Illustrazione del flusso atto a realizzare lo <i>Speech-To-Text</i> (fonte: medium.com)	83
5.3	Illustrazione dell'interfaccia di Amazon Transcribe	85
5.4	Illustrazione dell'interfaccia del primo microservizio di Amazon Transcribe	87
5.5	Illustrazione dell'elaborazione del giuramento di Carlo III senza filtri (figura superiore) e con filtri (figura inferiore)	89
5.6	Illustrazione dell'elaborazione di una scena del film " <i>Phas Gaye Re Obama</i> "	93
5.7	Illustrazione della configurazione della categoria "Angry" per il microservizio di Transcribe "Analisi delle chiamate"	94
5.8	Illustrazione della configurazione della categoria "Comprehend" per il microservizio di Transcribe "Analisi delle chiamate"	95
5.9	Illustrazione dello sviluppo temporale della conversazione e l'interfaccia di riepilogo generata dal microservizio <i>analisi delle chiamate</i> per il primo campione audio	96

5.10	Illustrazione dello sviluppo temporale della conversazione e l'interfaccia di riepilogo generata dal microservizio <i>analisi delle chiamate</i> per il secondo campione audio	97
6.1	Illustrazione delle mansioni lavorative che gli LLM potrebbero teoricamente eseguire (blu), confrontata con la stima di quante professioni siano effettivamente impattate (rosso), basata sui dati reali di utilizzo (fonte: anthropic.com)	105
6.2	Illustrazione dell'intelligenza a ragnatela (fonte: researchgate.net)	107

Elenco delle tabelle

1.1	Riepilogo dei principali sviluppatori di sistemi di Intelligenza Artificiale . . .	6
1.2	Riepilogo dei principali aspetti dei due approcci all'Intelligenza Artificiale . .	8
2.1	Riepilogo dei principali compromessi analizzati per AWS	34
3.1	Esempio di PII rilevati all'interno di una recensione reale	44
3.2	Stress test su indirizzi stradali metaforici	44
3.3	Stress test serie di codici numerici	44
3.4	Stress test recensioni con dati reclamo	45
3.5	Stress test nel codice fiscale italiano	45
3.6	Caso di studio: Analisi del feedback dei clienti su alcuni prodotti proposti in <i>Amazon.it</i> tramite Amazon Comprehend	46
6.1	Analisi comparativa dei vantaggi e degli svantaggi riscontrati nei tre servizi esaminati.	102

Da quando sono stati presentati al pubblico i primi modelli di Intelligenza Artificiale generativa nel 2022, essa si è evoluta in modo esponenziale, infatti, ogni modello risulta migliore del precedente per ordini di grandezza. Fondamentalmente, tale tecnologia è capace di generare contenuti basandosi sull'analisi massiva di grandi moli di dati passati. Oggi, ha trovato applicazione in vari ambiti, dai sistemi di sicurezza, ai moduli di accessibilità fino all'ottimizzazione della produttività, tutti settori che hanno permesso l'incremento dell'efficienza in ambito aziendale e non. Il futuro della disciplina si muove, infatti, verso traguardi ambiziosi, che vanno dal superamento dei limiti dei modelli attuali, con l'obiettivo di rendere le capacità dei sistemi sempre più prossime ad un'Intelligenza Artificiale Generale (AGI), fino all'integrazione dei computer quantistici. Sebbene i sistemi di calcolo quantistico rappresentino una realtà hardware già parzialmente esistente e in fase di sviluppo, una delle sfide dei prossimi decenni sarà proprio la loro integrazione con i modelli di *Machine Learning*, un'unione che porterà un'evoluzione ancora più forte rispetto a quella che abbiamo vissuto negli ultimi anni.

In questo contesto, il presente lavoro nasce dalla crescente consapevolezza dell'importanza che l'Intelligenza Artificiale ha assunto nella nostra quotidianità, e si sviluppa per due motivi principali. In primo luogo, c'è la volontà di comprendere come funzionino, sia dal punto di vista teorico che pratico, gli strumenti che stanno guidando questa rivoluzione, analizzando anche le tappe storiche che ne hanno permesso la nascita. In secondo luogo, vi è l'intenzione di rimanere al passo con i tempi, un fattore fondamentale per migliorare la produttività personale e mantenere un'elevata competitività all'interno del mercato aziendale attuale.

Partendo da questi obiettivi, nella presente tesi costruiamo un percorso che, a partire dalle panoramiche teoriche dell'Intelligenza Artificiale e dei relativi modelli da essa utilizzati, ci consenta di effettuare specifici case study tramite la piattaforma Amazon Web Services (AWS). Questo ecosistema è stato selezionato per la sua popolarità sul mercato, per la molteplicità dei servizi e microservizi da esso offerti e per la possibilità di verificare direttamente sul campo il funzionamento di soluzioni SaaS "pronte all'uso". L'utilizzo di AWS permette, infatti, di analizzare queste tecnologie accostandole alla teoria e agli aspetti tecnici che le compongono, senza la necessità di dover applicare complesse configurazioni che, in altre piattaforme, sarebbero state necessarie.

Il presente lavoro si sviluppa, quindi, mostrando, in primo luogo, la panoramica teorica della disciplina per poi, in secondo luogo, valutarne l'efficacia sul campo attraverso gli strumenti di AWS. Il nostro obiettivo è analizzarne le funzionalità offerte come applicazioni pratiche di concetti chiave quali il *Deep Learning*, il *Machine Learning*, le reti neurali profonde,

la *Sentiment Analysis*, l'*Image Recognition* e lo *Speech-To-Text*.

Su queste premesse, abbiamo sviluppato la parte centrale della tesi concentrandoci sulla progettazione di tre specifici *case study*, scelti per esplorare tre differenti settori. Il primo case study si concentra su Amazon Comprehend per l'analisi testuale e l'estrazione delle emozioni (*sentiment analysis*), accostato allo sfruttamento del microservizio di rilevamento delle informazioni sensibili (*PII Detection*). Per tale case study, ci si è serviti del servizio Amazon Transcribe. Il secondo case study sfrutta le capacità di Amazon Rekognition, attraverso i suoi microservizi, per il riconoscimento degli oggetti (*Label Detection*), per l'analisi facciale e il riconoscimento delle emozioni (*Face Detection and Analysis*), per la *Celebrity Recognition* e per il tracciamento dei video. Il terzo e ultimo case study si sposta, infine, sul settore dei flussi audio tramite Amazon Transcribe, esaminandone sia la trascrizione standard sia le logiche comportamentali del modulo di *Analisi delle chiamate*. Ciascun servizio è stato sottoposto a sessioni incrementalmente di stress test, introducendo elementi di disturbo testuale, visivo e acustico, come testi ambigui e ironici, inquadrature notturne, occlusioni e forti rumori di fondo, al fine di individuare i reali punti di rottura dei servizi cloud offerti da AWS.

I dati che abbiamo raccolto evidenziano che, sebbene le soluzioni *SaaS* offrano oggi un'immediata integrabilità e un immediato utilizzo, per quanto riguarda ambienti più articolati e complessi le vulnerabilità strutturali dei servizi impongono lo sviluppo di strategie software di mitigazione precedenti al loro utilizzo.

L'intero percorso conoscitivo che abbiamo organizzato è stato strutturato in sei capitoli secondo la sequenza di seguito riportata:

- *Capitolo 1*: poniamo le basi concettuali del lavoro, definendo l'Intelligenza Artificiale, ripercorrendone la sua nascita, la sua evoluzione, le metodologie di approccio, i fondamenti della disciplina, le tecniche per attuarla e, infine, alcuni contesti applicativi in cui essa si rende indispensabile.
- *Capitolo 2*: in questo capitolo ci concentriamo sull'infrastruttura globale di *cloud computing* offerta da Amazon Web Services (AWS), delineandone la nascita, descrivendone l'architettura e l'infrastruttura, mostrandone l'ecosistema di soluzioni nate per l'elaborazione dei dati e per il *Machine Learning*, e infine, soffermandoci su alcune soluzioni dedicate all'Intelligenza Artificiale, con relativi vantaggi e svantaggi.
- *Capitolo 3*: introduciamo le logiche dell'elaborazione del linguaggio naturale e della *sentiment analysis* mediante l'impiego di Amazon Comprehend. Tale servizio ci ha consentito di effettuare, oltre all'analisi del testo, il riconoscimento e la redazione delle informazioni sensibili, mettendo in luce l'affidabilità e i vincoli del servizio.
- *Capitolo 4*: analizziamo il settore dell'*Image Recognition* attraverso la suite Amazon Rekognition, documentando l'accuratezza dei suoi microservizi, quali, *Label Detection*, *Face Detection and Analysis*, *Celebrity Recognition* e analisi dei video, illustrando le principali funzionalità e mostrandone i rispettivi risultati di fronte a variabili progressivamente più sfidanti.
- *Capitolo 5*: dedichiamo il capitolo interamente alle tecnologie di *Speech-to-Text* e ai sistemi di *Call Analytics* di Amazon Transcribe, delineandone la panoramica teorica, per poi testare i limiti e i vantaggi del servizio.
- *Capitolo 6*: proponiamo una discussione critica dell'intero lavoro. Questo capitolo mette in relazione i risultati per delineare le migliori strategie software di mitigazione dei rischi e dei limiti, concludendo la tesi con un'analisi sull'impatto di tali architetture sul futuro del lavoro e sulla cooperazione tra l'uomo e le macchine.

Introduzione all'Intelligenza Artificiale

Il presente capitolo spiega cos'è l'Intelligenza Artificiale, analizzando le definizioni principali e gli obiettivi che essa si pone. Il capitolo vuole mettere in luce i diversi approcci, facendo una distinzione, e allo stesso tempo trovando le comunanze, tra il criterio del Test di Turing, gli approcci basati sulla razionalità e sul comportamento umano. Viene poi approfondita anche la natura interdisciplinare dell'Intelligenza Artificiale, sottolineando il contributo che hanno avuto varie discipline quali la statistica, le neuroscienze e la linguistica computazionale. Successivamente, il capitolo si concentra sulle tecniche di Machine Learning e Deep Learning, facendo una distinzione, con dovizia di particolari, dei modelli di apprendimento e delle architetture delle reti neurali artificiali. Il capitolo esplora, inoltre, delle applicazioni pratiche, con un particolare approfondimento relativo ai sistemi di Natural Language Processing, Computer Vision e Speech Processing, contestualizzando anche i loro benefici, i limiti tecnologici e i rischi etici discussi e messi in risalto da quando questa tecnologia ha preso piede a livello globale. Infine, viene trattata l'evoluzione dei sistemi di Intelligenza Artificiale all'interno delle più grandi infrastrutture Cloud.

1.1 Definizione e concetti fondamentali dell'Intelligenza Artificiale

Uno degli interessi principali della comunità scientifica è da sempre quello di comprendere i processi cognitivi umani. L'utilizzo di questi processi applicati alla realtà ha permesso alla specie di sviluppare se stessa e l'ambiente circostante negli anni. L'intelligenza diventa, così, uno strumento essenziale per interagire con la realtà e per evolvere. Gli obiettivi della comunità scientifica risiedono, perciò, nel creare sistemi in grado di replicare e migliorare funzioni umane, come l'apprendimento, la risoluzione di problemi e il decision making. Proprio per questi motivi è nato questo spiccato interesse per l'Intelligenza Artificiale, disciplina che si prepone di realizzare questi ambiziosi compiti.

L'Intelligenza Artificiale non deve essere vista come una semplice imitazione delle facoltà umane, ma piuttosto come lo sviluppo di modelli capaci di operare in modo autonomo all'interno di scenari caratterizzati da elevata complessità e incertezza. Attualmente, questo settore costituisce una delle frontiere più avanzate dell'innovazione tecnologica, con riflessi economici e sociali di portata storica. La disciplina abbraccia una vasta gamma di campi di studio, che includono l'elaborazione del linguaggio naturale, la visione artificiale e la robotica, influenzando ogni attività di tipo intellettuale. Identificare una definizione tecnica univoca per un sistema di Intelligenza Artificiale è, perciò, un compito articolato. Storicamente, il dibattito scientifico si è diviso tra chi privilegia il confronto con le prestazioni umane e chi, invece, si focalizza su una concezione astratta di efficienza, definita razionalità.

Tali differenze teoriche si sviluppano in due tesi principali: la contrapposizione tra processi interni di pensiero e comportamento esterno osservabile, e quella tra la fedeltà al modello umano e il mantenere un ideale di logica perfetta. Dall'intersezione di tali dimensioni emergono i due paradigmi fondamentali della disciplina, che verranno approfonditi nelle sezioni successive.

1.1.1 Che cos'è l'Intelligenza Artificiale

L'Intelligenza Artificiale è la capacità di una macchina di mostrare abilità umane, quali il ragionamento, l'apprendimento, la pianificazione e la creatività.

Essa permette ai sistemi sui quali è installata di capirne l'ambiente, mettersi in relazione con quello con cui viene interfacciata, risolvere problemi e agire verso un obiettivo specifico da essa stabilito o fornito da chi la controlla. Il sistema riceve i dati o li raccoglie, li processa e, infine, risponde.

I sistemi di Intelligenza Artificiale sono capaci di adattare il proprio comportamento agli input esterni, analizzando gli effetti delle loro azioni precedenti e lavorando senza necessità di un supervisore.

Alcuni tipi di Intelligenza Artificiale esistono da più di cinquanta anni, ma l'evoluzione dell'informatica, delle tecnologie e dello sviluppo di nuovi algoritmi ha condotto la disciplina a fare enormi salti in avanti negli ultimi anni. Applicazioni future di essa porteranno sicuramente grandi cambiamenti alla nostra vita di tutti i giorni, ma non dobbiamo dimenticare che l'Intelligenza Artificiale è già presente nelle nostre vite.

1.1.2 Modelli di Intelligenza Artificiale più diffusi

Nella Figura 1.1 viene illustrata l'evoluzione temporale dei sistemi di Intelligenza Artificiale dal 1940 a oggi.

Risulta evidente come la società contemporanea sia attraversata da una rivoluzione digitale senza precedenti, guidata dalla rapidissima evoluzione dell'Intelligenza Artificiale. Questa risorsa, integrata spesso in modo invisibile all'utente nella maggior parte delle applicazioni odierne, può essere suddivisa in:

- *Intelligenza Artificiale Generativa*: risulta attualmente la più diffusa nel dibattito pubblico (*ChatGPT*, *Midjourney*). Il suo scopo è creare nuovi contenuti: testi, immagini, musica, video o codice informatico. Prende i dati che ha imparato e li rielabora per generare qualcosa di nuovo che prima non esisteva.
- *Intelligenza Artificiale Discriminativa o Analitica*: questa è l'Intelligenza Artificiale che viene utilizzata prevalentemente senza accorgercene. Invece di creare, il suo compito è analizzare e decidere.

Le soluzioni di Intelligenza Artificiale oggi sono più di 1500; quindi, di seguito, verranno mostrati i principali modelli sia linguistici che di generazioni immagini.

I principali modelli linguistici sono:

- *serie GPT (OpenAI)*: ha rilasciato vari modelli ed è considerato tra i più utilizzati per le sue capacità multimodali (testo, audio, immagini);
- *Gemini (Google)*: integrato nell'ecosistema Google, è efficace nella ricerca di informazioni e nell'analisi di contenuti ampi;

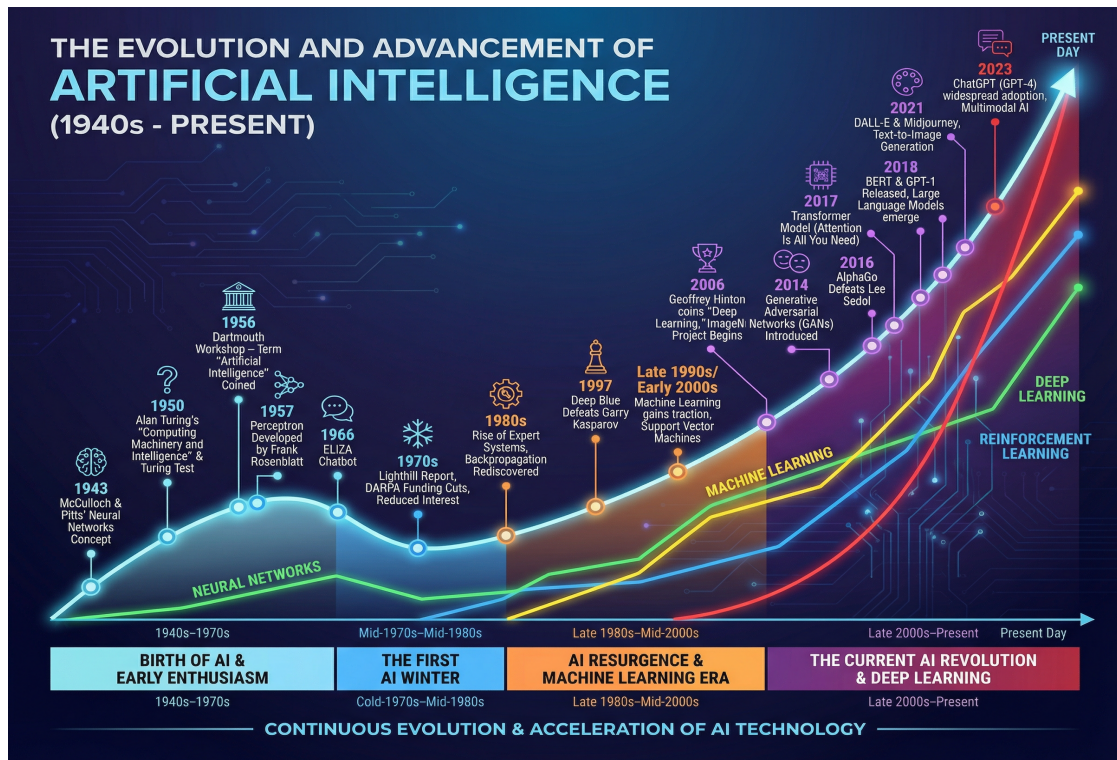


Figura 1.1: Linea temporale dell'evoluzione dell'Intelligenza Artificiale dal 1943 ai giorni nostri (fonte:easy-peasy.ai)

- *Claude (Anthropic)*: fondato da ex dipendenti di OpenAI, sono sistemi di Intelligenza Artificiale che si basano sul concetto di "AI costituzionale", un documento che serve a guidare la condotta dei dipendenti, ma anche degli stessi modelli di Claude;
- *Llama (Meta)*: il modello "open-source" più significativo; permette a sviluppatori e ricercatori di creare versioni personalizzate;
- *DeepSeek (DeepSeek-AI)*: è un modello emergente ed efficiente, valido nel ragionamento logico e matematico.

I principali modelli per generazione di immagini sono:

- *DALL-E (OpenAI)*: riconosciuto per la semplicità d'uso e l'integrazione con *ChatGPT*;
- *Midjourney (Midjourney Inc.)*: considerato di alta qualità per il fotorealismo;
- *Stable Diffusion (Stability AI)*: un modello open-source che può essere installato localmente, consentendo il controllo creativo.

Nella Tabella 1.1 sono riepilogati i principali sviluppatori e i relativi punti di forza.

1.1.3 Obiettivi dell'Intelligenza Artificiale

L'Intelligenza Artificiale non è da intendersi come la sola esecuzione di comandi, bensì come lo sviluppo di sistemi dotati di una propria autonomia decisionale. L'obiettivo risiede nel dare alle macchine la capacità di operare in modo analogo alla specie umana, superando il limite dei comandi diretti per raggiungere un'autonomia simile a quella che caratterizza la specie umana.

sviluppatore	modello principale	punti di forza
OpenAI	GPT-5	Versatilità e ragionamento complesso
Google	Gemini 3 / 3.1	Integrazione e analisi di grandi quantità di dati
Anthropic	Claude Opus 4.7	Scrittura naturale e sicurezza
Meta	Llama 4	Accessibilità e ricerca open-source
Mistral AI	Mistral Large 3	Efficienza e velocità

Tabella 1.1: Riepilogo dei principali sviluppatori di sistemi di Intelligenza Artificiale

Invece di limitarsi a fare calcoli rapidi, l'obiettivo profondo è permettere ai computer di comprendere il contesto. Proprio come accade con l'essere umano; quest'ultimo non necessita di istruzioni per riconoscere un animale o intuire le emozioni che sta provando un conoscente: ha un'intelligenza emotiva sviluppata e adibita a farlo. L'Intelligenza Artificiale punta a replicare questa capacità di analisi e intuizione per aiutare la specie umana dove essa trova più difficoltà, come l'analisi di quantità ingenti di dati in tempi brevi o la gestione di compiti ripetitivi senza mai dover riposare.

Oggi, però, gli obiettivi stanno evolvendo: non si vogliono più solo "assistenti", ma veri e propri collaboratori creativi. Lo scopo si è spostato sulla capacità di generare idee, scrivere testi o risolvere problemi, agendo come un moltiplicatore delle abilità.

Infine, il traguardo più ambizioso è rappresentato dall'Intelligenza Artificiale Generale. L'idea è quella di creare un sistema che non sappia svolgere solo determinati compiti, ma che, al contrario, possa imparare a fare qualsiasi cosa venga chiesto, adattandosi a situazioni nuove proprio come farebbe un essere umano.

In breve, l'obiettivo finale è passare da macchine che "eseguono" a macchine che "comprendono" il contesto e "risolvono"

1.2 Tipi di approccio all'Intelligenza Artificiale

Gli approcci all'Intelligenza Artificiale sono generalmente classificati in quattro categorie principali, definite nel celebre testo di Stuart Russell e Peter Norvig. *"Intelligenza artificiale. Un approccio moderno"*. Questi approcci si distinguono in base a due dimensioni: la prima tratta lo scopo dell'Intelligenza Artificiale, ovvero se l'obiettivo sia imitare l'essere umano o seguire una razionalità ideale, mentre la seconda definisce se il campo d'indagine sia il pensiero o l'azione.

1.2.1 L'approccio di Turing

Questo approccio si concentra sulla capacità di una macchina di comportarsi in modo indistinguibile da un essere umano.

Denominato come "approccio di Turing" poiché proposto da Alan Turing nel 1950, secondo questo metodo una macchina è "intelligente" se un esaminatore umano, conversando tramite testo, non riesce a distinguere se sta parlando con una persona umana o con un computer. Il problema di fondo di questo approccio è che i test che vengono proposti hanno delle regole e, quindi, sono inevitabilmente imperfetti. Ad esempio, la durata della conversazione con il soggetto è rilevante, poiché risulta più facile imitare un umano per cinque minuti piuttosto che per cento ore di conversazione. È importante osservare che la macchina potrebbe ingannare un essere umano in un breve lasso di tempo, mentre, in periodi di test più lunghi, ciò risulterebbe più complesso. Inoltre uno scienziato addestrato nel mettere alla prova la macchina risulta più difficile da ingannare rispetto ad un inesperto.

Questo tipo di test si concentra sui risultati, non sulle modalità con cui essi vengono ottenuti; superarlo può risultare notevole, soprattutto se è condotto da esperti e per un lungo periodo di tempo.

Nonostante i grandi progressi nel *Natural Language Processing*, ovvero la capacità di un sistema di comprendere il contesto di una frase pronunciata, la distanza da un'intelligenza generale che riesca ad imitare perfettamente l'essere umano è ancora ampia. Ciò è dovuto maggiormente dal fatto che, per riuscire nell'intento di imitare l'uomo, serve una quantità significativa di dati necessaria ad addestrare il sistema di Intelligenza Artificiale.

1.2.2 Approccio umano

L'approccio umano si riferisce al tentativo di creare un modello in grado di emulare il comportamento umano nella risoluzione dei problemi; esso non si limita ad imitare il comportamento esterno, ma prova a replicare anche i processi cognitivi interni del cervello umano basandosi sulla *modellazione cognitiva*, una tecnica che utilizza l'introspezione o elementi psicologici per capire come pensano gli umani e per trasferire le informazioni ricavate ad un programma informatico. Appartiene a questo approccio il Test di Turing.

1.2.3 Approccio razionale

L'approccio razionale è l'approccio prevalente oggi; si basa sull'utilizzo di modelli differenti da quelli che l'uomo utilizza per interfacciarsi ai problemi. È caratterizzato dalla volontà di creare modelli che abbiano la capacità di "*pensare razionalmente*". I modelli utilizzano una propria rappresentazione del mondo. In questo tipo di approccio la macchina agisce razionalmente affrontando problemi e cercandone la soluzione ottimale. I corrispettivi sistemi sono molto di più di semplici algoritmi; sono sistemi complicati e complessi che sbagliano, imparano e migliorano.

1.2.4 Differenze tra i due approcci

La differenza fondamentale tra i due approcci è da ricercare nell'essere umano, quest'ultimo nella vita di tutti i giorni, non è sempre razionale; al contrario, prende decisioni basate su emozioni, intuizioni rapide, piaceri e scorciatoie mentali che a volte conducono a conclusioni errate.

- un'Intelligenza Artificiale con approccio umano potrebbe sbagliare un calcolo complesso perché "distratta", proprio come una persona;
- un'Intelligenza Artificiale con approccio razionale risolverà il calcolo perfettamente, poiché il suo scopo è la perfezione tecnica, non la somiglianza con l'essere umano.

Ad oggi, l'informatica predilige l'approccio razionale, poiché si ritiene più utile avere sistemi che risolvono i problemi in modo ottimale ed efficiente piuttosto che macchine che simulano i nostri comportamenti.

Nella Tabella 1.2 vengono riepilogati i principali aspetti dei due approcci all'Intelligenza Artificiale.

1.3 Fondamenti interdisciplinari dell'Intelligenza Artificiale

L'Intelligenza Artificiale non è considerabile come una disciplina scientifica, ma come un'area di ricerca tecnologica che nasce dall'incontro di varie discipline che vanno a fondersi

Caratteristica	Approccio Umano (Empirico)	Approccio Razionale (Ideale)
Punto di riferimento	L'essere umano nella sua totalità	La logica e l'efficacia assoluta
Obiettivo	Imitare il pensiero delle persone	Ottenere il miglior risultato possibile
Criterio di successo	Se la macchina pensa come l'umano	Se raggiunge l'obiettivo con efficacia
Gestione degli errori	Tende a replicare gli errori umani	Evita errori utilizzando regole matematiche
Metodo di valutazione	Confronto diretto (es. Test di Turing)	Analisi delle prestazioni e della correttezza

Tabella 1.2: Riepilogo dei principali aspetti dei due approcci all'Intelligenza Artificiale

per sviluppare questo settore e arrivare a simulare l'intelligenza naturale. Le principali discipline che contribuiscono all'evoluzione dei sistemi di Intelligenza Artificiale sono riportati nelle prossime sezioni.

1.3.1 Matematica e statistica

La matematica da sempre rappresenta una diversa modalità di descrivere la realtà che ci circonda. Nell'ambito dei problemi di machine learning, i dati non sono dei valori singoli, ma vettori in N dimensioni. Quindi la conoscenza di strumenti come le matrici diventa fondamentale per organizzare e manipolare dati provenienti da sistemi di Intelligenza Artificiale.

Un esempio chiave è la regressione lineare, in cui si usano matrici e operazioni come la trasposizione per trovare parametri ottimali che consentono di minimizzare l'errore. Queste conoscenze sono necessarie, ad esempio, per algoritmi di classificazione e regressioni potenti, come SVM (Support Vector Machine) che supportano il *Machine Learning*.

Il calcolo differenziale entra, invece, in gioco quando si ha la necessità di ottimizzare le prestazioni di un modello. In particolare viene utilizzata la discesa del gradiente (vettore delle derivate parziali) per muoversi nella direzione in cui la funzione diminuisce.

Nella Figura 1.2 viene mostrata la definizione di gradiente di una funzione $f(x)$ in N variabili:

$$\nabla f(\mathbf{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(\mathbf{x}) \\ \vdots \\ \frac{\partial f}{\partial x_N}(\mathbf{x}) \end{bmatrix}$$

Figura 1.2: Definizione di gradiente di una funzione $f(x)$ in N variabili (fonte: ai4business.it)

Il calcolo integrale risulta importante soprattutto per comprendere la probabilità nel caso continuo. È essenziale capire concetti come il valore atteso, distinguendo tra variabili discrete (sommatorie) e continue (integrali). Questo aiuta a interpretare correttamente i modelli probabilistici. Nelle Figure 1.3 e 1.4 viene mostrato, rispettivamente, il calcolo del valore atteso di una variabile nel discreto e nel continuo.

$$\mathbb{E}(X) = \sum_{x \in \Omega} x P(X = x)$$

Figura 1.3: Calcolo del valore atteso di una variabile nel discreto, data la sua funzione di probabilità $P(X = x)$ (fonte: ai4business.it)

La probabilità e la statistica rappresentano il cuore del *Machine Learning*. Sono utilizzate per analizzare i dati e gestire le prestazioni dei modelli individuandone errori e accuratezza. È fondamentale l'argomento della distribuzione di probabilità utilizzato all'interno delle reti

$$\mathbb{E}(X) = \int_{\mathbb{R}} xp(x)dx$$

Figura 1.4: Calcolo del valore atteso di una variabile nel continuo, data la sua funzione di probabilità $p(x)$ (fonte: ai4business.it)

Bayesiane, che trovano fondamento nel teorema di Bayes. Nella Figura 1.5 viene mostrato il teorema di Bayes nella sua forma più semplice:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Figura 1.5: Teorema di Bayes nella sua forma più semplice (fonte: ai4business.it)

La formula sopra riportata, fornisce la possibilità di calcolare un evento A condizionato al fatto che si sia verificato l'evento B (anche detta probabilità di A condizionata a B).

1.3.2 Informatica e ingegneria del software

Tradizionalmente, l'ingegneria del software si occupa di tutto quello che è il ciclo di vita di un'applicazione, dalla progettazione alla manutenzione. Fornisce, quindi, all'Intelligenza Artificiale il contesto e gli strumenti per essere sviluppata e resa utilizzabile nei sistemi reali. Nello specifico, il contributo di tale disciplina si articola in diverse aree chiave:

- *architetture scalabili e robuste*: fornisce competenze per gestire architetture capaci di modellare carichi di dati elevati e garantire soluzioni sicure;
- *qualità e testing*: introduce un rigore per la validazione e il collaudo del software, assicurando che i sistemi di Intelligenza Artificiale funzionino come previsto;
- *integrazione e Deployment*: gestisce l'integrazione di algoritmi complessi all'interno di infrastrutture informatiche esistenti, come sistemi ERP (Enterprise Resource Planning);
- *affidabilità e manutenzione*: si occupa dell'evoluzione dei sistemi nel tempo, garantendo che le prestazioni rimangano costanti tramite processi di manutenzione predittiva.

1.3.3 Neuroscienze e modelli cognitivi

La relazione tra neuroscienze e Intelligenza Artificiale è sempre più forte. I sistemi di Intelligenza Artificiale traggono ispirazione dalle neuroscienze cognitive, avendo dall'inizio cercato un modello per gli algoritmi all'interno del funzionamento del cervello umano. A testimonianza di questo, nel 1943 McCulloch e Pitts crearono quello che venne ritenuto il primo lavoro inerente all'Intelligenza Artificiale; tale sistema impiegava, infatti, un modello di neuroni artificiali aventi uno stato "acceso" o "spento". Tale che, in presenza di stimoli causati dai neuroni circostanti, si otteneva un passaggio allo stato "acceso". La macchina ingegnata fu un successo in quanto si riuscì a dimostrare che qualsiasi funzione computabile poteva essere rappresentata da qualche rete di neuroni, e che tutti i connettivi logici "e", "o", potevano essere implementati da una struttura neurale.

Le ricerche dei due scienziati si sono basate sull'analisi dei meccanismi della ragione, evidenziando il ruolo della corteccia prefrontale, il cui funzionamento è stato descritto efficacemente attraverso un modello di tipo binario.

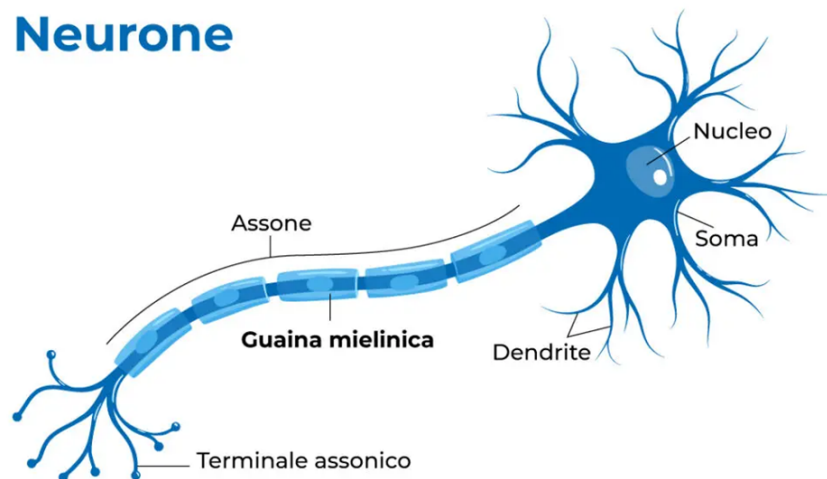


Figura 1.6: Neurone naturale (fonte: skillbook.it)

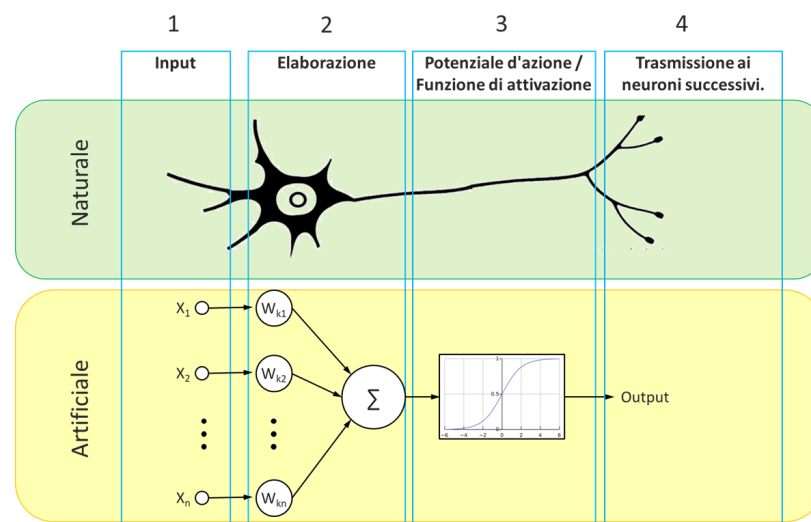


Figura 1.7: Confronto tra neurone e schema neurale per gli algoritmi (fonte: skillbook.it)

Nella Figura 1.6 viene rappresentato un neurone biologico nel dettaglio, mentre nella Figura 1.7 è rappresentato uno schema di confronto esaustivo tra un neurone e lo schema neurale inventato da McCulloch e Pitts.

La figura mostra chiaramente l'analogia che concorre tra neurone e schema neurale artificiale, evidenziando come gli input artificiali siano rappresentati da delle variabili che vengono conseguentemente, elaborate attraverso dei pesi sinaptici (W_{kn}). Questi pesi determinano l'importanza di ogni singolo input e vengono poi sommati nel nucleo del neurone (indicato dal simbolo Σ), simulando l'integrazione dei segnali che avviene nel corpo cellulare biologico.

Il valore ottenuto dalla somma passa attraverso una funzione di attivazione (spesso una sigmoide, come mostrato nel grafico). Questa decide se il segnale è abbastanza forte da "attivare" il neurone, replicando il meccanismo del potenziale d'azione naturale.

Infine, il risultato viene emesso come output. Questo segnale finale viene trasmesso ai neuroni successivi della rete, esattamente come l'assone trasmette l'impulso elettrico verso le sinapsi per comunicare con altre cellule.

1.3.4 Linguistica e elaborazione del linguaggio

L'elaborazione del linguaggio naturale (NLP, da *Natural Language Processing*) è una disciplina dell'Intelligenza Artificiale che descrive l'interazione tra macchine e linguaggio umano. L'obiettivo è rendere il computer in grado di comprendere il contenuto dei documenti e il loro contesto, in modo da estrapolare informazioni e categorizzare i file analizzati.

Per via di questo processo, particolarmente complesso a causa delle ambiguità che appartengono al linguaggio umano, la fase di elaborazione viene suddivisa in differenti fasi:

- *analisi lessicale*: la parola viene scomposta in token;
- *analisi grammaticale*: associazione delle parti del discorso a ciascuna parola nel testo;
- *analisi sintattica*: riorganizzazione dei token in una struttura sintattica, definita come "albero sintattico";
- *analisi semantica*: assegnazione di un significato alla parola.

Nella Figura 1.8 viene mostrata una rappresentazione vettoriale di come sono distribuite le parole all'interno di un documento, evidenziando le relazioni esistenti tra i vari termini di un documento

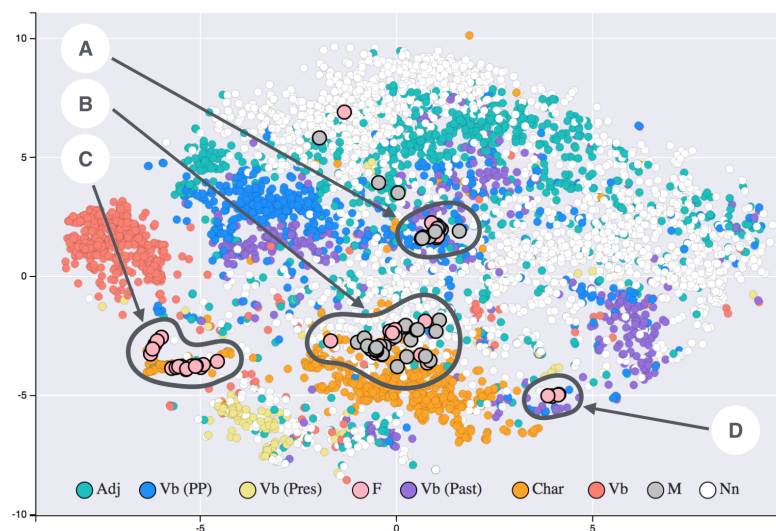


Figura 1.8: Rappresentazione vettoriale di come sono distribuite le parole all'interno di un documento con evidenziate le relazioni tra i termini (fonte: wikipedia.it)

Nonostante la suddivisione mostrata, l'elaborazione del linguaggio naturale rimane un compito arduo, dato che risulta intrinsecamente ambiguo.

Nel corso degli anni, per affrontare le problematiche menzionate, sono stati adottati vari approcci, dai modelli statistici fino alle più recenti reti neurali. Queste tecniche hanno trovato applicazione in numerosi settori, tra cui traduzione automatica, assistenti virtuali e analisi del sentimento.

1.4 Tecniche e approcci dell'Intelligenza Artificiale

Le tecniche e gli approcci all'Intelligenza Artificiale rappresentano le diverse strategie metodologiche e tecnologiche utilizzate per permettere alle macchine di emulare capacità umane, come l'apprendimento, il ragionamento e la risoluzione di problemi.

L'Intelligenza Artificiale viene storicamente suddivisa in due filosofie di pensiero:

- *Approccio simbolico (Top-Down)*: si basa sulla logica ad alto livello e sulla manipolazione di simboli combinabili in strutture simboliche. È tipico dei "sistemi esperti", dove la conoscenza umana viene codificata in regole rigide e architetture dichiarative; ad esempio: "SE il paziente ha la febbre E ha mal di gola, ALLORA controlla la presenza di placche", oppure "SE ci sono placche, ALLORA la probabilità di infezione batterica è dell'80%" e ancora "SE l'infezione è batterica, ALLORA consiglia l'uso di un antibiotico specifico". In sostanza, il computer non è consapevole di cosa sia il dolore, ma sa che qualora il simbolo "febbre" è presente, allora deve attivare la ricerca del simbolo "mal di gola" seguendo le regole della logica formale.
- *Approccio connessionista (Bottom-Up)*: risulta di più basso livello rispetto all'approccio Top-Down; tenta di imitare il funzionamento del cervello biologico tramite reti neurali artificiali. Invece di seguire regole preimpostate, il sistema impara dai dati e dall'esperienza. Su questa metodologia di approccio si fondano il *Machine Learning* e il *Deep Learning*.

La Figura 1.9, le relazioni tra Intelligenza Artificiale, *Machine Learning* e *Deep Learning*

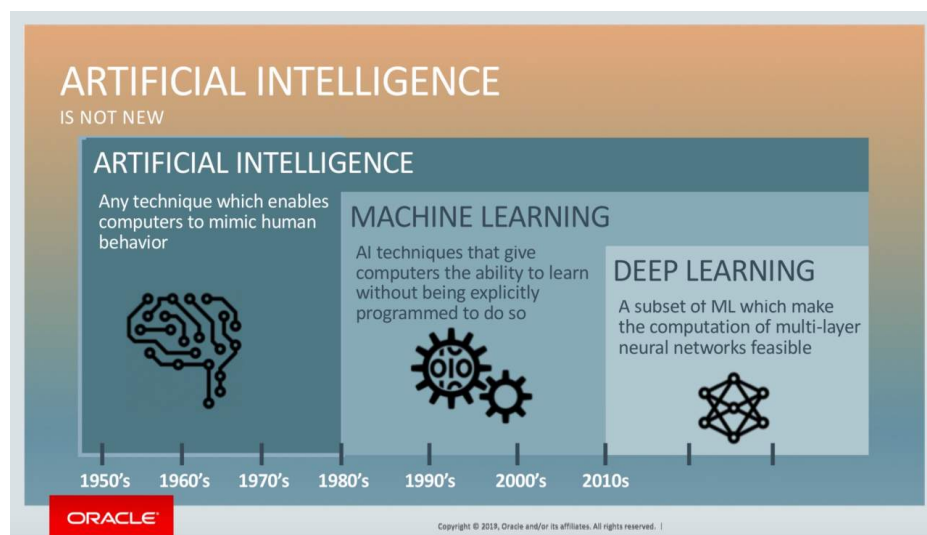


Figura 1.9: Rappresentazione delle relazioni tra Intelligenza Artificiale, *Machine Learning* e *Deep Learning* (fonte: oracle.it)

1.4.1 Introduzione al *Machine Learning*

Il *Machine Learning* (apprendimento automatico) è un ramo dell'Intelligenza Artificiale evoluzione del paradigma *Bottom-UP*, che sposta il paradigma della programmazione dallo spiegare alla macchina cosa fare alla macchina che impara ad agire osservando l'esperienza.

Nella programmazione tradizionale, il programmatore scrive algoritmi in successione per risolvere problemi, come se stesse seguendo una ricetta. Nel *Machine Learning* questo schema viene completamente sostituito. Invece di utilizzare regole rigide, si fornisce al computer una grande mole di dati, tramite la quale il sistema cercherà di individuare schemi ricorrenti per perseguire il suo obiettivo finale, ossia costruire un modello previsionale.

Il cuore del *Machine Learning* è un processo iterativo, in cui la macchina fa una prima previsione e, attraverso una funzione di perdita, misura di quanto si è discostata dalla risposta corretta, modificando i propri parametri interni per evitare di commettere gli stessi errori in

futuro. È un concetto quasi biologico: la macchina "impara" dai propri errori per ottimizzare una prestazione.

Il vero obiettivo finale del *Machine Learning* non è memorizzare i dati che vengono da esso, ma interiorizzare la capacità di generalizzare; se si ha la necessità di addestrare un algoritmo a riconoscere un determinato oggetto o persona, si pretende che tale sistema riconosca tutte le generalità relative ai dati su cui è stato addestrato.

Nella Figura 1.10 viene mostrato il ciclo di vita di un sistema di *Machine Learning*.

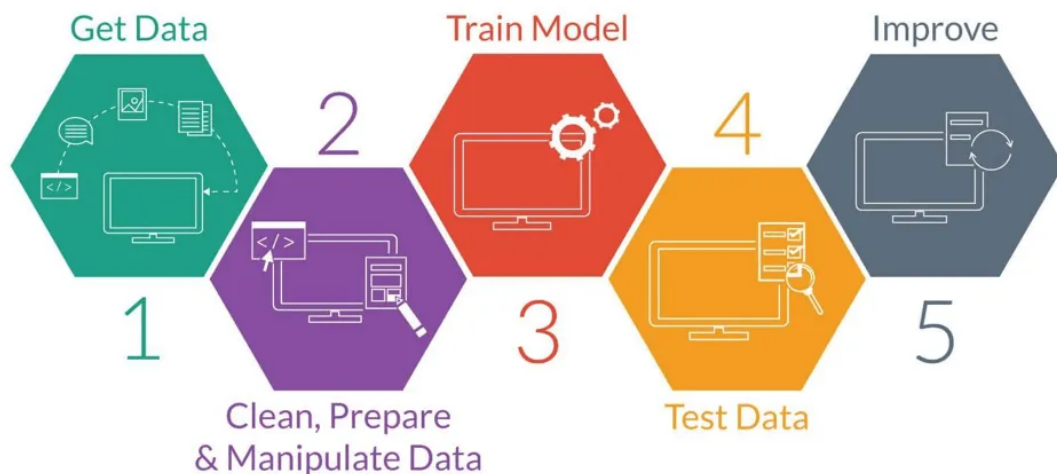


Figura 1.10: Rappresentazione del ciclo di vita di un sistema di *Machine Learning* (fonte: medium.com)

1.4.2 Tipologie di apprendimento

Nell'universo dell'Intelligenza Artificiale, trovano realizzazione più modalità di apprendere. La distinzione che viene riportata, dipende principalmente dalla natura dei dati a disposizione e dal ruolo che l'essere umano esercita nel processo di istruzione della macchina.

- *Apprendimento Supervisionato (Supervised Learning)*: è la forma di apprendimento più comune e intuitiva, paragonabile a uno studente guidato dal suo insegnante. La macchina viene addestrata su un'insieme di dati etichettato, ovvero dati di cui conosciamo già la risposta corretta, con l'obiettivo di creare una funzione capace di mappare nuovi input basandosi sulla storicità delle informazioni passate. Esempi di apprendimento supervisionato riguardano, fornire come "addestramento" una serie di e-mail etichettate come "spam" o "non spam", catalogarle e riconoscere i pattern già interiorizzati per la futura posta in arrivo.
- *Apprendimento Non Supervisionato (Unsupervised Learning)*: se in precedenza la forma di apprendimento era paragonabile a uno studente guidato, in questo caso la macchina agisce senza guida. Infatti, l'algoritmo riceve dati non etichettati, ossia senza risposte predefinite, in modo tale che il sistema debba trovare in totale autonomia la struttura logica che connette i dati, con lo scopo di trovare percorsi logici nascosti, che sarebbero passati inosservati a un'eventuale guida. Esempi di apprendimento "non supervisionato" riguardano fornire come dataset un insieme di clienti, così che il sistema possa analizzare i loro comportamenti d'acquisto e raggrupparli in categorie senza che siano state definite a priori.
- *Apprendimento Semi-Supervisionato (Semi-Supervised Learning)*: questa metodologia rappresenta un punto di incontro tra le precedenti, ideale quando etichettare i dati diventa

una procedura che si protrae nel tempo. Consiste nell'utilizzo di un piccolo set di dati etichettati unito a una enorme mole di dati non etichettati. In questa casistica, vengono sfruttati i pochi esempi certi per coordinare l'esplorazione dei dati incogniti, avendo un miglioramento delle prestazioni notevole rispetto al caso "non supervisionato". È una tecnica che viene utilizzata da colossi del web, come Google Photo, per il riconoscimento facciale; viene etichettato manualmente un piccolo insieme di persone e l'Intelligenza Artificiale successivamente identificherà automaticamente gli stessi volti in tutte le altre immagini.

- *Apprendimento per Rinforzo (Reinforced Learning)*: questo approccio si basa sul concetto di "trial and error". Funziona con dei feedback che possono essere positivi o negativi in base alle azioni che compie, con l'obiettivo di apprendere una strategia che ottimizzi il punteggio totale nel lungo periodo; ad esempio, si consideri un software che impara a giocare a scacchi; ogni mossa che lo avvicina alla vittoria viene associata ad un feedback positivo, rinforzando quel comportamento.

1.4.3 Introduzione al Deep Learning

Il Deep Learning è una sottocategoria del *Machine Learning* ispirata al funzionamento del cervello umano. Come mostrato nella sezione 1.3.3 le reti neurali artificiali sono composte da unità fondamentali di elaborazione chiamate "neuroni artificiali" perché imitano il funzionamento di quelli biologici.

L'aggettivo "Deep" (profondo) si riferisce alla struttura dell'algoritmo, che utilizza reti neurali artificiali con vari strati di elaborazione che vengono esplorati in profondità. A differenza del *Machine Learning*, il *Deep Learning* utilizza una struttura che arriva a contare fino a dieci o venti strati. Ogni layer (strato) è connesso agli altri e astrae in maniera sempre più precisa i dati, passando dal riconoscere oggetti semplicissimi (linee o bordi) a individuare oggetti interi (occhi, musi, volti).

È stato mostrato come il *Machine Learning*, nella maggior parte dei casi, richieda l'intervento umano; il *Deep Learning*, invece, è completamente autonomo e si basa su reti neurali artificiali composte da:

- *input layer*: luogo in cui entrano i dati grezzi;
- *hidden layers*: strati nascosti il cui numero risulta proporzionale all'apprendimento della rete neurale;
- *output layer*: rappresenta il risultato finale della rete.

Nel Deep Learning il processo di addestramento opera con una sequenza definita di fasi, ovvero:

- *inizializzazione*: la rete viene inizializzata con connessioni casuali, proprio come accade nei primi anni di vita dei bambini, periodi in cui non hanno la piena capacità di distinguere le cose;
- *alimentazione dei dati*: vengono fatti processare alla rete migliaia di esempi che attraverseranno la rete nella sua interezza producendo una risposta;
- *calcolo dell'errore*: viene confrontato il risultato con la risposta corretta;
- *retro propagazione (Backpropagation)*: è l'attimo chiave; l'errore viene propagato all'indietro attraversando la rete e aggiustando ogni connessione per non ripetere l'errore stesso in futuro;

- *ripetizione*: il processo viene ripetuto fino a che la rete non risulta accurata nei suoi risultati.

Un esempio celebre di sistema che utilizza le reti neurali è *ChatGPT*, il quale, usando un meccanismo chiamato *Attention*, permette alla rete di "concentrarsi" sulle parti importanti di un prompt esteso.

Il *Deep Learning* si è diffuso massicciamente solo nell'ultimo decennio grazie all'avvento dei Big Data e all'evoluzione delle GPU. La crescita esponenziale dei dati, unita alla capacità delle schede video di gestire i complessi calcoli paralleli richiesti dalle reti neurali, ha reso possibile l'attuale rivoluzione tecnologica.

1.5 Applicazioni e impatti dell'Intelligenza Artificiale

L'adozione dei sistemi di Intelligenza Artificiale ha trasformato radicalmente diversi settori, diventando uno strumento chiave nel mondo attuale. Queste tecnologie risultano così efficaci per via della capacità di elaborare informazioni eterogenee, consentendo l'automazione di compiti difficili da rendere tali fino a qualche anno fa, fornendo un prezioso contributo alla gestione dei dati.

L'utilizzo di questi modelli di apprendimento ha permesso di sviluppare soluzioni capaci di interagire con il mondo esterno. Questa interazione si mostra principalmente su tre aree di applicazione: l'elaborazione del linguaggio, la visione artificiale e il trattamento dei segnali vocali, argomenti che verranno trattati nelle prossime sezioni.

1.5.1 Applicazioni nel Natural Language Processing (NLP)

Il *Natural Language Processing (NLP)*, o elaborazione del linguaggio naturale, è il ramo dell'Intelligenza Artificiale che permette ai computer di comprendere, interpretare e generare il linguaggio umano, sia scritto che parlato, ossia rappresenta un ponte tra la comunicazione umana disordinata e ambigua e la comprensione tipica delle macchine.

Il linguaggio umano è una sfida ardua per un computer poiché risulta saturo di:

- *ambiguità*: ad esempio, "porta la borsa" per "porta" si intende l'oggetto o il verbo?;
- *sfumature e sarcasmo*: ad esempio, "Bello questo film!" può essere un complimento o una critica feroce, a seconda del contesto;
- *contesto*: il significato di una parola cambia in base a quelle che la circondano.

Questa branca dell'Intelligenza Artificiale utilizza algoritmi di *Machine Learning* per analizzare parole, strutture grammaticali e, persino, il sentimento dietro i termini utilizzati.

Come è stato già mostrato nei capitoli precedenti, nel *Natural Language Processing*, le parole vengono decomposte in *token*, eliminate di elementi comuni che non portano significato e ridotte alla loro radice; ad esempio, la frase "Il gatto sta correndo velocemente sui tetti." viene:

- *trasformata in token* in ["Il", "gatto", "sta", "correndo", "velocemente", "sui", "tetti", "."];
- *eliminata degli elementi superflui*: vengono rimosse ["il", "sta", "sui"], rimangono ["gatto", "correndo", "velocemente", "tetti"];
- *lemmatizzata*: "correndo" diventa "correre", "tetti" diventa "tetto".

Il risultato finale è ["gatto", "correre", "velocemente", "tetto"].

Come è facilmente osservabile, invece di analizzare una frase complessa, il computer ora ha una lista pulita di concetti chiave: gatto + correre + velocemente + tetto. Questo rende molto più facile per l'algoritmo confrontare questa frase con altre o classificarne l'argomento.

Alcune delle applicazioni più comuni del Natural Language Processing sono:

- *motori di ricerca*: Google, oltre a comprendere le parole chiave, capisce l'intento del nostro input;
- *filtri spam*: Gmail comprende se un'email è considerabile come "indesiderata", analizzandone il contenuto;
- *traduzione*: Google Translate utilizza il Natural Language Processing per rendere il testo fluido e naturale in un'altra lingua;
- *assistenza clienti*: i chatbot moderni risolvono i problemi senza la necessità dell'intervento umano;

L'*NLP* è, quindi, ciò che trasforma una macchina da una semplice calcolatrice a un interlocutore capace di dialogare, scrivere codici, riassumere libri e fornire assistenza. Il futuro di questa tecnologia si sta dirigendo verso la capacità di generare risposte che sembrano umane e, nel capire al meglio l'input dell'essere umano.

1.5.2 Applicazioni nella Computer Vision (Image Recognition)

La Computer Vision è un campo dell'Intelligenza Artificiale che insegna alle macchine a comprendere immagini e video allo stesso modo degli esseri umani, per poi prendere decisioni basate sulle informazioni estrapolate. Se per un sistema l'acquisizione di materiale, ad esempio tramite telecamere o sensori, risulta banale, il deliberare decisioni sulle informazioni estratte è un compito arduo, che viene supportato dalla Computer Vision. L'obiettivo principale di questa tecnologia è il riconoscimento, la classificazione e il tracciamento di elementi multimediali.

Il funzionamento della *Computer Vision* può essere diviso in cinque fasi:

- *acquisizione*: il primo passo consiste nell'ottenere l'input tramite qualsiasi tipo d'ingresso multimediale;
- *elaborazione*: l'oggetto in ingresso viene pulito dal "rumore", normalizzato e ridimensionato per facilitare l'analisi da parte del modello di Intelligenza Artificiale;
- *segmentazione*: l'input viene suddiviso in oggetti di interesse per permettere di mantenere in considerazione soprattutto le parti rilevanti;
- *estrazione*: identificazione di bordi, colori, *texture*, e infine, forme geometriche complesse;
- *riconoscimento e classificazione*: tramite algoritmi di *Machine Learning* vengono confrontate queste forme con i dati ricevuti durante l'addestramento per comprenderne l'output.

Anche la *Computer Vision*, come tanti altri sistemi di Intelligenza Artificiale, si è diffusa nella maggior parte delle attività reali; quali:

- *guida autonoma*: Tesla e altre auto usano la visione artificiale per riconoscere pedoni, cartelli e corsie in tempo reale;
- *sanità*: in alcuni stati l'analisi di radiografie è stata automatizzata tramite questi strumenti per ottimizzare l'individuazione di tumori o anomalie;

- *sicurezza*: sistemi di sorveglianza rilevano abbigliamento e comportamenti sospetti o intrusioni non autorizzate;
- *industria 4.0*: viene controllata la qualità nelle catene di montaggio per identificare difetti impercettibili all'occhio umano nei prodotti.

1.5.3 Applicazioni nello Speech Processing (Speech-To-Text)

Lo *Speech-To-Text*, noto anche come riconoscimento vocale automatico, è una tecnologia che si basa sull'Intelligenza Artificiale, capace di convertire il linguaggio parlato in testo scritto. È un sottoinsieme del *Natural Language Processing* e funziona trasformando le onde sonore catturate tramite input in dati digitali che un computer può interpretare e trascrivere in caratteri leggibili.

Questa tecnologia si basa su un processo complesso che sfrutta l'Intelligenza Artificiale e algoritmi avanzati di *Machine Learning*. Le fasi di cui si compone sono le seguenti:

- *acquisizione del segnale audio*: le vibrazioni sonore vengono interpolate come una serie di dati numerici, tramite dispositivi di input, come microfoni, cellulari o altre piattaforme di comunicazione, e archiviate per una successiva elaborazione;
- *pulizia dell'audio*: il sistema pulisce il segnale audio, riducendo il rumore di fondo e normalizzando il volume per migliorare la chiarezza dell'input;
- *elaborazione del segnale*: il suono viene trasformato in segnale digitale utilizzando la digitalizzazione audio che prevede la segmentazione della traccia in unità più piccole, come fonemi, che permettono al sistema l'elaborazione;
- *modellazione acustica e linguistica*:
 - *modello acustico*: associa i segmenti di suono ai fonemi corrispondenti;
 - *modello linguistico*: analizza il contesto per decidere quale parola sia la più probabile, ad esempio, aiuta il sistema a distinguere tra "hanno" e "anno" basandosi sulla grammatica della frase;
- *generazione dell'output*: il sistema assembla le parole identificate in una trascrizione testuale finale.

Lo *Speech-To-Text* viene utilizzato in molti ambiti anche conosciuti, quali:

- *assistenti vocali*: Siri di Apple, Alexa di Amazon e Google Assistant di Google utilizzano questa tecnologia;
- *dettatura e produttività*: il software Google Documenti, permette di scrivere testi parlando;
- *business*: programmi come Teams permettono la trascrizione automatica di riunioni e interviste.

Nella Figura 1.11 viene mostrato come grazie alla tecnologia *Speech-To-Text*, lo strumento "Amazon Echo" riesce a comprendere la richiesta dell'utente e ad eseguire il compito richiesto.

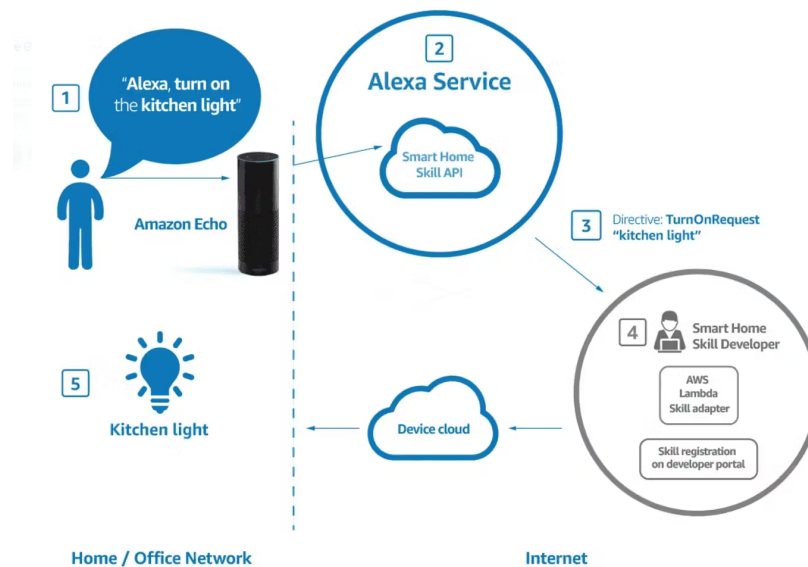


Figura 1.11: Rappresentazione dell'interazione con "Amazon Echo" tramite la tecnologia *Speech-To-Text* (fonte: enerzai.com)

1.5.4 Impatti, rischi, implicazioni etiche e prospettive future

Per il mercato occupazionale attuale, l'Intelligenza Artificiale non è solo una rivoluzione tecnologica, ma un cambiamento strutturale profondo che ridefinisce il concetto di "lavoro".

L'impatto di questi sistemi innovativi si manifesta sotto due principali aspetti: la sostituzione di mansioni specifiche e il potenziamento delle prestazioni umane.

Risulta largamente noto che l'Intelligenza Artificiale eccelle nell'automatizzare attività di routine, sia fisiche che cognitive.

Contrariamente a quanto si pensa, le nuove tecnologie non stanno sostituendo intere professioni, ma cambiando le singole mansioni. Se, da una parte, i lavori puramente esecutivi sono a rischio, dall'altra emergono ruoli di supervisione dell'Intelligenza Artificiale. In ambiti come la sanità, i computer sono in grado di agire come assistenti diagnostici, permettendo ai medici di avere un supporto ulteriore relativamente alle decisioni più critiche; alcuni esempi estremamente rilevanti riguardano:

- *La diagnosi precoce di sepsi:* l'insorgenza di *sepsi* (una condizione pericolosa per la vita del paziente) può essere prevista con varie ore di anticipo dai sistemi di *Intelligenza Artificiale*, consentendo un intervento tempestivo da parte dei medici.
- *La diagnosi del tumore al seno:* l'IA può supportare i radiologi umani nell'individuazione dei primi segni di tumore al seno con una notevole precisione.

Mantenendosi al passo con i tempi, è necessario sviluppare una riorganizzazione continua delle proprie competenze. Abilità prettamente umane come l'empatia, il pensiero critico, la creatività e l'etica diventano centrali, essendo difficilmente replicabili dagli algoritmi. Anche il saper interagire in modo ottimale con i sistemi intelligenti diventa una necessità trasversale a ogni settore, tenendo in considerazione la velocità di cambiamento che riduce notevolmente la durata utile delle conoscenze tecniche.

L'attuale ricerca si è direzionata sulla *Artificial General Intelligence*, ovvero sistemi capaci di eguagliare o superare l'intelletto umano in qualsiasi compito cognitivo. Il raggiungimento di un'Intelligenza Artificiale generale rappresenterebbe un punto di singolarità, garantendo sia

incrementi della produttività umana, ma anche sollevando sfide esistenziali sull'occupazione dell'uomo.

Nonostante i progressi, l'Intelligenza Artificiale presenta diverse barriere strutturali; ad esempio, non comprende il mondo allo stesso livello dell'essere umano; al contrario, funziona su base statistica e probabilistica; se viene posta davanti ad una situazione singolare, probabilmente questa la porterà al fallimento. Nei modelli di linguaggio (come *ChatGPT*), l'Intelligenza Artificiale può generare informazioni false, ma espresse in modo convincente, poiché il suo obiettivo è la coerenza testuale, non necessariamente la veridicità dei fatti. Addestrare e far funzionare in modo corretto i grandi modelli di *IA* richiede un impatto ambientale significativo in termini di elettricità e di consumo di acqua per il raffreddamento delle macchine ospitanti.

Per quanto concerne i rischi e le implicazioni etiche, questi argomenti rappresentano un punto critico di questa tecnologia innovativa; ad esempio, le tecnologie di riconoscimento facciale e l'analisi predittiva dei comportamenti sollevano molti dubbi sulla libertà della persona e sul controllo di massa da parte dei governi. La capacità di creare immagini, video e audio falsi sempre più verosimili rende difficile distinguere la realtà dalla manipolazione, con rischi gravi per la democrazia e la reputazione delle persone coinvolte. L'introduzione di sistemi decisionali autonomi è sia un vantaggio in termini di tempo che un rischio a livello etico, questo perché una valutazione puramente algoritmica potrebbe prevalere sulla moralità necessaria da adottare in contesti come questo.

In conclusione, si osserva come l'Intelligenza Artificiale possa liberare tempo e aumentare la produttività aziendale, ma anche prendere decisioni basate unicamente sui numeri, tralasciando completamente l'aspetto emotivo ed etico, generando criticità e discussioni sulla sua adozione.

1.6 L'evoluzione dell'Intelligenza Artificiale nel Cloud

L'Intelligenza Artificiale unendosi al *Cloud* ha segnato una vera e propria evoluzione, diventando una tecnologia d'élite invisibile e pronta all'uso.

Nelle prime fasi dell'evoluzione, le aziende acquistavano potenza di calcolo in *Cloud* per addestrare modelli che, altrimenti, avrebbero fuso i computer dell'ufficio. In questa fase era necessario conoscere perfettamente come configurare un server per poter ospitare anche il più piccolo algoritmo.

I grandi provider (AWS, Google Cloud, Azure) compresero, vista la incrementale diffusione, la necessità di servizi cognitivi pronti all'uso; in questo modo, qualsiasi sviluppatore avrebbe avuto la possibilità di aggiungere alla propria applicazione sistemi di Intelligenza Artificiale senza avere la minima conoscenza relativa alle reti neurali.

Oggi, ci troviamo nel momento più espansivo. Con l'introduzione dei Large Language Model (come *GPT-5* o *Claude*), il Cloud non offre più solo strumenti per mettere in piedi queste tecnologie, ma vere e proprie interfacce utilizzabili da chiunque.

Questo capitolo tratta la tecnologia Amazon Web Services (AWS), introducendone la genesi storica e l'infrastruttura globale per poi approfondire il concetto di regioni, zone di disponibilità e sicurezza dei dati nel sistema Cloud. Successivamente, viene proposta una disamina dell'insieme di sistemi tecnologici dedicati all'Intelligenza Artificiale, con dovizia di particolari, relativi a: risorse di calcolo accelerato, piattaforma Amazon SageMaker e servizio Amazon Bedrock. Il capitolo prosegue con i principali servizi cognitivi pronti all'uso, ossia, Amazon Comprehend, Amazon Rekognition e Transcribe; essi costituiscono la base tecnologica per i casi studio che verranno trattati nei capitoli successivi. Infine, lo studio si conclude con una valutazione dei vantaggi strategici e dei limiti nell'adozione del Cloud per lo sviluppo di sistemi intelligenti.

2.1 Genesi ed evoluzione del modello *Cloud* di Amazon

L'evoluzione di Amazon Web Services (AWS) è uno dei casi studio per eccellenza della tecnologia moderna, poiché ha rappresentato la trasformazione di un costo operativo (l'infrastruttura IT) in un business da miliardi di dollari.

L'ascesa di AWS iniziò quando l'azienda si rese conto che l'aggiunta di nuove funzionalità al suo sito web lo rendeva estremamente lento e inutilizzabile. I team di Ingegneria del Software del tempo si dedicarono, quindi, a costruire infrastrutture che fossero in grado di supportare e scalare un colosso come Amazon; tuttavia, formare questi elementi era complesso, a un livello tale che i nuovi progetti costituirono solo una base per le nuove funzionalità interessanti che i clienti stavano cercando.

Selipsky, oggi CEO di AWS, ricorda che i dirigenti dell'azienda iniziarono a pensare di costruire un livello condiviso di servizi infrastrutturali sui quali tutti i team potessero fare affidamento; in questo modo, nessuno di loro avrebbe dovuto dedicare tempo a funzionalità generali come lo storage nei database o la capacità di calcolo.

Un ulteriore problema riguardava i siti web terzi che volevano aggiungere collegamenti ai prodotti Amazon nelle loro pagine. Ad esempio, ai portali che consigliavano un prodotto veniva inviato un link di affiliazione, tramite il quale potevano guadagnare una commissione sugli acquisti. Tuttavia, con il passare del tempo, gli sviluppatori vollero sempre più libertà; da qui nacque un'intuizione importante: gli sviluppatori esterni avevano gli stessi problemi di quelli interni ad Amazon.

Nel 2003, durante una riunione, i dirigenti compresero i problemi e capirono che quello che stavano affrontando sarebbe potuto diventare un modello di business.

La tecnologia che si posizionava nei retroscena di Amazon venne così evoluta in AWS, una suite di servizi totalmente dedicata al mercato degli sviluppatori. Tramite essa, le aziende non dovevano più acquistare interi server costosi, ma potevano utilizzare il cloud di Amazon pagando esclusivamente ciò di cui usufruivano.

La successiva evoluzione del modello cloud di Amazon si divide in queste fasi:

- 2002: viene lanciato il primo servizio, Amazon SQS (Simple Queue Service), un sistema per gestire i messaggi tra applicazioni;
- 2006: AWS debutta ufficialmente con Amazon S3 (storage) e poco dopo con Amazon EC2 (potenza di calcolo);
- *espansione Globale (2008-2012)*: AWS ha iniziato a diversificare geograficamente e a offrire servizi più complessi, come database gestiti e reti di distribuzione contenuti (CloudFront);
- *dall'Infrastruttura all'Intelligenza (2014-Oggi)*: l'evoluzione si è spostata verso servizi di livello superiore, come il Serverless (AWS Lambda nel 2014) e, più recentemente, l'Intelligenza Artificiale Generativa.

La Figura 2.1 mostra la timeline dell'evoluzione di AWS con alcuni dei servizi sviluppati.

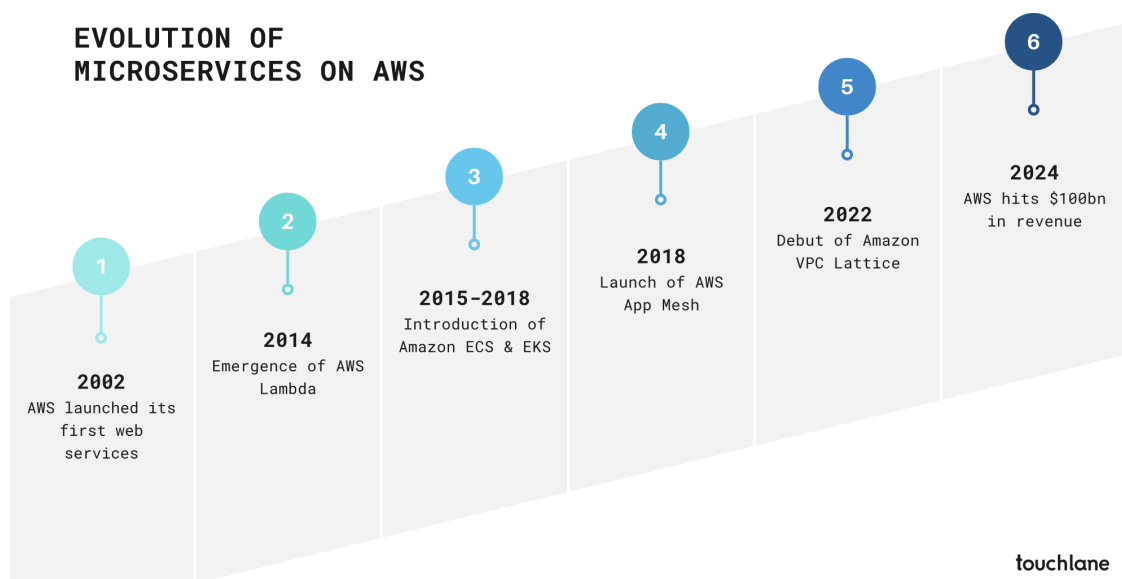


Figura 2.1: Schema temporale dell'evoluzione di AWS fino ad oggi (fonte: touchlane.com)

2.2 Architettura e infrastruttura globale

Amazon Web Services si sviluppa tramite una divisione in regioni, questa architettura è stata progettata per offrire massima affidabilità, scalabilità e bassa latenza. Si basa su un modello gerarchico che parte da aree geografiche vaste (regioni), fino ad arrivare a punti di consegna vicini agli utenti finali, collegati da oltre 9 milioni di chilometri di cavi in fibra ottica. Per ogni regione, poi, sono presenti tre zone di disponibilità.

Il Cloud AWS comprende in tutto 123 zone di disponibilità all'interno di 39 Regioni Geografiche, con piani già annunciati per altre 7 zone di disponibilità e altre 2 regioni AWS nel Regno dell'Arabia Saudita e in Cile.

2.2.1 Concetti di regioni e zone di disponibilità

Una regione è un'area geografica fisica isolata nel mondo che raggruppa diversi data center, che ci permette di ospitare un'applicazione, ad esempio, Milano, Dublino o Tokyo; queste, oltre ad essere completamente autonome per garantire che un guasto non influenzi le altre, permettono ai clienti di scegliere dove archiviare i dati garantendo il massimo rispetto delle normative locali sulla privacy, fornendo, anche, la sicurezza agli utenti che i loro dati non escano dalla regione scelta.

Le zone di disponibilità (*Availability Zones*) sono data center ad alta disponibilità all'interno di ciascuna regione; ognuna di queste zone ha alimentazione, raffreddamento e rete indipendenti. Quando un'intera *Availability Zone* va fuori uso, AWS è in grado di trasferire i carichi di lavoro su una delle altre zone che si trovano nella stessa regione; questa particolare capacità è nota come "Multi-AZ".

Nella Figura 2.2 viene mostrata una panoramica dell'architettura delle regioni e delle zone di disponibilità.

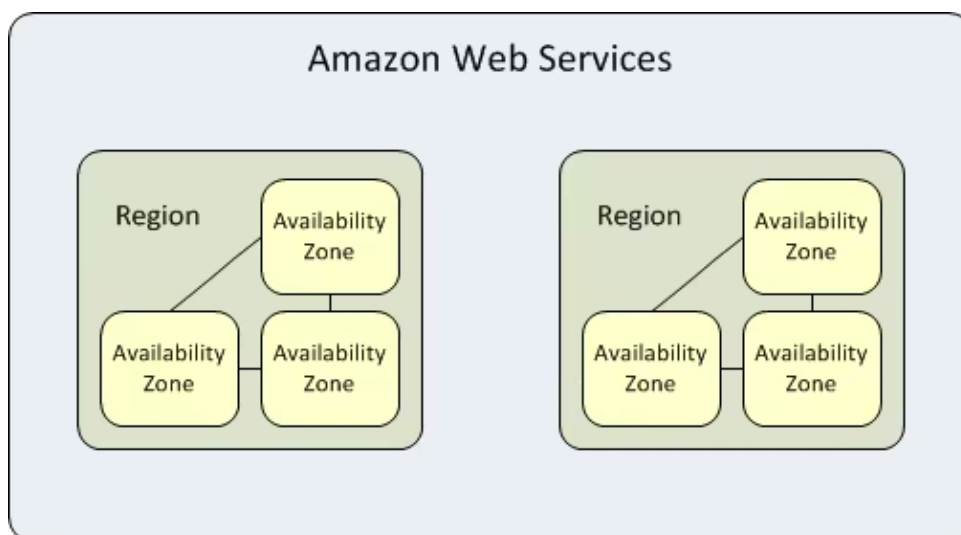


Figura 2.2: Panoramica dell'architettura delle regioni e delle zone di disponibilità di AWS (fonte: docs.aws.amazon.com)

Prima di scegliere una regione AWS per ospitare la propria applicazione è importante fare alcune considerazioni tenendo presente alcuni parametri:

- *latenza e prossimità*: scegliere una regione vicina al cliente permette di avere una connessione veloce al server, garantendo, quindi, una migliore esperienza utente;
- *costo*: le regioni hanno costi differenti dipendenti dalle tasse dei paesi ospitanti; risulta, quindi, opportuno effettuare una stima del miglior rapporto prezzo-prestazioni;
- *conformità normativa e sicurezza*: ogni paese ospitante ha le proprie norme e regole per proteggere i dati degli utenti; risulta, quindi, fondamentale rispettarle.

2.2.2 Sicurezza e conformità dei dati nel Cloud

La sicurezza e conformità dei dati nel Cloud all'interno del mondo AWS è fondamentale, poiché questi due principi si basano su un approccio strutturato che combina infrastruttura fisica, strumenti software avanzati e una chiara suddivisione delle responsabilità tra il fornitore e l'utente.

Il concetto fondamentale in AWS è che la sicurezza è rappresentata da uno sforzo congiunto tra:

- *sicurezza "DEL" Cloud (responsabilità di AWS)*: Amazon Web Service è il principale responsabile della protezione dell'infrastruttura globale che esegue tutti i servizi (hardware, software, networking e strutture fisiche, come i data center);
- *sicurezza "NEL" Cloud (responsabilità del cliente)*: l'utente è responsabile della gestione dei dati, delle configurazioni di rete (firewall), della gestione delle identità (IAM) e della crittografia dei dati che inserisce nel Cloud.

La Figura 2.3 illustra la differenziazione della responsabilità tra utente e Amazon.

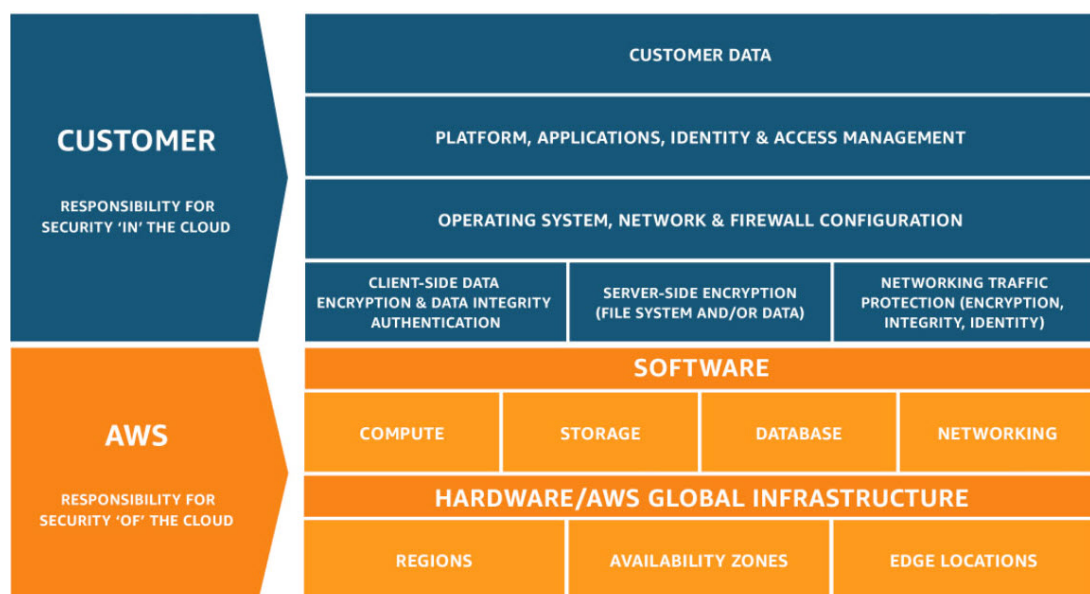


Figura 2.3: Illustrazione della responsabilità "DEL" Cloud (responsabilità di AWS) e "NEL" Cloud (responsabilità del cliente) (fonte: aws.amazon.com)

Per quanto concerne la sicurezza dei dati, AWS fornisce diversi servizi dedicati; uno di questi è la gestione delle identità IAM (*Identity and Access Management*). Questo servizio permette al cliente di controllare con precisione chi può accedere a quali risorse, utilizzando account AWS multipli, credenziali di sicurezza temporanee e autorizzazioni eseguendo il principio del "minimo privilegio", ossia concedendo solo i permessi strettamente necessari.

Alcuni degli ulteriori strumenti dedicati alla protezione dell'infrastruttura sono rappresentati da AWS Shield, progettato per contrastare attacchi DDoS (*Distributed Denial of Service*), e AWS WAF (*Web Application Firewall*) utilizzato per proteggere le applicazioni web dai comuni bot che possono influire sulla disponibilità o compromettere la sicurezza.

Per dimostrare la propria affidabilità, AWS si sottopone regolarmente a audit di terze parti e supporta oltre 143 standard di sicurezza e certificazioni. Tra questi citiamo:

- *certificazioni globali*: includono standard come ISO 27001 relativi alla gestione della sicurezza;
- *regolamentazioni locali*: AWS fornisce strumenti e contratti (come il *Data Processing Addendum*) per aiutare le aziende a rispettare il GDPR in Europa.

2.3 Modelli di servizio di AWS

Per comprendere appieno il mondo AWS è fondamentale analizzare i diversi livelli di astrazione del *Cloud Computing*: *Infrastructure as a Service (IaaS)*, *Platform as a Service (PaaS)* e *Software as a Service (SaaS)*. Questi paradigmi rappresentano dei veri e propri strumenti necessari per gestire l'elevato carico computazionale richiesto dai modelli di Intelligenza Artificiale, poiché più ci si direziona verso il SaaS, meno responsabilità tecniche ricadono sull'azienda e maggiore è la semplicità d'uso.

- *IaaS (Infrastructure as a Service)*: è il livello base di AWS; infatti, tramite questa tecnologia, vengono forniti server virtuali, storage e rete. L'utente ha il controllo totale sul sistema operativo e sulle applicazioni, ma ha anche il compito di gestire gli aggiornamenti e la sicurezza del software. Ad esempio, utilizzare il servizio Amazon EC2 (*Elasti Computer Cloud*) significa affittare un vero e proprio computer all'interno di un data center, con annesse spese relative all'utilizzo.
- *PaaS (Platform as a Service)*: in questo caso, AWS gestisce l'infrastruttura sottostante il server, ossia, sistema operativo e hardware; l'utente si deve preoccupare solo di caricare il codice della propria applicazione. Questo servizio rappresenta l'ideale per sviluppatori che vogliono essere veloci e hanno necessità di lasciare ad Amazon il compito di configurare le patch di sicurezza del server o della rete. Ad esempio, *AWS Elastic Beanstalk* lascia la possibilità al programmatore di caricare il proprio codice e impegna AWS nella gestione del bilancio del carico e della scalabilità.
- *SaaS (Software as a Service)*: è il modello che descrive i prodotti finiti che mette a disposizione AWS; non è compito del cliente preoccuparsi di come il servizio venga sostenuto o di come sia costruito; il software distribuito viene semplicemente utilizzato. Alcuni esempi sono rappresentati da Amazon Chime (per le videoconferenze) o Amazon Connect (per i call center), anche se molti SaaS famosi sono esterni ad AWS (come Gmail).

La Figura 2.4 rappresenta i tre modelli di servizio di AWS con una breve descrizione dell'utilizzo relativo.

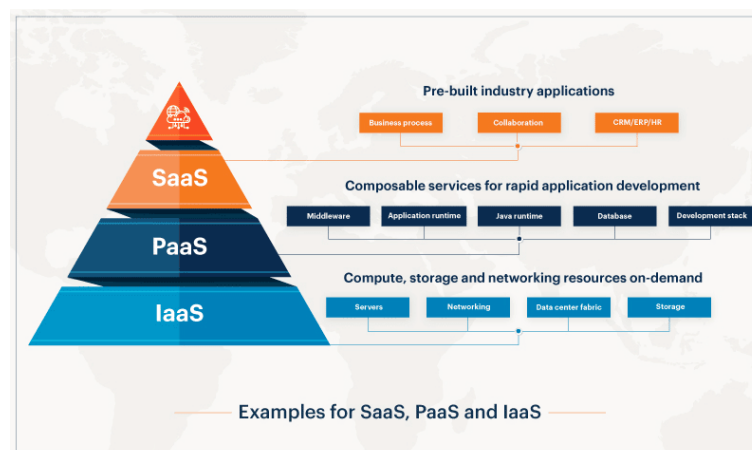


Figura 2.4: Illustrazione dei tre modelli di servizio di AWS con annessa breve descrizione dell'utilizzo (fonte: kinsta.com)

2.4 L'ecosistema per l'Intelligenza Artificiale e il Machine Learning

AWS non è solo un'infrastruttura, ma un vero e proprio ecosistema completo per il *Machine Learning*.

La filosofia di AWS si basa sull'idea di rendere le tecnologie di Intelligenza Artificiale e *Machine Learning* accessibili a ogni sviluppatore, data scientist e azienda, indipendentemente dalle loro dimensioni o competenze tecniche iniziali. Questa visione mira a rimuovere le barriere avvicinando anche i meno esperti al settore.

2.4.1 Gestione della potenza computazionale tramite istanze scalabili e acceleratori hardware

AWS gestisce la potenza di calcolo principalmente attraverso il servizio Amazon EC2 (*Elastic Compute Cloud*) facente parte del livello *Infrastructure as a Service (IaaS)*; questo offre oltre 750 istanze di processori, sistemi di archiviazione, reti e sistemi operativi.

Amazon EC2 è nato per soddisfare al meglio le esigenze di supporto del grande carico di lavoro delle aziende, con una rete che raggiunge i 400 Gbps; ad oggi è il miglior provider di servizi per il training di *Machine Learning*.

Le caratteristiche principali di questo servizio sono:

- *scalabilità*: è possibile aumentare o diminuire il numero di server in pochi minuti per gestire i picchi di traffico o ridurre i costi quando la domanda diminuisce;
- *pagamento a consumo*: il servizio è stato pensato per far pagare ai clienti solo quello che utilizzano effettivamente, evitando, così, investimenti iniziali importanti;
- *controllo totale*: l'amministratore ha accesso ad ogni istanza, proprio come se ne fosse il proprietario.
- *varietà di opzioni*: le istanze messe a disposizione coprono un'ampia varietà; ognuna di queste è ottimizzata per scopi diversi.

Questo servizio offre server virtuali di ogni tipo chiamati "istanze" e modelli preconfigurati (template) che includono il sistema operativo e il software necessario per avviarle, permettendo al cliente in totale autonomia di scegliere quella più adatta a lui; tra le istanze troviamo:

- *Istanze per uso generico (general purpose)*: questo tipo di istanze assicurano un equilibrio tra le varie risorse hardware e possono essere utilizzate per una vasta gamma di carichi di lavoro differenti.
- *Istanze ottimizzate per il calcolo (computed optimized)*: sono le migliori per le applicazioni che traggono vantaggio da processori ad elevate prestazioni, ad esempio, appartengono a questa categoria applicazioni che utilizzano l'elaborazione in batch, la transcodifica multimediale e i server di gioco dedicati.
- *Istanze ottimizzate per la memoria (memory optimized)*: progettate per erogare prestazioni a carichi di lavoro che utilizzano grandi set di dati; ad esempio, software per analisi dei dati e applicazioni aziendali.
- *Istanze per calcolo accelerato (accelerated computing)*: utilizzano acceleratori hardware o coprocessori per eseguire le funzioni in modo più efficiente; ad esempio, le GPU (*Graphic Processing Unit*) sono ideali per il calcolo dei numeri in virgola mobile.

- *Istanze ottimizzate per l’archiviazione (storage optimized)*: sono istanze ottimizzate per l’archiviazione e forniscono milioni di operazioni di input e output a bassa latenza al secondo.
- *Istanze ottimizzate per HPC (hpc optimized)*: create appositamente per offrire il miglior rapporto qualità-prezzo; esse sono ideali per applicazioni che traggono vantaggio da processori ad alte prestazioni, come simulazioni complesse, deep learning e rendering di effetti visivi.

Ogni famiglia di istanze EC2 è progettata con un hardware specifico per "accelerare" un carico di lavoro particolare. In AWS non esiste un acceleratore universale; la scelta finale dipende dal tipo di esigenza del cliente e dall’uso che quest’ultimo ne vuole fare.

Scegliere e configurare manualmente queste risorse per il ciclo di vita completo del *Machine Learning* risulta complesso e dispersivo. Per questo motivo, AWS fornisce una serie di prodotti *Platform as a Service (PaaS)* che garantiscono un’elevata adattabilità ai carichi di lavoro eseguiti dal cliente.

2.4.2 La piattaforma Amazon SageMaker per lo sviluppo di modelli

Amazon SageMaker è un servizio *PaaS* di Amazon Web Services (AWS) completamente gestito che semplifica l’intero ciclo di vita del *Machine Learning*. Il servizio include diverse *Application Program Interface (API)* che consentono a sviluppatori e data scientist di preparare, creare, addestrare e distribuire modelli di alta qualità su scala in modo più veloce e preciso.

Per comprendere chiaramente come funzioni SageMaker e come semplifichi il ciclo di vita del *Machine Learning*, viene riportato di seguito un elenco di priorità che caratterizzano lo sviluppo di questo tipo di modelli, rispettivamente senza e con il supporto della piattaforma:

- *Caratteristiche di vita senza SageMaker*:
 - *Dati dispersi*: necessità di reperire informazioni da sorgenti eterogenee, con operazioni manuali di normalizzazione e archiviazione locale.
 - *Ambiente frammentato*: l’installazione di numerosi pacchetti incrementa il rischio di conflitti tra versioni e incompatibilità tra librerie.
 - *Training manuale*: l’addestramento richiede ripetuti cicli di esecuzione, debugging e analisi manuale dei risultati.
 - *Inferenza complessa*: sono richieste conversione manuale dei modelli e la configurazione dei server senza automazione.
 - *Monitoraggio assente*: vi è mancanza di visibilità sullo stato dei modelli in produzione e sui relativi indicatori di performance.
- *Ciclo di vita con SageMaker*:
 - *Dati centralizzati*: l’integrazione nativa con Amazon S3 permette un accesso automatizzato ai dataset, eliminando i trasferimenti manuali.
 - *Ambiente unificato*: SageMaker Studio fornisce un IDE con notebook e strumenti preconfigurati e pronti all’uso.
 - *Training orchestrato*: la piattaforma automatizza l’approvvigionamento delle risorse e il debugging, garantendo il tracciamento di ogni esecuzione.
 - *Pubblicazione semplice*: è possibile creare rapidamente degli endpoint scalabili con gestione automatica dell’infrastruttura.

- *Governance integrata*: è possibile il monitoraggio automatico delle versioni e dei metadati, assicurando il controllo di qualità su ogni operazione.
- *Monitoraggio continuo*: integrazione con CloudWatch per l’osservabilità delle performance e la gestione degli avvisi in tempo reale.

Nella Figura 2.5 viene riportato il ciclo di vita del *Machine Learning* con SageMaker in modo chiaro.

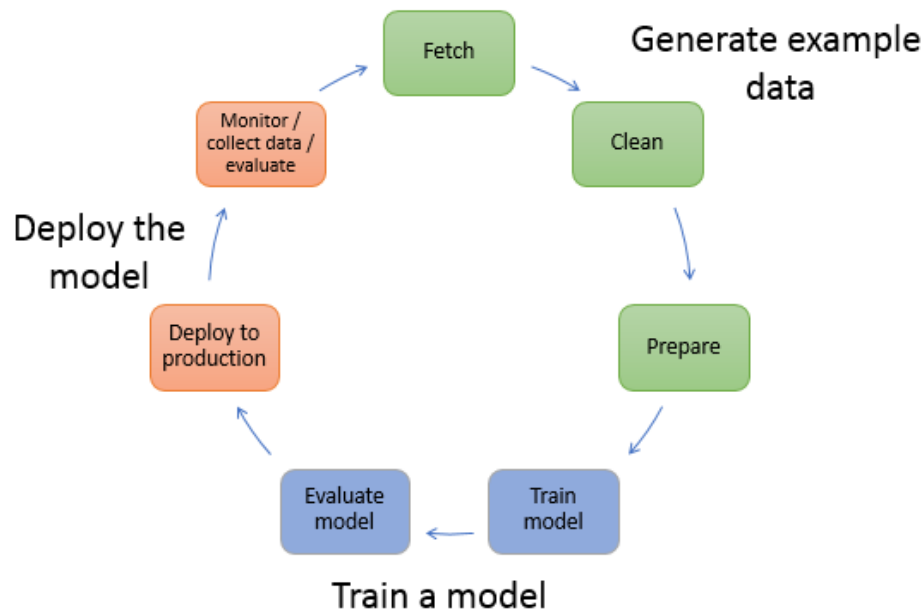


Figura 2.5: Ciclo di vita del *Machine Learning* con Amazon SageMaker (fonte: aws.amazon.com)

Ricordando che un software di *Machine Learning* risulta molto utile in casi in cui si ha la necessità di automatizzare abilità umane quali il riconoscimento di determinati pattern, e sottolineando che per configurare questo tipo di sistemi complessi sono necessari molto tempo e abilità, SageMaker risulta una valida soluzione per via della sua facilità di utilizzo, sviluppo, configurazione e individuazione di pattern complessi in volumi di dati estremamente grandi. Infatti, esso viene utilizzato in diversi ambiti come:

- *sanità*: analisi dei dati relativi ai pazienti, previsione dei risultati e personalizzazione delle terapie;
- *finanza*: rilevamento delle frodi e valutazione o rischi di investimento;
- *vendita*: analitiche predittive per gli inventari, personalizzazione del cliente e ottimizzazione dei prezzi.

2.4.3 Intelligenza Artificiale generativa ready-to-use: *Amazon Bedrock*

Amazon Bedrock si configura come un ulteriore servizio gestito in modalità *PaaS*. Utilizzando questa risorsa, gli sviluppatori aziendali non devono gestire architetture di calcolo o configurare complessi parametri infrastrutturali; l’ambiente astrae l’intera architettura hardware, permettendo agli sviluppatori di integrarsi direttamente con i modelli pre-addestrati tramite semplici chiamate API.

Bedrock fornisce la possibilità di accedere a modelli di Intelligenza Artificiale generativa sviluppati dalle organizzazioni più note, permettendo la creazione di soluzioni su misura.

Le aziende, utilizzando questo servizio, hanno la possibilità di ottimizzare i modelli di Intelligenza Artificiale per soddisfare determinate loro specifiche senza dover sviluppare una soluzione da zero. Ciò permette l'accelerazione dei tempi di lancio dei nuovi prodotti e servizi.

I provider principali che mettono a disposizione i loro modelli "di fondazione" alle aziende che utilizzano il servizio Amazon Bedrock, e sui quali esse possono operare, sono:

- *Amazon* con i seguenti modelli: Amazon Nova Multimodal Embeddings, Nova 2 Lite, Nova 2 Sonic, Nova Canvas, Nova Lite, Nova Micro, Nova Premier, Nova Pro, Nova Reel, Nova Sonic, Titan Embeddings G1 - Text ecc.;
- *Anthropic* che mette a disposizione: Claude Opus 4.7, Claude Opus 4.6, Claude Sonnet 4.6, Claude Haiku 4.5, Claude Opus 4.5, Claude Sonnet 4.5, Claude Sonnet 4, Claude Opus 4.1 ecc.;
- *DeepSeek* con: DeepSeek V3.2, DeepSeek-V3.1, e DeepSeek-R1;
- *Google* che fornisce: Gemma 3 12B IT, Gemma 3 27B PT e Gemma 3 4B IT;
- *Meta* che fornisce: Llama 3.3 70B Instruct, Llama 3.2 11B Instruct, Llama 3.2 1B Instruct, Llama 3.2 3B Instruct, Llama 3.2 90B Instruct, Llama 3.1 405B Instruct, Llama 3.1 70B Instruct, Llama 3.1 8B Instruct ecc. ;
- *OpenAI* con: GPT OSS Safeguard 120B, GPT OSS Safeguard 20B, gpt-oss-120b e gpt-oss-20b.

Amazon Bedrock prevede tre modalità principali di utilizzo:

- *pronto all'uso via API*: è la modalità ideale per un utilizzo immediato dei modelli di fondazione così come sono; inviando un prompt tramite interfaccia API, si riceve la risposta del modello pre-addestrato;
- *personalizzazione*: entra in gioco quando il modello di base non è sufficiente; esso viene riaddestrato fornendo un insieme di dati etichettati per modificarne i pesi e far apprendere al modello compiti specifici;
- *knowledge bases*: collega il modello ai database aziendali per estrarre informazioni aggiornate dai documenti; questa tecnica non modifica direttamente il modello, ma ne arricchisce il contesto.

I casi d'uso per questo servizio sono molteplici poiché, partendo dai modelli di fondazione, è possibile generare qualsiasi soluzione su misura. Uno dei casi d'uso più rilevanti riguarda l'assistenza clienti; in questa modalità Amazon Bedrock è capace di superare i chatbot tradizionali eseguendo azioni reali nei sistemi aziendali. Ad esempio, se un cliente chiedesse di gestire il reso di un prodotto acquistato online, l'agente ne comprenderebbe l'intento e interrogherebbe il datababse per verificare l'acquisto e fornire assistenza su di esso.

2.4.4 SageMaker vs Bedrock

Entrambi i servizi forniscono una piattaforma software e hardware pronta all'uso per sviluppare applicazioni, eliminando la necessità di configurare server fisici, sistemi operativi o reti da zero (compiti tipici dello *IaaS*). Tuttavia, il livello di astrazione della loro infrastruttura è nettamente differente. Infatti:

- *Amazon SageMaker*: pur essendo una piattaforma di sviluppo (*PaaS*), mantiene un forte legame con l'infrastruttura hardware tipica dello *IaaS*. Per poter programmare il codice, lo sviluppatore ha la responsabilità diretta di scegliere le istanze virtuali (CPU e GPU) necessarie per l'addestramento e il rilascio dei modelli. Esso offre un controllo totale sull'architettura computazionale, il che lo rende più vicino alla gestione infrastrutturale.
- *Personalizzazione*: pur rimanendo uno strumento per sviluppatori (*PaaS*), adotta un'architettura interamente serverless che elimina ogni visibilità sull'hardware, avvicinandosi all'esperienza d'uso di un *SaaS*. I modelli di Intelligenza Artificiale vengono consumati direttamente come funzionalità pronte all'uso tramite chiamate API. Non essendoci server da configurare o monitorare, il servizio offre un livello di astrazione così elevato da sfiorare il modello del software come servizio.

2.5 Servizi cognitivi e strumenti di Intelligenza Artificiale pronti all'uso

Sviluppare questo tipo di modelli di *Machine Learning* tramite le soluzioni *PaaS* citate può risultare comunque molto complicato e complesso per le realtà prive di competenze nel settore. Per questo motivo, piattaforme come Amazon Web Services consentono un ulteriore livello di astrazione e semplificazione, liberando le aziende dalla gestione dell'infrastruttura e dei modelli di *Deep Learning* ed estendendo il proprio catalogo con servizi cognitivi e strumenti di Intelligenza Artificiale pronti all'uso. Queste soluzioni si configurano, di fatto, come strumenti *Software as a Service (SaaS)*

2.5.1 Analisi del testo e dei sentimenti: *Amazon Comprehend*

Amazon Comprehend è uno dei servizi *SaaS* pronti all'uso che AWS mette a disposizione del cliente; è un servizio di *Natural Language Processing* il cui scopo è utilizzare il *Machine Learning* per estrarre informazioni e relazioni significative da testi anche non strutturati sviluppando analitiche basate su entità, frasi chiave, linguaggio, sentimenti e altri elementi comuni in un documento. Il sistema, essendo continuamente addestrato su un ampio corpus di testo, è progettato per due tipologie di utenti: da un lato, chi non ha competenze di *Machine Learning* permettendo di analizzare documenti, email e recensioni tramite semplici API; dall'altro, esperti del settore fornendo la possibilità di sviluppare nuovi prodotti basati sulla comprensione della struttura dei documenti. Ad esempio, con Amazon Comprehend è possibile cercare nei post dei social network le menzioni di prodotti o scansionare un intero archivio di documenti alla ricerca di frasi chiave.

In questo contesto, l'utente ha la possibilità di utilizzare modelli pre-addestrati forniti da Amazon, oppure addestrare i propri modelli personalizzati per la classificazione e il riconoscimento delle entità di testo.

Amazon Comprehend ha la possibilità di analizzare i seguenti tipi di informazioni:

- *entità*: riferimenti ai nomi di persone, luoghi e oggetti contenuti in un documento;
- *frasi chiave*: frasi fondamentali; ad esempio, in un documento riepilogativo di una partita, potrebbe riportare i nomi e i punteggi delle squadre;
- *informazioni di identificazione personale (PII)*: dati identificativi di un individuo, quali indirizzo, numero di telefono, nome e cognome;
- *lingua*: lingua prevalente in un documento;

- *sentimento*: sentimento predominante di un testo;
- *sentimento mirato*: sentimenti associati a entità specifiche;
- *sintassi*: la funzione grammaticale relativa a ogni parola del documento.

Comprehend utilizzando i modelli pre-addestrati consente l'analisi di documenti di testo UTF-8, mentre utilizzando modelli addestrati dall'utente supporta molteplici formati, tra cui immagini, PDF, file Word e documenti di testo UTF-8.

Per quel che concerne i casi d'uso generali, Amazon Comprehend è in grado di interpretare il sentimento e le tematiche delle recensioni attraverso l'analisi del testo fornendo un feedback positivo, negativo o neutrale sulla loro opinione. Inoltre, inviandogli testi ricavati dalle discussioni online, può monitorare i forum per capire gli interessi dei clienti e analizzare grandi volumi di documenti per estrarne i concetti chiave.

La Figura 2.6 rappresenta il flusso generale di Amazon Comprehend.

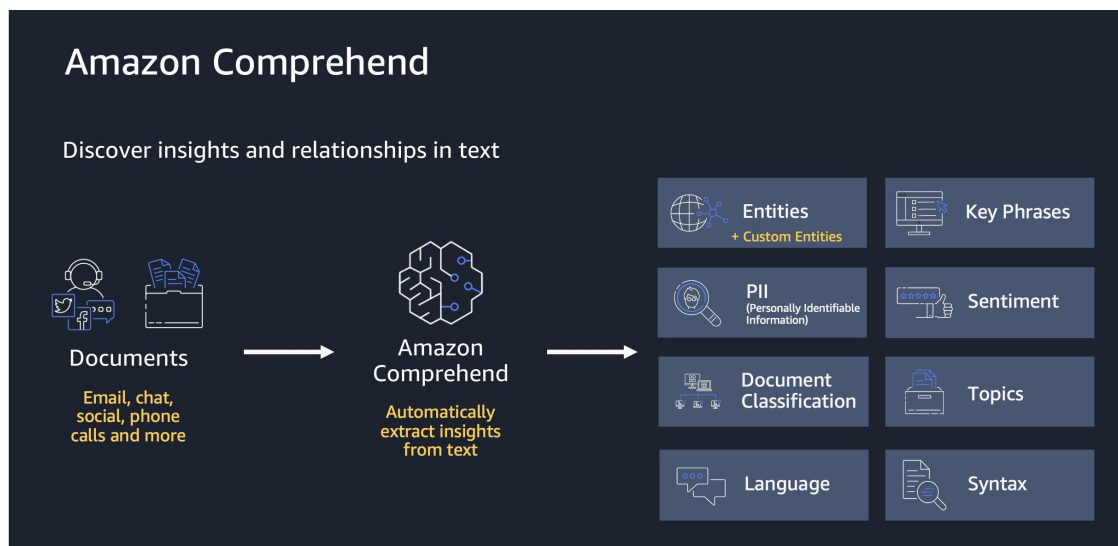


Figura 2.6: Illustrazione del flusso generale di Amazon Comprehend (fonte: aws.amazon.com)

2.5.2 Visione artificiale e analisi d'immagine: *Amazon Rekognition*

Amazon Rekognition è un ulteriore servizio *SaaS* pronto all'uso che AWS mette a disposizione del cliente.

Questo tipo di sistema analizza immagini e video con formati come PNG, JPEG e MP4. Si basa su una comprovata tecnologia di *Deep Learning*, anch'essa non richiedente competenze di *Machine Learning* per essere utilizzata; infatti, include un'API semplice in grado di analizzare rapidamente qualsiasi documento multimediale.

Il servizio si basa sulla stessa tecnologia di *Deep Learning* collaudata e altamente scalabile sviluppata dagli esperti di visione artificiale di Amazon. Tale tecnologia è in grado di analizzare miliardi di immagini e video ogni giorno, con il vantaggio di apprendere continuamente e quotidianamente tramite nuovi dati ed etichette aggiunte dagli sviluppatori di Amazon.

Allo stesso modo di Comprehend, Rekognition mette a disposizione modelli pre-addestrati offrendo la possibilità di creare classificatori personalizzati per riconoscere oggetti, scene e volti specifici rilevanti per il proprio caso d'uso.

Nell'elenco sottostante vengono riportate una serie di funzionalità chiave di Rekognition sia per l'analisi delle immagini che per l'analisi dei video:

- *analisi delle immagini:*
 - *oggetti, scene e concetti:* rileva e classifica migliaia di oggetti, contesti ambientali e categorie specifiche all'interno delle immagini;
 - *rilevamento del testo:* riconosce e trascrive testo stampato e scritto a mano in diverse lingue e diversi stili;
 - *moderazione dei contenuti:* individua e filtra immagini esplicite o inappropriate attraverso etichette granulari per la sicurezza dei contenuti;
 - *riconoscimento celebrità:* identifica decine di migliaia di figure pubbliche note, tra cui politici, atleti, attori e musicisti;
 - *analisi facciale:* rileva i volti e ne analizza attributi, come sesso, età ed emozioni, per scopi di verifica, catalogazione o sicurezza;
 - *etichette personalizzate:* permette di creare classificatori su misura per identificare oggetti specifici, come loghi, prodotti o personaggi;
 - *proprietà dell'immagine:* analizza le caratteristiche tecniche del file, come qualità visiva, colori, nitidezza e contrasto.
- *analisi dei video:*
 - *oggetti, scene e concetti:* rileva e classifica elementi, contesti ambientali e categorie specifiche all'interno dei flussi video;
 - *rilevamento del testo:* riconosce e trascrive il testo stampato e manoscritto presente nei fotogrammi in diverse lingue;
 - *percorso delle persone:* monitora gli spostamenti degli individui identificati attraverso i diversi fotogrammi del video;
 - *analisi facciale:* esegue il rilevamento e il confronto dei volti, analizzandone gli attributi sia in streaming che in video archiviati;
 - *riconoscimento di celebrità:* identifica figure pubbliche note appartenenti a diverse categorie, come politica, sport e spettacolo, all'interno dei filmati;
 - *contenuti non sicuri:* individua e segnala scene esplicite, inappropriate o violente per la moderazione automatica dei contenuti video;
 - *segmentazione video:* identifica automaticamente i segmenti tecnici rilevanti, come la presenza di fotogrammi neri o i titoli di coda;
 - *vivacità del viso:* verifica la presenza reale di un utente dal vivo durante i processi di autenticazione facciale.

Per quanto riguarda i casi d'uso di Amazon Rekognition, il servizio consente la verifica dell'identità degli utenti basandosi sui loro volti; esso permette, tra le altre cose, la *Rekognition Face Liveness*; quest'ultima è una funzionalità progettata per aiutare gli sviluppatori a scoraggiare le frodi durante il processo di verifica dell'identità; essa consente di verificare che non sia un malintenzionato a utilizzare il volto di un utente reale con l'intento di svolgere tentativi di appropriazione indebita.

Una seconda applicazione è rappresentata dalla ricerca facciale; con essa si possono ricercare immagini e video archiviati per trovare volti che corrispondano a quelli presenti in un dataset.

Ulteriori casi d'uso includono: il rilevamento di contenuti non sicuri, il rilevamento di dispositivi di protezione individuale, il riconoscimento delle celebrità e il riconoscimento del testo.

Nella Figura 2.7 viene mostrato un possibile flusso che può essere ottenuto tramite l'utilizzo di Amazon Rekognition.

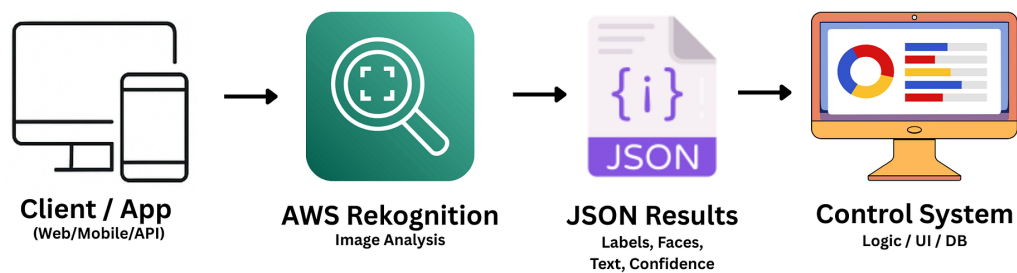


Figura 2.7: Uno dei possibili flussi che può essere ottenuto tramite l'utilizzo di Amazon Rekognition (fonte: aws.plainenglish.io)

2.5.3 Trascrizione vocale: *Amazon Transcribe*

Amazon Transcribe è un ulteriore servizio *SaaS* pronto all'uso, capace di riconoscere il segnale vocale in modo automatico (*Speech-to-Text*) e convertirlo in testo scritto.

Transcribe ha la capacità di convertire il parlato in un documento scritto utilizzando modelli di apprendimento automatico e aggiungendo dati sul contenuto trascritto, compresi i punteggi di affidabilità per ogni parola o segno di punteggiatura. È possibile utilizzare Amazon Transcribe come servizio di trascrizione autonomo o per aggiungere delle funzionalità di *Speech-to-Text* a qualsiasi applicazione.

I metodi di trascrizione che utilizza questo servizio possono essere suddivisi in due categorie principali:

- *Trascrizioni in batch*: è la modalità ideale per interviste e riunioni registrate. Tramite tale modalità è possibile caricare i file audio già registrati sulla piattaforma Amazon S3, in modo tale che, in autonomia, Transcribe possa analizzarli, leggerli e restituire un documento contenente tutto ciò che è stato oggetto di discussione.
- *Trascrizioni in streaming*: questa modalità è perfetta per trascrivere mentre un soggetto sta esponendo un argomento. Al posto di caricare un file, il flusso viene inoltrato direttamente al server di Amazon, in modo tale che il testo trascritto appaia istantaneamente sullo schermo mentre l'evento è in corso.

I tipi di file multimediali supportati da fornire in input al servizio differiscono tra le trascrizioni in batch e le trascrizioni in streaming; tra i formati consigliati per il batch sono presenti FLAC e WAV con codifica PCM a 16 bit, mentre per le trascrizioni in streaming sono presenti FLAC e audio PCM Little-endian firmato a 16 bit.

Relativamente ai casi d'uso, il servizio risulta particolarmente utile per la trascrizione automatica di riunioni e conferenze, permettendo di generare verbali testuali senza intervento manuale. Un'altra applicazione è la creazione di sottotitoli per filmati in tempo reale, il che fornisce la possibilità ai soggetti non udenti di visualizzare il contenuto in modo ottimale. Il servizio trova, poi, applicazione in call center per la trascrizione post chiamata, facendo da supporto per le analisi delle interazioni con i clienti e il monitoraggio della qualità del servizio offerto dall'azienda usufruttrice della soluzione Transcribe. Un'altra capacità è rappresentata da un'evoluzione del servizio denominato *Amazon Transcribe Toxicity Detection*, che consente di utilizzare segnali audio e testuali per identificare e classificare i contenuti inappropriati nei flussi audio, utile per il monitoraggio dei call center o piattaforme di live streaming.

Nella Figura 2.8 vengono rappresentati i due possibili flussi di *Amazon Transcribe*: trascrizione in batch e trascrizione in streaming.

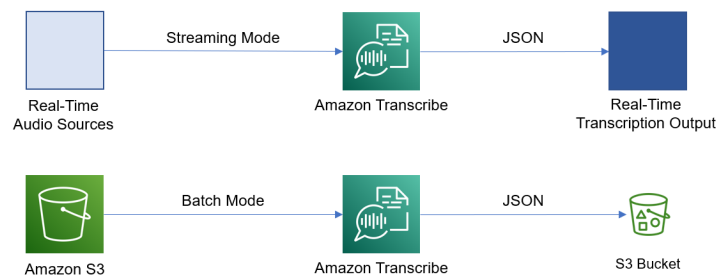


Figura 2.8: Illustrazione dei due possibili flussi di *Amazon Transcribe*: trascrizione in batch e trascrizione in streaming (fonte: aws.amazon.com)

2.6 Vantaggi strategici e limiti nell'adozione di soluzioni AWS

L'adozione di *Amazon Web Services* (AWS) consente di utilizzare una piattaforma *cloud* matura, resiliente e scalabile, ma comporta compromessi gestionali e finanziari precisi.

2.6.1 Vantaggi Strategici ed Efficientamento Architettuale

Dal punto di vista ingegneristico, i principali benefici derivanti dall'adozione dell'ecosistema AWS si riflettono sulla flessibilità e sulla resilienza del servizio. In particolare, si hanno i seguenti vantaggi:

- *Pay-per-use*: la politica di AWS consiste nel far pagare al cliente solo in base a quanto consuma, evitandogli, così, investimenti iniziali ingenti e magari, inutili.
- *Sfruttamento dell'economia di scala*: l'ampia base di clienti consente ad AWS di ammortizzare le spese di gestione, garantendo un costo del servizio notevolmente ridotto rispetto a quello che riuscirebbe a fare un singolo cliente in autonomia.
- *Capacità adattata al bisogno*: AWS permette di partire da una capacità base e di scalare verso l'alto o verso il basso secondo la necessità dell'utente, eliminando, così, le insicurezze decisionali iniziali relative a questo argomento.
- *Velocità e agilità*: AWS consente di allocare le risorse computazionali necessarie in pochi minuti rivoluzionando l'approccio tradizionale in cui le varie fasi per ottenere l'hardware necessario richiedevano complessi cicli di approvvigionamento, con tempistiche diventate ad oggi inaccettabili.
- *Gestione totalmente a carico del provider*: Amazon Web Services, facendosi carico delle configurazioni, delle manutenzioni e dei problemi infrastrutturali, consente al cliente finale di concentrarsi sul suo core business senza distrazioni.
- *Globalità delle applicazioni*: tramite AWS è possibile distribuire facilmente la propria applicazione in più regioni del mondo fornendo una latenza inferiore e un'esperienza migliore al cliente finale.

2.6.2 Limiti Tecnici, Vincoli e Fattori di Rischio

Contestualmente ai vantaggi, l'integrazione nell'ecosistema AWS introduce una complessa gestione dei costi, una curva di apprendimento ripida, vincoli architettureali, nonché operativi che richiedono un'attenta valutazione ingegneristica e finanziaria. In particolare, si hanno i seguenti vantaggi:

- *Pay-per-use*: questo modello di fatturazione presenta un'intrinseca natura ambivalente: se da un lato offre flessibilità, dall'altro, sotto un punto di vista di dimenticanze o di traffico anomalo, può esporre l'utente a costi imprevisti.
- *Costi di trasferimento dati*: spostare grandi volumi di dati esternamente all'ecosistema AWS comporta tariffe aggiuntive che possono risultare piuttosto ingenti.
- *Curva di apprendimento ripida*: Amazon Web Service offre centinaia di servizi unici; questo, in assenza di competenze verticali o certificazioni dedicate, rende difficoltoso orientarsi tra le opzioni e le relative configurazioni.
- *Gestione della sicurezza complessa*: configurare correttamente i permessi tramite l'*Identity and Access Management (IAM)* è un processo lungo e complesso, dove un errore o la mancanza di competenze può esporre pubblicamente dati sensibili.
- *Vendor Lock-in*: molti servizi sono basati su tecnologie proprietarie di Amazon, ad esempio, AWS Lambda o DynamoD; questa situazione comporta una riscrittura totale del codice nel momento in cui un'azienda decide di migrare verso un altro provider.
- *Rischio di downtime centralizzato*: sebbene AWS garantisca un'elevata resilienza, disservizi legati a intere "Region" comportano il crash contemporaneo di migliaia di applicazioni e servizi globali.
- *Controllo limitato*: l'azienda che acquista il servizio non ha nessun tipo di controllo sull'hardware fisico, dovendosi affidare totalmente alla gestione del provider.

2.6.3 Sintesi dei compromessi

La scelta di adottare AWS rappresenta il risultato di un bilanciamento strategico tra innovazione, resilienza e controllo dell'infrastruttura. La Tabella 2.1 riassume i principali compromessi ingegneristici e finanziari analizzati.

Caratteristica	Vantaggio	Svantaggio o rischio
Modello economico	<i>Pay-per-use</i> ed economia di scala: nessun investimento iniziale.	Volatilità dei costi, rischio di spese impreviste per traffico anomalo o dimenticanze.
Scalabilità e capacità	Adattamento dinamico al bisogno e allocazione delle risorse in pochi minuti.	Tariffe elevate e impreviste per spostare dati fuori da AWS.
Gestione infrastrutturale	Gestione totalmente a carico del provider (Focus sul core business).	Controllo limitato sull'hardware fisico e dipendenza totale in caso di downtime della <i>Region</i>
Ampiezza ecosistema	Velocità, agilità e accesso a centinaia di servizi innovativi pronti all'uso.	Curva di apprendimento ripida e necessità di competenze verticali per evitare errori.
Sicurezza e permessi	Strumenti di controllo granulari e avanzati (AWS IAM).	Gestione della sicurezza complessa; alto rischio di data leak per errori di configurazione.
Architettura e codice	Servizi nativi ottimizzati (es. Lambda, DynamoDB) per massima efficienza.	<i>Vendor Lock-in</i> : legame a tecnologie proprietarie e costi enormi in caso di futura migrazione.

Tabella 2.1: Riepilogo dei principali compromessi analizzati per AWS

Un case study nel contesto della Sentiment Analysis

La Sentiment Analysis è una tecnologia diventata fondamentale all'interno delle organizzazioni il cui obiettivo è ottimizzare i tempi e le risorse; tale processo viene attuato automatizzando l'estrazione delle opinioni dei clienti, delle informazioni sensibili, delle intenzioni d'acquisto e delle emozioni dominanti provenienti da testi non strutturati. Nei capitoli precedenti è stata analizzata l'infrastruttura AWS ed è stato definito il servizio Amazon Comprehend, strumento per il Natural Language Processing; nel capitolo corrente verrà dimostrato come Comprehend funziona nella pratica. A tal fine la trattazione delineerà la definizione, gli aspetti tecnici e attuali della Sentiment Analysis, presentando due scenari aziendali concreti (anonimizzazione dei dati sensibili per la conformità GDPR e analisi di recensioni di e-commerce) per illustrare come applicare Amazon Comprehend nella risoluzione di problemi reali di business.

3.1 Sentiment Analysis

Nell'era moderna, ogni giorno miliardi di persone esprimono opinioni differenti su piattaforme di e-commerce e social media, scambiando, al contempo, informazioni sensibili. Diventa, quindi, fondamentale un meccanismo capace di controllare e segnalare eventuali problemi nei prodotti, di riportare informazioni preziose sul comportamento, sulle preferenze dei clienti e, al tempo stesso, di tutelarne la privacy. In questo contesto, la Sentiment Analysis si offre come strumento risolutivo alle questioni, applicando tecniche di Natural Language Processing (NLP) per classificare automaticamente il tono, il sentimento, le frasi chiave e le informazioni di identificazione personale dei testi, garantendo la sicurezza dei dati e la conformità alle normative vigenti.

Grazie alla capacità strategica della Sentiment Analysis di estrarre contenuti dalle varie piattaforme su cui viene applicata, essa è stata capace di rendere obsoleta ogni organizzazione che non utilizzasse questo metodo per automatizzare il laborioso processo di raccolta e analisi delle informazioni.

Le applicazioni di tale tecnologia sono molteplici, dal marketing e business con le analisi dei feedback provenienti dai clienti, alle previsioni di mercato per analizzare notizie finanziarie con l'obiettivo di anticipare i trend della borsa.

Nel presente capitolo, la Sentiment Analysis costituisce la base teorica su cui costruire i case study del servizio Amazon Comprehend. Tale introduzione viene fatta per avere una panoramica più chiara del contesto, con l'obiettivo di far comprendere in modo ottimale le potenzialità e le criticità del servizio Comprehend.

3.1.1 Definizione di Sentiment Analysis

L'analisi del sentiment, o *opinion mining*, è il processo che permette di analizzare grandi quantità di dati relativi ad e-mail, tweet, risposte ai sondaggi online, chat con gli operatori del servizio clienti e recensioni, per determinare se il tono di tali messaggi risulta essere positivo, negativo o neutro.

3.1.2 Aspetti tecnici sulla Sentiment Analysis

L'analisi del sentiment utilizza tecnologie di elaborazione del linguaggio naturale e il *Machine Learning* per addestrare i software ad analizzare e, soprattutto, interpretare il testo nel modo che più si avvicina a quello dell'essere umano.

Il software addestrato utilizza tre principali approcci per raggiungere i propri obiettivi, ovvero:

- *Analisi del sentiment basata su regole*: in questo approccio, il software viene addestrato per classificare determinate parole chiave all'interno di un blocco di testo sulla base di insiemi di parole che descrivono il sentiment dell'autore. Ad esempio, all'interno di un lessico positivo è possibile trovare termini quali: "conveniente", "veloce" e "ben fatto"; mentre all'interno di un lessico negativo sono fondamentali parole come: "costoso", "lento" e "di scarsa qualità". Il software adibito all'analisi individua le parole del lessico positivo o negativo e, infine, calcola un punteggio totale del sentiment in base alle occorrenze dei termini utilizzati e al punteggio di sentiment di ciascuna categoria.
- *Analisi del sentiment basata sul Machine Learning*: il nome di questa tipologia di approccio deriva dal fatto che i modelli di *Machine Learning* utilizzati per l'analisi dei testi continuano ad "imparare" dai dati che ricevono. In questa metodologia, gli sviluppatori utilizzano algoritmi di analisi del sentiment per insegnare al software come identificare le emozioni all'interno dei testi in modo simile agli esseri umani. Alcuni degli algoritmi più comunemente utilizzati sono:
 - *regressione lineare*: è un algoritmo statistico che descrive un valore Y in base a un insieme di funzioni X .
 - *Naïve Bayes*: questo è un algoritmo che utilizza il teorema di Bayes, principio già mostrato nella Sezione 1.3.1, necessario per categorizzare le parole in un blocco di testo;
 - *macchine a vettori di supporto*: è un algoritmo di classificazione veloce ed efficiente utilizzato per risolvere problemi di classificazione a due gruppi;
 - *deep learning (DL)*: tecnica avanzata già ampiamente discussa nella Sezione ??.
- *Approccio ibrido*: questa tecnica combina le funzionalità del *Machine Learning* e si basa su regole per ottimizzare la precisione e la velocità.

In aggiunta ai vari approcci menzionati per creare strumenti di analisi del sentiment, prendono spazio all'interno del contesto anche vari tipi di analisi del sentimento a cui le organizzazioni fanno riferimento dipendentemente dai loro bisogni.

I tre tipi di analisi più popolari hanno tutti in comune il fondamento sulla capacità del software di misurare la polarità, ovvero la metrica per misurare le emozioni testuali, espressa con una valutazione numerica su una scala da 0 a 100, in cui 0 rappresenta il neutro, mentre 100 rappresenta il sentimento più estremo; i tre tipi menzionati sono:

- *Dettagliata (graduata)*: è un tipo di analisi del sentiment che raggruppa il testo in base alle emozioni individuate, valutate su una scala da 0 a 100.
- *Basata sugli aspetti (ABSA)*: tale tipologia analizza il testo a livello di singolo prodotto o servizio.
- *Rilevamento delle emozioni*: rappresenta la tipologia più complessa, poiché cerca di comprendere lo stato psicologico dell'autore del testo; invece di utilizzare la polarità, il rilevamento emotivo può identificare emozioni specifiche, come la frustrazione, l'indifferenza, l'irrequietezza e lo shock.

3.1.3 Applicazioni attuali della Sentiment Analysis

Il problema attuale delle aziende è rappresentato dall'enorme mole di dati in loro possesso; tali dati sono soggetti ad analisi per estrarre insight significativi con l'obiettivo di calibrare le decisioni aziendali. In questo contesto, l'analisi del sentiment aiuta le organizzazioni a comprendere in maniera ottimale i propri clienti e a creare un rapporto più solido con questi ultimi. Tra i principali casi d'uso moderni troviamo:

- *Migliorare il supporto clienti*: le aziende utilizzano l'analisi del sentiment per fornire ai clienti risposte più personalizzate che mostrano in modo più accurato lo stato d'animo del cliente con cui si sta interagendo. Le organizzazioni che usano questa tipologia di approccio hanno la capacità di dare priorità e di individuare i problemi in maniera più efficiente. Gli algoritmi presenti sui vari forum aiutano a classificare gli argomenti in base al livello di urgenza e hanno persino la capacità di individuare la frustrazione del cliente relativamente a un particolare prodotto o servizio.
- *Costruire una presenza del marchio più forte*: con l'analisi del sentiment i vari brand possono comprendere l'opinione che hanno i clienti su di loro; in questo contesto la Sentiment Analysis gioca un ruolo fondamentale monitorando le vendite e le recensioni degli utenti sui social media per identificare le emozioni espresse riguardo a un determinato prodotto.
- *Condurre ricerche di mercato*: tramite l'analisi del sentiment è possibile anche analizzare le mosse dei vari concorrenti, individuando ciò che sta avendo effetto sul pubblico e ciò che, invece, non sta funzionando, in modo tale che i decisori delle organizzazioni possano prendere le migliori decisioni a livello di marketing in tempo reale.

La Figura 3.1 mostra una rappresentazione delle applicazioni attuali della Sentiment Analysis.

3.2 Contesto e metodologia per attuare la Sentiment Analysis

Per l'applicazione delle tecniche di Sentiment Analysis è stato scelto Amazon Comprehend, un servizio della suite AWS già presentato nella Sezione 2.5.1, il quale utilizza principalmente l'analisi del sentimento basata sugli aspetti (ABSA) affiancandola a un'analisi graduata a 4 categorie. Al fine di valutarne l'efficacia nella risoluzione di problematiche aziendali concrete, lo strumento è stato applicato allo scenario reale delle recensioni dei prodotti dei siti di e-commerce.

L'iter metodologico adottato si articola come segue:

1. definizione dello scenario aziendale e del problema da risolvere;

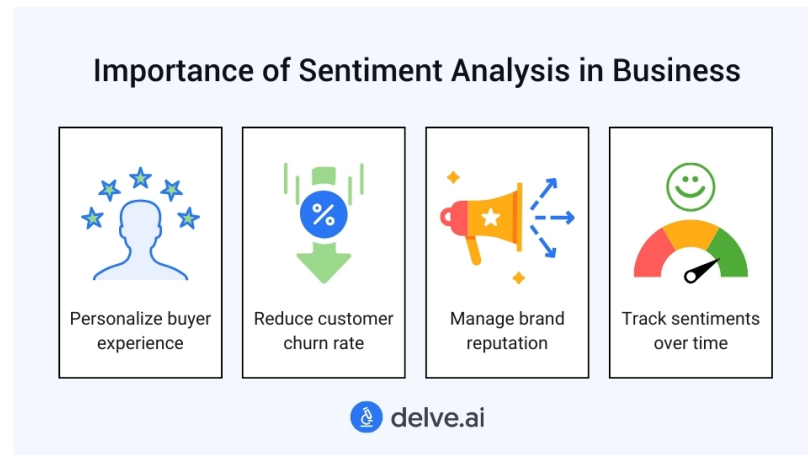


Figura 3.1: Rappresentazione delle applicazioni attuali della Sentiment Analysis (fonte: www.delve.ai)

2. preparazione, pulizia, normalizzazione e caratterizzazione del dataset;
3. estrazione e post-processing dei risultati;
4. analisi qualitativa dei risultati ottenuti;
5. identificazione dei punti di forza e delle limitazioni.

Nella sezione successiva verrà mostrato in dettaglio come verrà configurato Comprehend al fine di ottenere un'analisi del case study ottimale.

3.2.1 Implementazione con Amazon Comprehend

Per analizzare i dataset con Amazon Comprehend, verrà effettuato l'accesso alla AWS Management Console, verrà scelto il servizio "Amazon Comprehend" e verrà selezionato "Launch Amazon Comprehend" dal menù principale.

Nonostante questo servizio abbia anche la possibilità di funzionare caricando un file CSV, JSON o importandolo da Amazon S3, viene scelto di procedere con la modalità più semplice che AWS predispone, ossia selezionare le informazioni da analizzare e caricarle nell'apposita casella di testo.

È necessario, poi, specificare se l'analisi è "Built-in" o "Custom"; la prima, ovvero quella utilizzata per eseguire i case study, prevede l'utilizzo di modelli proprietari di AWS built-in, mentre la seconda fornisce la possibilità di utilizzare, per l'analisi, modelli personalizzati con endpoint creati dall'utente.

Nella Figura 3.2 viene mostrata la configurazione di Comprehend sopra spiegata.

Una volta finita la configurazione ed inserito il testo, è possibile cliccare su "Analyze" e attendere il completamento dell'elaborazione. Amazon Comprehend restituirà il risultato in due diversi formati, ossia tramite API e tramite interfaccia utente.

- Selezionando come output "Entities" si ottengono queste tre caratteristiche per ciascuna entità rilevata:
 - *Entity*: l'entità registrata.
 - *Type*: il tipo dell'entità registrata.
 - *Confidence*: il confidence score (da 0 a 1) per ogni categoria.

Input data
Supported languages [↗](#)

Analysis type [Info](#)

☒ **Built-in**
View real-time insights based on AWS built-in models.

☐ **Custom**
View real-time insights based on custom models from an endpoint you've created.

Input text

Hello Zhang Wei, I am John. Your AnyCompany Financial Services, LLC credit card account 1111-0000-1111-0008 has a minimum payment of \$24.53 that is due by July 31st. Based on your autopay settings, we will withdraw your payment on the due date from your bank account number XXXXXX1111 with the routing number XXXXX0000.
Customer feedback for Sunshine Spa, 123 Main St, Anywhere. Send comments to Alice at sunspa@mail.com.
I enjoyed visiting the spa. It was very comfortable but it was also very expensive. The amenities were ok but the service made the spa a great experience.

578 of 5000 characters used.

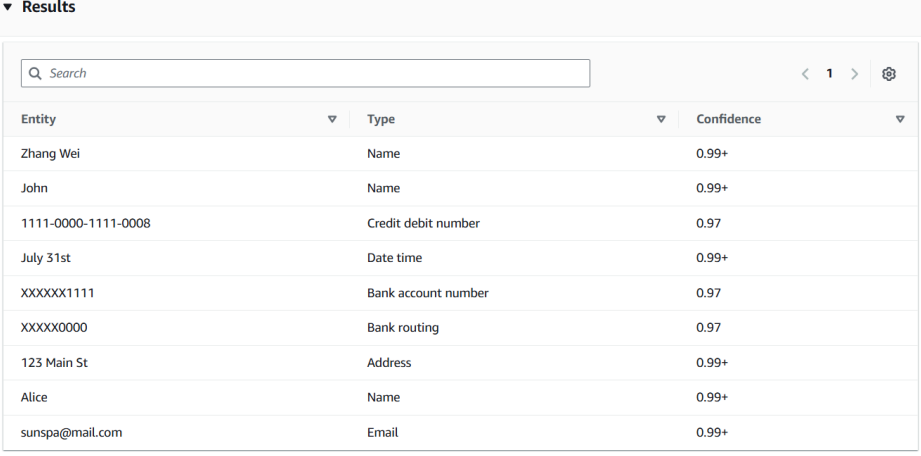
[Clear text](#) [Analyze](#)

Figura 3.2: Rappresentazione della configurazione di Comprehend

- Selezionando come output "Key phrases" si ottengono queste due caratteristiche per ogni frase chiave rilevata:
 - *Key phrases*: la frase chiave registrata.
 - *Confidence*: il confidence score (da 0 a 1) per ogni frase chiave.
- Selezionando come output "Language" si ottiene come output il linguaggio della frase inserita.
- Selezionando come output "PII" è possibile ottenere informazioni relative all'identificazione (non disponibile in Italiano); in particolare vengono fornite le seguenti caratteristiche:
 - *Entity*: l'entità chiave registrata.
 - *Type*: il tipo di informazione identificativa registrata, ad esempio, indirizzo di casa.
 - *Confidence*: il confidence score (da 0 a 1) per l'informazione identificativa.
- Selezionando come output "Sentiment" si ottengono queste due caratteristiche:
 - *Sentiment*: la classificazione in POSITIVE, NEGATIVE, NEUTRAL, MIXED.
 - *SentimentScore*: il confidence score (da 0 a 1) per ogni categoria.
- Selezionando come output "Targeted sentiment" si ottengono queste caratteristiche specifiche per ogni entità rilevata:
 - *Entity*: la specifica entità a cui è rivolto il sentiment.
 - *Type*: il tipo di entità identificata.
 - *Sentiment*: il sentiment mirato espresso verso quella specifica entità (POSITIVE, NEGATIVE, NEUTRAL, MIXED).
 - *Confidence*: il confidence score (da 0 a 1) sia per il riconoscimento dell'entità che per il relativo sentiment assegnato.
- Selezionando come output "Syntax" si ottengono queste caratteristiche per ciascuna parola del testo:
 - *Word*: la singola parola o elemento di punteggiatura analizzato.

- *PartOfSpeech*: l'analisi grammaticale dell'elemento, identificata tramite un tag, ad esempio, *NOUN*, *VERB*, *ADJ*.
- *Confidence*: il confidence score (da 0 a 1) per la correttezza dell'assegnazione grammaticale.

Con l'obiettivo di far focalizzare il lettore su quello che verrà spiegato all'interno del capitolo, le Figure 3.3 e 3.4, rappresentano, rispettivamente, i risultati via interfaccia utente e via API che si ottengono selezionando "PII" all'interno di Comprehend. Invece, mentre nelle Figure 3.5 e 3.6 si presentano, rispettivamente, i risultati via interfaccia utente e via API che si ricevono selezionando "Sentiment". Entrambe le casistiche analizzate, "Sentiment" e "PII", sono state mostrate utilizzando l'input di default fornito da Comprehend: *"Hello Zhang Wei, I am John. Your AnyCompany Financial Services, LLC credit card account 1111-0000-1111-0008 has a minimum payment of \$24.53 that is due by July 31st. Based on your autopay settings, we will withdraw your payment on the due date from your bank account number XXXXXX1111 with the routing number XXXXX0000. Customer feedback for Sunshine Spa, 123 Main St, Anywhere. Send comments to Alice at sunspa@mail.com. I enjoyed visiting the spa. It was very comfortable but it was also very expensive. The amenities were ok but the service made the spa a great experience"*.



▼ Results		
Search		
Entity ▼	Type ▼	Confidence ▼
Zhang Wei	Name	0.99+
John	Name	0.99+
1111-0000-1111-0008	Credit debit number	0.97
July 31st	Date time	0.99+
XXXXXX1111	Bank account number	0.97
XXXXX0000	Bank routing	0.97
123 Main St	Address	0.99+
Alice	Name	0.99+
sunspa@mail.com	Email	0.99+

Figura 3.3: Rappresentazione dei risultati di Comprehend via interfaccia utente selezionando "PII"

3.3 Case study: analisi di recensioni di prodotti

Il case study si concentra sul settore dell'e-commerce, un ambito in cui la comprensione del livello di soddisfazione del cliente è fondamentale per una crescita aziendale; i feedback dei clienti rappresentano, infatti, un enorme vantaggio competitivo chiave per ottimizzare la gestione del catalogo. Applicare la Sentiment Analysis in questo contesto significa trovare una soluzione al monitoraggio manuale delle recensioni scritte dagli acquirenti, riducendo i tempi di lettura da parte di dipendenti e, allo stesso tempo, permettendo l'identificazione tempestiva dei prodotti difettosi o di picchi di insoddisfazione.

L'esperimento mira a testare la capacità di Amazon Comprehend di interpretare la polarità di testi commerciali in lingua italiana e di estrarre le parole chiave più significative. Inoltre, il lavoro esplora l'efficacia della piattaforma nella protezione della privacy degli utenti tramite il rilevamento dei dati sensibili. Di seguito vengono analizzate nel dettaglio le fasi di selezione e preparazione del dataset utilizzato, i parametri metrici emersi dall'elaborazione e una successiva valutazione qualitativa degli output restituiti dalla piattaforma cloud.

API response

```

1  {
2    "Entities": [
3      {
4        "Score": 0.9997738003730774,
5        "Type": "NAME",
6        "BeginOffset": 6,
7        "EndOffset": 15
8      },
9      {
10       "Score": 0.9998970031738281,
11       "Type": "NAME",
12       "BeginOffset": 22,
13       "EndOffset": 26
14     },
15     {
16       "Score": 0.9723346829414368,
17       "Type": "CREDIT_DEBIT_NUMBER",
18       "BeginOffset": 88,
19       "EndOffset": 107
20     },
21     {
22       "Score": 0.9956514835357666,
23       "Type": "DATE_TIME",
24       "BeginOffset": 155,
25       "EndOffset": 164
26     }
27   ]
28 }

```

Figura 3.4: Rappresentazione dei risultati di Comprehend via API selezionando "PII"

▼ Results

Sentiment

Neutral
0.23 confidence
Positive
0.62 confidence
Negative
0.00 confidence
Mixed
0.13 confidence

Figura 3.5: Rappresentazione dei risultati di Comprehend via interfaccia utente selezionando "Sentiment"

3.3.1 Dataset e preparazione

Per questo esperimento è stato selezionato un dataset composto da recensioni di prodotti commerciali in lingua italiana estratte dal sito web *Amazon.it*.

Ogni elemento del dataset è caratterizzato da due elementi principali: un testo scritto in linguaggio naturale dall'utente e una valutazione qualitativa espressa in stelle (da 1 a 5). Il voto in stelle è stato volutamente inserito all'interno dell'insieme di dati in quanto funge da oggettiva verità di riferimento e permette, tra le altre cose, la distinzione analitica di feedback "ironici"; ad esempio, alcuni utenti potrebbero scrivere:

- *"L'utilizzo di questa lavatrice è perfetto per socializzare: grazie ai continui guasti, ho stretto amicizia con tutti i tecnici della provincia"* → recensione a 1 stella → Sentiment atteso: *NEGATIVE* → Sentiment riscontrato: *POSITIVE*.

API response

```

1  {
2    "Sentiment": {
3      "Sentiment":
4        "POSITIVE",
5        "SentimentScore":
6          {
7            "Positive": 0
8              .625441074371
9              3379,
10             "Negative": 0
11               .001183692016
12               6388154,
13             "Neutral": 0
14               .234653919935
15               22644,
16             "Mixed": 0
17               .138721317052
18               8412
19           }
20        }
21    }
22  }
```

Figura 3.6: Rappresentazione dei risultati di Comprehend via interfaccia API selezionando "Sentiment"

3.3.2 Data Cleaning e Pre-processing

La fase di preparazione ha previsto una pulizia preliminare per rimuovere eventuali residui o stringhe vuote. Non è stata, invece, effettuata la rimozione della punteggiatura o delle maiuscole, poiché i modelli *built-in* di Amazon Comprehend sono progettati per analizzare il testo nel suo contesto grammaticale originario. Ai fini del test, è stato estratto un campione rappresentativo di recensioni bilanciato tra le diverse valutazioni in stelle e sono stati evidenziati i casi in cui il servizio valutava i testi in maniera differente dalle aspettative.

3.3.3 Il problema della Privacy: Integrazione della PII Detection

Nel contesto aziendale moderno, i feedback degli utenti non possono prescindere dal rispetto delle normative sulla protezione dei dati personali. Sebbene le recensioni pubbliche di e-commerce siano teoricamente prive di dati sensibili, poiché analizzate da strumenti come Comprehend prima di essere pubblicate, nella pratica aziendale gli utenti inseriscono frequentemente nei testi liberi informazioni di identificazione personale (PII), come il proprio nome di battesimo, numeri di telefono per sollecitare l'assistenza, o indirizzi di spedizione per lamentarsi di un ritardo.

Consentire la pubblicazione di questi testi grezzi nelle piattaforme senza un controllo preventivo espone l'azienda a gravi violazioni della privacy. In realtà, all'interno delle piattaforme di e-commerce pubbliche come *Amazon.it*, i testi delle recensioni subiscono già una moderazione e una normalizzazione preventiva prima della condivisione.

Tuttavia, all'interno della preparazione di questo case study, si è scelto di integrare una sezione propedeutica di *Text Sanitization* a scopo puramente illustrativo. L'obiettivo è mostrare come le realtà moderne operano quando si trovano ad affrontare questo tipo di problemi in contesti come quello delle recensioni. Per questo motivo, prima di sottoporre il testo all'algoritmo di Sentiment Analysis, i dati vengono processati tramite la *PII Detection* di Amazon Comprehend.

Proprio a causa della pulizia preventiva già effettuata a monte dalle piattaforme di e-commerce, i testi reali estratti risultano naturalmente privi di criticità. Di conseguenza, qualora si rendesse necessario evidenziare le capacità di rilevamento del servizio cloud, alcune recensioni del campione verranno integrate inserendo intenzionalmente informazioni personali e finanziarie inventate. Questo processo di arricchimento artificiale permette di

simulare un flusso di dati grezzi non filtrati, mostrando concretamente il meccanismo di identificazione delle entità sensibili operato dalla piattaforma.

Emerge, però, un ulteriore problema: la funzionalità di *PII Detection* non è supportata nativamente per la lingua italiana. Al fine di superare questo vincolo tecnologico si è optato per un processo sequenziale basato su una doppia traduzione. La procedura si articola in quattro fasi distinte:

1. Le recensioni italiane originali vengono tradotte in lingua inglese tramite il servizio *Amazon Translate*, un servizio *SaaS* di AWS che permette di eseguire traduzioni nella lingua desiderata partendo da un testo.
2. Il testo in lingua inglese viene sottoposto manualmente alla *PII Detection* di Amazon Comprehend, la quale individua i dati sensibili.
3. Il testo inglese viene manualmente normalizzato dai dati sensibili e tradotto nuovamente in lingua italiana da *Amazon Translate*.
4. Il testo italiano risultante, privato degli elementi identificativi originari, viene infine sottoposto al modulo di *Sentiment Analysis* di Amazon Comprehend.

Per comodità e per testare l'efficacia del sistema, la sostituzione delle informazioni personali è stata simulata manualmente. Tuttavia, integrando le API di Amazon Comprehend all'interno di un software, questo intero procedimento di mascheramento viene eseguito in modo completamente automatico.

Al fine di fornire maggiore chiarezza, nella Figura 3.7 si mostrano i vari passi del processo sopra descritto.

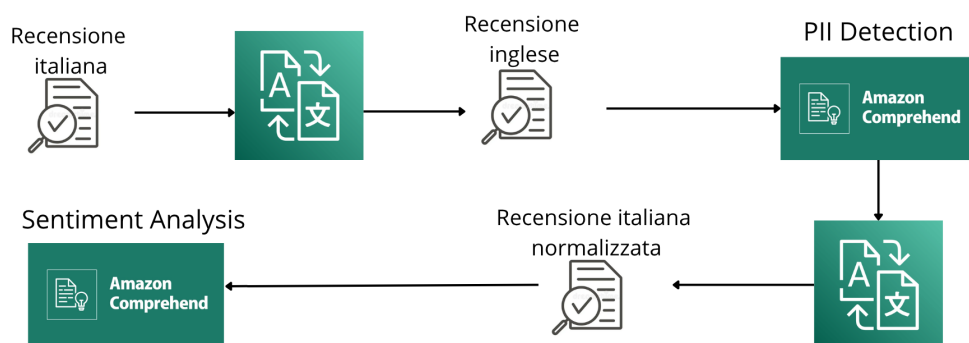


Figura 3.7: Illustrazione dei passaggi eseguiti per effettuare la Sentiment Analysis

3.3.4 Stress Test di Amazon Comprehend sulle entità PII

Il dataset di recensioni reali estratto da *Amazon.it* è stato inizialmente sottoposto all'intero processo di doppia traduzione e mascheramento delle PII. In questo scenario standard, il sistema ha dimostrato di funzionare correttamente sui dati testuali comuni, riuscendo a

identificare entità lineari, (ad esempio, i nomi propri tradizionali degli acquirenti o amici di questi ultimi presenti nel testo), consentendo, poi, la successiva analisi del sentiment sul testo italiano normalizzato. Nella Tabella 3.1 si riporta un esempio di PII rilevati all'interno di una recensione.

Fasi	Stato del testo
Input Italiano originale	Grazie a Marco che mi ha aiutato ad installare la Fire sul mio TV non smart. Mi ha dedicato più di un'ora e mezza ma alla fine è stato in grado di installare la fire sul mio Tv. Grazie Marco !
Traduzione in Inglese	Thanks to Marco who helped me install the Fire on my non-smart TV. He dedicated more than an hour and a half to me but in the end he was able to install the fire on my TV. Thanks Marco!
Rilevamento PII di Comprehend	Marco - Name 0.99+, Marco - Name 0.99+
Testo italiano normalizzato delle informazioni sensibili	Grazie a [NAME] che mi ha aiutato ad installare la Fire sul mio TV non smart. Mi ha dedicato più di un'ora e mezza ma alla fine è stato in grado di installare la fire sul mio Tv. Grazie [NAME] !

Tabella 3.1: Esempio di PII rilevati all'interno di una recensione reale

Come già accennato, al fine di verificare il corretto funzionamento del *Personally identifiable information (PII) analysis mode*, sono stati condotti degli stress test inserendo nel dataset stringhe ideate per mettere in difficoltà il sistema. Di seguito vengono riportati i tre principali scenari di fallimento riscontrati durante le simulazioni.

- *Stress test 1, indirizzi stradali metaforici:* nella Tabella 3.2 si evidenzia il limite del modello davanti a un indirizzo composto da un nome proprio monumentale e da una via espressa in modo discorsivo. L'algoritmo dimostra una dipendenza eccessiva dalle parole chiave strutturali, ad esempio "Lane", con un 41% di confidenza, ignorando la prima parte dell'informazione poiché associata statisticamente a un'entità di tipo istituzionale.

Descrizione	
Input	"I am currently staying at The White House down by Penny Lane and i have not received the package."
Rilevamento PII di Comprehend	Lane - Address 0.41

Tabella 3.2: Stress test su indirizzi stradali metaforici

- *Stress test 2, confusione tra una serie di codici numerici:* come mostrato nella Tabella 3.3, il modello evidenzia i suoi limiti quando si trova davanti a due codici diversi con la stessa struttura. Essendo stato addestrato su codici e standard americani, Comprehend non riesce a distinguerli correttamente e isola solo parzialmente la prima stringa con un 78% di confidenza.

Descrizione	
Input	"My personal customer ID is 987-222-123 and the broken product serial number is 123-456-789."
Rilevamento PII di Comprehend	456 - Ca social insurance number 0.78

Tabella 3.3: Stress test serie di codici numerici

- *Stress test 3, recensioni con dati reclamo*: nel caso visibile in Tabella 3.4, Comprehend ha successo al 100%: riesce a estrarre tutte le PII sensibili nonostante il testo sia scritto con un tono colloquiale, arrabbiato e privo di una formattazione rigida.

Descrizione	
Input	<i>"Terrible experience with this laptop! It arrived broken. I tried calling customer care from my number 555-0143 but no one answered. The courier John Doe left the package outside my house at 42 Maple Street. I want a full refund immediately on my visa card 4321-5678-9012-3456. My order email is customer.help@email.com."</i>
Rilevamento PII di Comprehend	555-0143 - Phone 0.96, John Doe - Name 0.96, 42 Maple Street - Address 0.99+, 4321-5678-9012-3456 - Credit debit number 0.99+, customer.help@email.com - Email 0.99+

Tabella 3.4: Stress test recensioni con dati reclamo

- *Stress test 4, difficoltà nel riconoscimento di codici burocratici diversi da quelli americani*: Nel caso visibile in Tabella 3.5, Amazon Comprehend, pur individuando un pattern particolare, non riconosce il formato del codice fiscale italiano e classifica l'entità come "Other". Tale limite è riconducibile alla dipendenza del modello dalla lingua inglese.

Descrizione	
Input	<i>"Please update my profile with the official tax code RSSMRA80A01F205X for the billing section."</i>
Rilevamento PII di Comprehend	RSSMRA80A01F205X - Other 0.95

Tabella 3.5: Stress test nel codice fiscale italiano

3.3.5 Metriche di accuratezza del mascheramento PII

Durante le prove di *Text Sanitization*, si è proceduto a un mascheramento manuale delle informazioni reputate "sensibili" da Comprehend. Tuttavia, se si volesse valutare e misurare la precisione del modello nel contesto della protezione dei dati, sarebbe necessario fare riferimento a metriche statistiche standardizzate prese in prestito dall'ambito del *Machine Learning*. Tali metriche sono:

- *Precisione (Precision)*: misura la percentuale di elementi mascherati che erano effettivamente PII. Una precisione elevata riduce i falsi positivi, ossia evita di mascherare parole comuni o dati non sensibili.
- *Richiamo (Recall)*: indica la percentuale di PII reali che il sistema è riuscito a intercettare e mascherare rispetto al totale presente. Un richiamo elevato riduce i falsi negativi, ossia garantisce che nessuna PII sfugga al mascheramento.
- *Punteggio F1 (F1-Score)*: rappresenta la metrica di riferimento globale ed equilibra la precisione e il richiamo. Varia da 0 a 1 (o da 0 a 100), dove 1 indica la massima accuratezza nel mascheramento.

3.3.6 Risultati e metriche della Sentiment Analysis

Conclusa la valutazione del modulo di protezione dei dati, questa sezione si concentra sull'analisi quantitativa della classificazione del sentimento operata da Amazon Comprehend sulle recensioni in lingua italiana normalizzate. L'analisi della polarità emotiva è stata eseguita, come già anticipato nella Sezione 3.3.1, misurando la corrispondenza tra il sentiment estratto dal modello (*POSITIVE*, *NEGATIVE*, *NEUTRAL*, *MIXED*) e la valutazione oggettiva in stelle lasciata originariamente dagli utenti sulla piattaforma di e-commerce. La Tabella 3.6 mostra i risultati dell'analisi del servizio di Sentiment Analysis di Amazon Comprehend.

Tabella 3.6: Caso di studio: Analisi del feedback dei clienti su alcuni prodotti proposti in *Amazon.it* tramite Amazon Comprehend

N°	Testo	Frase chiave	Entità	stelle	Sentiment atteso	Sentiment riscontrato
1	La chiavetta e' facile da configurare e funziona molto bene, anche il suo telecomando non perde mai un colpo	La chiavetta 0.99+, anche il suo telecomando 0.96, un colpo 0.99+	No entities to display	5	Positivo	Positive (0.99)
2	Ottimo prodotto, ha trasformato la mia TV (in un quarto d'ora) da pezzo da discarica a dispositivo ancora perfettamente utilizzabile!	Ottimo prodotto 0.99+, la mia TV 0.99+, un quarto d'ora 0.86, pezzo da discarica 0.98, dispositivo ancora perfettamente utilizzabile 0.86	un quarto d'ora - Quantity (0.92)	5	Positivo	Positive (0.99)
3	Grazie a [Name] che mi ha aiutato ad installare la Fire sul mio TV non smart. Mi ha dedicato più di un'ora e mezza ma alla fine è stato in grado di installare la fire sul mio Tv. Grazie [Name] !	che 0.99+, mi 0.99+, la Fire 0.99+, TV non smart 0.98, Mi 0.99+, un'ora e mezza 0.76, fine 0.99+, grado 0.99+, la fire 0.99+	Fire - Commercial item (0.96), più di un'ora e mezza - Quantity (0.99+)	5	Positivo	Positive (0.99)
4	La mia tv già smart, per un qualsiasi motivo, ha cessato di funzionare. Non prendeva più il Wi-Fi e non funzionava neanche con il cavo di rete. Al posto di aggiustare la tv ho deciso di comprare questa Fire e devo dire che ho risolto il problema.	La mia tv già smart 0.99+, un qualsiasi motivo 0.99+, il Wi-Fi 0.99+, neanche con il cavo di rete 0.98, posto 0.99+, la tv 0.99+, questa Fire 0.99+, il problema 0.99+	Wi-Fi - Other (0.88), Fire - Commercial item (0.99+)	5	Positivo	Negative (0.99)
5	Collegata molto facilmente alla mia tv. ottima l'idea di inserire una prolunga per raggiungere le porte usb .Dopo averla aggiornata ho recuperato i programmi che la mia tv non riusciva a vedere alla perfezione,davvero ottima !	tv 0.99+, l'idea 0.82, una prolunga 0.99+, le porte usb 0.97, i programmi 0.99+, che 0.99+, la mia tv 0.99+, perfezione 0.98	usb - Other (0.77)	5	Positivo	Positive (0.99)
6	La confezione è arrivata x due volte con il packaging rotto ma la penna sembra vada bene.	La confezione 0.99+, due volte 0.99+, il packaging rotto 0.99+, la penna 0.99+	due volte - Quantity (0.98)	3	Mixed	Mixed (0.99)
7	Prodotto validissimo, consigliato. Spedizione pessima, ritardi con la stessa azienda di spedizioni, pessimi in tutto.	Prodotto validissimo, consigliato 0.74, Spedizione pessima 0.95, ritardi con la stessa azienda di spedizioni 0.99+, pessimi in tutto 0.91	No entities to display	3	Negative	Mixed (0.95), Negative (0.02), Positive (0.01)

Continua nella pagina successiva...

Tabella 3.6 – continua dalla pagina precedente

N°	Testo	Frase chiave	Entità	stelle	Sentiment atteso	Sentiment riscontrato
8	Sicuramente meglio dei box cinesi che spesso sono anche più costosi. Fa il suo lavoro e dalla sua parte c'è la praticità ed il poco ingombro. Peccato che sia macchinosa quando passi da un app all'altra. Alle volte si inchioda proprio e devo riavviare ma tutto sommato per il lavoro che deve svolgere va benino	box cinesi 0.99+, che 0.80, il suo lavoro 0.99+, parte 0.99+, c'è 0.99+, la praticità 0.99+, il poco ingombro 0.99+, Peccato 0.95, un app all' 0.88, altra 0.63	cinesi - Other (0.94)	3	Mixed	Mixed (0.98), Positive (0.01)
9	Non ho capito perché non funziona il riconoscimento vocale	il riconoscimento vocale 0.99+	No entities to display	2	Negative	Negative
10	Pratica, ma a mio avviso è utile a svecchiare vecchie tv. Per la mia è stato un downgrade e quindi l'ho restituita.	Pratica 0.99+, avviso 0.99+, vecchie tv 0.99+, la mia 0.99+, un downgrade 0.87, l' 0.99+	No entities to display	2	Negative	Mixed (0.93), Negative (0.07)
11	Comprata qualche mese fa all'inizio tutto bene da qualche settimana continua a riavviarsi da sola prodotto poco valido non consiglio	qualche mese fa all' 0.93, inizio tutto bene 0.77, qualche settimana 0.99+, prodotto poco valido 0.99+	qualche mese fa - Date (0.99+), inizio - Date (0.75), qualche - Quantity (0.47), settimana - Date (0.52)	1	Negative	Negative (0.99)
12	Telecomando non funzionante! Non emette alcun impulso	Telecomando 0.99+, alcun impulso 0.99+	alcun impulso - Quantity (0.83)	1	Negative	Negative (0.99)
13	Il prodotto si blocca frequentemente e, nonostante dichiaro di avere 8 GB di memoria, ne ha solo 5 GB. La scheda Wi-Fi integrata è pessima.	Il prodotto 0.99+, si 0.99+, 8 GB di memoria 0.99+, ne 0.99+, solo 5 GB 0.96, La scheda Wi-Fi integrata 0.99+	8 GB - Quantity (0.99+), 5 GB - Quantity (0.99+)	1	Negative	Negative (0.99)
14	I primi due mesi tutto bene, poi è successo che la mia start tv non la legge più, ho provato su altri dispositivi ma niente.	I primi due mesi 0.99+, tutto 0.58, la mia start tv 0.98, non 0.96, la legge 0.93, altri dispositivi 0.99+, niente 0.99+	I primi due - Date (0.55), mesi - Quantity (0.57)	1	Mixed	Negative (0.68), Mixed (0.30)
15	La spedizione è stata così veloce che il pacco è arrivato dopo un mese, non aggiungo altro	La spedizione 0.99+, il pacco 0.99+, un mese 0.99+, altro 0.97	dopo un - Date (0.62) mese - Quantity (0.73)	1	Negative	Positive (0.99)

I testi selezionati sono stati elaborati registrando per ciascuno, oltre a "testo", "frasi chiave", "entità" e "sentiment atteso", il sentiment predominante e il relativo *SentimentScore* (il livello di confidenza restituito dall'algoritmo, indicato come "Sentimento riscontrato" in tabella). I risultati complessivi mostrati nella Tabella 3.6 mostrano una forte correlazione tra il voto in stelle e la classificazione di Comprehend.

Nella maggior parte delle recensioni fortemente polarizzate (1 stella o 5 stelle), lo strumento ha dimostrato un'elevata accuratezza, assegnando un *SentimentScore* spesso superiore allo 0.95 alla categoria corretta. Al contrario, la classe delle 3 stelle, ad esempio relativa alle recensioni 7 e 10, ha evidenziato una dispersione leggermente maggiore, con oscillazioni tra le etichette *NEUTRAL* e *MIXED*. Questo fenomeno è piuttosto comprensibile, poiché i testi associati a voti intermedi contengono spesso elementi sia di apprezzamento che di critica all'interno dello stesso periodo. All'interno di questo contesto è interessante notare come nei testi evidenziati, numero 4 e 15, Comprehend abbia commesso degli errori piuttosto notevoli;

tali inesattezze verranno commentate nella sezione successiva.

3.3.7 Analisi qualitativa

L'analisi qualitativa delle recensioni ha permesso di evidenziare come Amazon Comprehend interpreti la sintassi commerciale standard. Nei casi di successo, l'algoritmo identifica correttamente termini chiave fortemente evocativi, ad esempio, "ottimo", "pessimo" e "facile" associandoli alla classe corretta.

Tuttavia, emergono errori sistematici in presenza di costrutti linguistici più complessi. Un esempio tipico è rappresentato dalla presenza di una negazione che ribalta il senso di un aggettivo positivo, come ad esempio nella recensione numero 4 che presenta più negazioni relative al contesto del cliente, ma non alla recensione o al prodotto stesso che è al contrario valutato positivamente: *"La mia tv ... ha cessato di funzionare. Non prendeva più il WiFi e non funzionava ... e devo dire che ho risolto il problema"*. In questi scenari, Comprehend tende talvolta a farsi influenzare dalla presenza delle parole o frasi come: "cessato di funzionare" o "non funzionava" abbassando la confidenza del sentiment positivo complessivo. Un altro limite testato riguarda le recensioni ironiche o sarcastiche, ad esempio la recensione numero 15, *"Spedizione così veloce che il pacco è arrivato dopo un mese"*, dove lo strumento, mancando di un contesto semantico, si focalizza sulla parola "veloce" classificando erroneamente il testo come *POSITIVE*.

3.4 Punti di forza e limitazioni riscontrati

A conclusione dell'analisi condotta sia sulle funzionalità di *PII Detection* effettuata con recensioni reali e con test manuali mirati, sia sulle funzionalità di *Sentiment Analysis* eseguita sul dataset di recensioni di e-commerce, è possibile trarre un bilancio dei vantaggi e delle limitazioni riscontrate nell'utilizzo della piattaforma cloud di Amazon.

3.4.1 Punti di forza riscontrati

Il servizio ha dimostrato caratteristiche di alta affidabilità, riassumibili nei seguenti punti:

- *Interfaccia cloud e immediatezza d'uso*: la console web di Amazon Comprehend si offre come uno strumento estremamente intuitivo. Permette di testare complessi modelli di *Deep Learning* in tempo reale senza la necessità di scrivere codice o configurare infrastrutture hardware dedicate.
- *Ottimo rilevamento delle informazioni sensibili (Inglese)*: nei test condotti in lingua inglese, il modello *built-in* ha dimostrato un'eccellente accuratezza sintattica. Come evidenziato nello stress test 3, il sistema eccelle nell'isolare contemporaneamente entità sensibili come nomi, indirizzi e elementi finanziari, anche all'interno di testi colloquiali e non strutturati, come le recensioni di reclamo.¹
- *Comprensione della polarità standard*: per quanto riguarda la *Sentiment Analysis*, il servizio si è dimostrato robusto nell'analizzare correttamente i testi lineari in lingua italiana, fornendo percentuali di confidenza stabili che rispecchiano la reale valutazione in stelle lasciata dagli utenti nella maggior parte dei casi non ambigui.

¹Questa elevata affidabilità rispecchia casi d'uso moderni come Airbnb. Tali aziende implementano questo tipo di sistemi all'interno delle chat di messaggistica per impedire che l'ospite e l'host scambino indirettamente recapiti personali (social, e-mail o numeri di telefono) prima della conferma della prenotazione. L'identificazione e il blocco tempestivo di queste PII, obbliga gli utenti a concludere la transazione all'interno della piattaforma tutelando così le commissioni di ricavo di quest'ultima.

- *Continuo apprendimento del modello Cloud*: un elemento di forte interesse emerso durante le sessioni sperimentali risiede nella natura dinamica e in continua evoluzione della piattaforma SaaS di AWS. Test ripetuti a distanza di sole ventiquattro ore hanno dimostrato variazioni incrementali nell'accuratezza del modello, il quale ha iniziato a rilevare correttamente entità PII che il giorno precedente venivano ignorate. Questo comportamento dimostra l'efficacia delle logiche di *Continuous Learning* e dei processi di addestramento automatico implementati da Amazon, garantendo un servizio costantemente aggiornato e capace di affinare le proprie capacità.

3.4.2 Limitazioni riscontrate

Parallelamente, l'indagine ha fatto emergere limitazioni strutturali che richiedono attenzione in un'ottica di implementazione aziendale:

- *Dipendenza del servizio dalla lingua inglese ed asimmetria geografica*: il limite più rilevante per il mercato europeo risiede nella forte dipendenza della PII Detection dalla lingua inglese. L'assenza di un supporto nativo ottimizzato per la lingua italiana impedisce un'analisi diretta. Inoltre, come emerso dallo stress test 4, il sistema mostra difficoltà davanti a identificativi burocratici locali come il Codice Fiscale italiano: pur intercettando la stringa alfanumerica a tutela della privacy, l'algoritmo non è in grado di decodificarne la semantica specifica, classificandola come categoria generica *OTHER*.
- *Dipendenza regole rigide*: lo stress test ha evidenziato come l'algoritmo si affidi fortemente a parole chiave strutturali (*trigger words*). Come dimostrato all'interno dello stress test 1, la presenza del termine "*Lane*" attiva correttamente il rilevamento dell'indirizzo, mentre contesti residenziali privi di suffissi standard, ad esempio "*The White House*", confondono il modello causando falsi negativi. Inoltre, la medesima rigidità si riscontra nella gestione delle serie di codici numerici isolati: il sistema applica soglie di confidenza strettamente legate alla presenza di token espliciti, ad esempio "*Card*" o "*PIN*". Inoltre, in presenza di contesti ambigui, codici di prodotto seriali non sensibili rischiano di essere confusi con identificativi personali o finanziari, dimostrando che l'algoritmo non riesce ad attuare una reale comprensione logica dello stesso.
- *Vulnerabilità al sarcasmo e all'ironia*: sia per rilevamento delle informazioni che per il rilevamento del sentimento, le frasi ironiche rappresentano un punto di difficoltà per i modelli pre-addestrati. L'incapacità dell'algoritmo di decodificare il sarcasmo, come mostrato nella recensione numero 15, provoca errori di classificazione della polarità emotiva, invertendo l'output atteso rispetto alla reale intenzione dell'utente.
- *Influenza da termini prettamente negativi*: i test hanno evidenziato come la presenza di termini fortemente negativi tenda a influenzare eccessivamente l'output finale del servizio cloud. L'algoritmo mostra una sorta di sbilanciamento statistico: la presenza di queste espressioni negative all'interno del testo, anche se riferite a contesti passati o a problemi successivamente risolti dall'utente, genera un "effetto rumore" che abbassa la percentuale di confidenza del sentimento positivo complessivo.

È, quindi, possibile concludere che Amazon Comprehend rappresenta uno degli ottimi servizi della suite AWS. Esso permette di semplificare la vita delle organizzazioni e costituisce un vantaggio statistico nei processi più lunghi e complessi, come mostrato all'interno del case study; necessita però di una minima supervisione umana in grado di monitorare le situazioni più ambigue per consentire di affrontare le casistiche di incertezza e non commettere errori nell'output finale.

Un case study nel contesto dell'Image Recognition

Il capitolo introduce l'Image Recognition come una delle tecnologie più dinamiche e a maggior impatto nell'ambito dell'Intelligenza Artificiale computazionale. Questa disciplina è oggi adottata in numerose situazioni che spaziano dal rilevamento di oggetti e testi all'interno delle immagini fino all'analisi facciale avanzata. Il capitolo si pone due obiettivi: delineare una definizione degli aspetti tecnici della disciplina, e, parallelamente, valutarne l'efficacia attraverso un'analisi sperimentale. Nella prima parte del capitolo verranno presentati la definizione, gli aspetti tecnici e le principali applicazioni attuali. Successivamente verrà spiegata passo dopo passo la modalità utilizzata per implementare la tecnologia attraverso la piattaforma Amazon Rekognition. La sezione fondamentale del capitolo è rappresentata da un case study volto a testare e mostrare le funzionalità chiave del servizio, dalla label detentio alla videosorveglianza smart, con il fine ultimo di mostrare i principali punti di forza e le principali limitazioni della piattaforma.

4.1 Image Recognition

Al giorno d'oggi lo scambio di immagini e la loro analisi rappresenta un punto cruciale per la sicurezza delle persone, sia online che offline. Basti pensare alla vastità di effetti che un singolo scatto può comportare, dal ricordo di un momento felice e conviviale, al poter ledere la privacy dei soggetti in essa rappresentati. Diventa, per questo motivo, di fondamentale importanza la possibilità di analizzare le immagini al fine di segnalare eventuali problemi e irregolarità. Parallelamente, uno strumento di tale portata potrebbe trovare applicazione in compiti meno critici per la sicurezza, come il rilevamento delle emozioni tramite analisi facciale, la comparazione somatica tra individui o il riconoscimento di personaggi celebri. In questo scenario l'Image Recognition si pone come tecnologia risoltrice delle questioni discusse, applicando tecniche di *Computer Vision* e *Deep Learning* per poter catalogare in completa autonomia ogni tipologia di contenuto multimediale che passa attraverso la rete, garantendo così la sicurezza delle persone in ogni sua sfaccettatura e la massima tempestività nel rilevamento di potenziali minacce o anomalie.

Allo stesso modo della Sentiment Analysis, l'Image Recognition rappresenta un importante alleato strategico delle organizzazioni che utilizza determinati strumenti per adottarla. Senza tale tecnologia le aziende responsabili dell'analisi di enormi volumi di contenuti multimediali si troverebbero sommerse da carichi di lavoro insostenibili; ciò rende questo strumento un pilastro fondamentale per la conduzione delle attività operative.

Nel presente capitolo, l'Image Recognition costituisce la base teorica di partenza per poter comprendere il case study attuato tramite il servizio Amazon Rekognition. Tale introduzione

viene fatta per fornire una panoramica più chiara ed esaustiva del contesto, con il fine ultimo di far comprendere al lettore i punti di forza e le limitazioni della piattaforma.

4.1.1 Definizione di Image Recognition

Il riconoscimento di immagini (Image Recognition) è un'applicazione avanzata del *Machine Learning*, volta a far sì che i software e i dispositivi abbiano la possibilità di identificare oggetti, luoghi, persone, scritti e azioni in immagini o in video digitali. L'obiettivo principale di tale tecnologia non è la semplice scansione visiva, ma l'estrazione di informazioni dai contenuti multimediali, permettendo ai sistemi informatici di comprendere il contesto visivo allo stesso livello dell'essere umano.

La Figura 4.1 illustra il principio di funzionamento del sistema di *Image Recognition*.

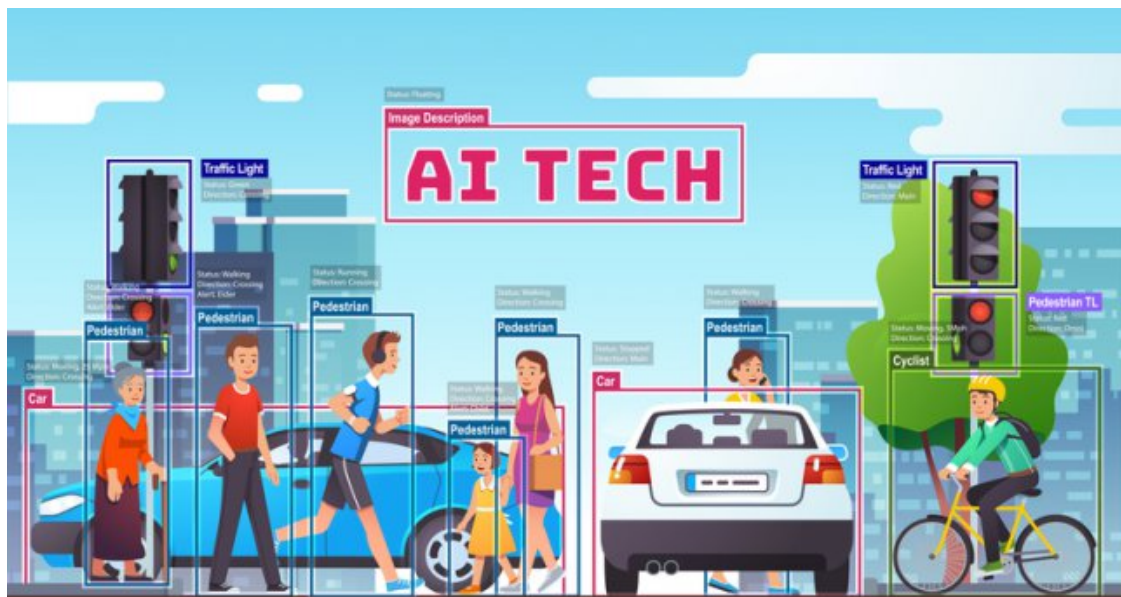


Figura 4.1: Illustrazione del principio di funzionamento dell'*Image Recognition* (fonte: fastweb.it)

4.1.2 Aspetti tecnici sull'Image Recognition

L'Image Recognition è uno dei campi più ampi dell'Intelligenza Artificiale; esso prende possibile ai software e alle macchine la comprensione dei dati e la reazione a questi ultimi.

Nell'ambito del riconoscimento delle immagini i progettisti integrano tecniche di *Machine Learning* e *Deep Learning* in modo separato o parallelo, a seconda della complessità del problema affrontato.

Di seguito viene mostrata la procedura tecnica con la quale gli scienziati attuano il riconoscimento dei contenuti multimediali con il *Machine Learning* tradizionale:

- *Normalizzazione*: la normalizzazione rappresenta la fase in cui i tecnici preparano le immagini scalando i pixel all'interno di un intervallo predefinito (tipicamente $[0,1]$ o $[-1,1]$) per ottimizzare lo step successivo di analisi. Questa fase include anche il ridimensionamento delle immagini, la conversione in scala di grigi o la rimozione del rumore¹ utilizzando tecniche di filtraggio gaussiane.

¹In questo caso per "rumore" si intende qualsiasi variazione indesiderata dei pixel, ad esempio un'immagine sfocata

- *Estrazione delle caratteristiche*: considerato che i modelli di apprendimento automatico si basano su caratteristiche estratte manualmente, in questa fase i tecnici hanno il compito di selezionare i dettagli più rilevanti, dai bordi, se si tratta di forme, all'intensità del colore, se si sta parlando della distinzione di oggetti basata sulla tonalità, permettendo, così, ai modelli di catalogare più facilmente oggetti specifici.
- *Codifica per il Machine Learning*: in questa fase i tecnici formattano in vettori numerici le caratteristiche estratte dallo step precedente riassumendo l'importanza di ogni dettaglio e facilitando l'elaborazione ai modelli.
- *Riconoscimento delle immagini tramite deep learning*: in questo step, tramite il *Deep Learning*, spiegato nella Sezione 1.4.3, i modelli utilizzano reti neurali a strati per apprendere direttamente dalle immagini, utilizzando, però, un costo computazionale più elevato.
- *Livello di input*: a questo livello il sistema riceve i dati grezzi dell'immagine multimediale. Nel *Machine Learning* tradizionale, questi dati richiedono una pre-elaborazione manuale da parte degli scienziati per estrarre le caratteristiche rilevanti; al contrario, nelle moderne reti neurali convoluzionali (CNN)², i pixel vengono trasmessi direttamente attraverso la rete, la quale estrae in autonomia gli schemi visivi necessari alle successive fasi di analisi.
- *Estrazione delle caratteristiche*: il livello convoluzionale applica dei filtri sull'immagine per rilevare pattern locali, come bordi o texture, eliminando l'estrazione manuale delle caratteristiche; queste ultime vengono, infatti, apprese direttamente dai dati in input. Alla fine di ogni convoluzione, grazie all'aggiunta di una funzione di attivazione, aumenta la capacità del modello di apprendere forme e oggetti via via più complessi.
- *Pooling e flattening*: questi livelli effettuano un sottocampionamento dell'immagine per ridurre le dimensioni preservando, allo stesso tempo, le caratteristiche fondamentali ed evitando la perdita di informazioni chiave. Successivamente queste ultime vengono incapsulate in un vettore e fatte passare attraverso dei livelli che integrano pattern appresi nelle fasi precedenti al fine di perfezionare la categorizzazione delle immagini.
- *Livello di output*: in questa fase finale, il consolidamento delle caratteristiche estratte permette di formulare una previsione. Il successivo confronto con il dataset di addestramento consente di quantificare l'errore e di ottimizzare i pesi della rete neurale. A seconda della disponibilità dei dati e della natura del problema di Image Recognition, il processo di ottimizzazione può seguire quattro principali paradigmi di apprendimento: apprendimento supervisionato, non supervisionato, autosupervisionato e per rinforzo, esplorati e discussi nella Sezione 1.4.1.

Un esempio concreto di quanto discusso è osservabile nella Figura 4.2, che mostra un sistema di *Image Recognition*, nello specifico un algoritmo di *Object Detection*, applicato al monitoraggio del traffico urbano in tempo reale.

Attraverso il processo analizzato, il modello elabora l'immagine e produce il livello di output sotto forma di rettangoli di selezione ed etichette predittive. Come si nota nell'immagine:

- Il modello identifica e categorizza i diversi elementi presenti sulla strada distinguendo in totale autonomia, ad esempio, i taxi, le auto private ("car"), i mezzi leggermente più grandi ("truck") e i mezzi di trasporto pubblico ("bus").

²Le Convolutional Network (reti neurali convoluzionali) sono architetture di *Deep Learning* che analizzano e apprendono autonomamente la natura strutturata dei dati visivi attraverso strati successivi.



Figura 4.2: Esempio di *Object Detection* applicato al monitoraggio del traffico urbano in tempo reale (fonte: ibm.com)

- Associa a ciascun rettangolo un'etichetta specifica dimostrando la capacità della rete neurale di comprendere il contesto visivo e di ricondurre gli elementi presenti al suo interno alla corretta classe di appartenenza.

Tale scenario riesce a far comprendere in modo chiaro ed esaustivo l'efficacia e l'efficienza di questa tecnologia. L'automazione del processo elimina la necessità di un'estrazione manuale delle caratteristiche che sarebbe stata, altrimenti, impossibile da gestire per un essere umano.

Esiste, poi, una differenza fondamentale tra i due compiti principali della *Computer Vision*, ovvero la classificazione delle immagini e il rilevamento degli oggetti. In particolare:

- *La classificazione delle immagini* risponde alla domanda "Cosa c'è in questa foto?", riuscendo ad analizzare l'immagine nel suo insieme e assegnando ad una singola etichetta generale.
- *Il rilevamento degli oggetti* approfondisce la semplice classificazione rispondendo alle domande "Dove si trova l'oggetto?" e "Quanti oggetti ci sono nell'immagine?", riuscendo ad identificare la posizione esatta, la dimensione e l'orientamento spaziale dell'oggetto rilevato.

Nonostante le tecniche di riconoscimento delle immagini siano diventate sempre più avanzate nel tempo, esse presentano, tuttora, forti margini di miglioramento riguardo alla precisione e alla stabilità. Tali immagini riguardano le seguenti problematiche:

- *Condizioni di illuminazione:* ombre, variazioni di luminosità improvvise o ambienti con poca luce potrebbero portare ad alterare caratteristiche critiche, facendo sì che il modello interpreti erroneamente quello a cui viene sottoposto. Esistono, tuttavia, metodi avanzati, ad esempio, l'equalizzazione adattiva dell'istogramma o l'integrazione dei dati di addestramento con condizioni di illuminazione variabili, che rendono i modelli più resilienti a determinate anomalie.

- *Limitazioni nei dati di addestramento*: i modelli addestrati su immagini prevalentemente ad alta qualità faranno, poi, difficoltà nel momento in cui verranno sottoposti ad immagini qualitativamente inferiori; per questo motivo si decide di optare all'addestramento utilizzando qualsiasi tipo di risoluzione degli elementi.
- *Dimensioni e prossimità dell'oggetto*: la vicinanza di un determinato oggetto alla fotocamera influenza notevolmente le capacità di questa tecnologia, in quanto, se l'entità risulta troppo lontana, il modello non riesce a estrarre un quantitativo di dati tale per cui risulta in grado di capire di cosa si tratta, mentre se si trova troppo vicino l'elemento apparirà distorto per essere correttamente classificato; per risolvere tali problemi gli ingegneri addestrano i modelli su set di dati che includono oggetti di varie dimensioni e a varie distanze.

4.1.3 Applicazioni attuali dell'Image Recognition

Il settore dell'Image Recognition, essendo uno dei più avanzati nell'era moderna, sta registrando una crescita esponenziale ad un livello tale da guidare lo sviluppo di un numero sempre maggiore di applicazioni software. Tali soluzioni trovano ampio impiego nell'ambito della sicurezza e nell'ottimizzazione della produttività giornaliera, contribuendo, in questo modo, a ridefinire radicalmente il concetto stesso di automazione applicata alla vita quotidiana. Le principali aree di applicazione comprendono:

- *Veicoli a guida autonoma*: in questo contesto, l'elaborazione visiva in tempo reale è fondamentale per la percezione dell'ambiente circostante. Attraverso l'analisi in tempo reale dei video il veicolo è in grado di mappare lo spazio, riconoscere la segnaletica e prevenire collisioni. Nella Figura 4.3 si presenta un esempio di questa casistica di applicazione.



What the vehicle sees (FSD V9 Beta)

T E S L A

Figura 4.3: Esempio di elaborazione in spazio vettoriale tramite la proiezione Bird's-Eye View nel sistema Tesla FSD. L'algoritmo di Computer Vision unifica i flussi video bidimensionali per generare una mappa tridimensionale predittiva dell'ambiente circostante (fonte: autoevolution.com)

- *Riconoscimento facciale*: nei settori della sicurezza e della sorveglianza la *Computer Vision* è ormai diventata una tecnologia fondamentale. Essa supporta le forze dell'ordine nel monitoraggio dei luoghi affollati e nell'identificazione di soggetti d'interesse, automatizzando, così, l'analisi di scenari complessi. L'assenza di tali sistemi automatizzati renderebbe notevolmente complessa la gestione della sicurezza pubblica, soprattutto durante eventi o parate su larga scala. Nella Figura 4.4 viene presentato un esempio di questa casistica di applicazione.

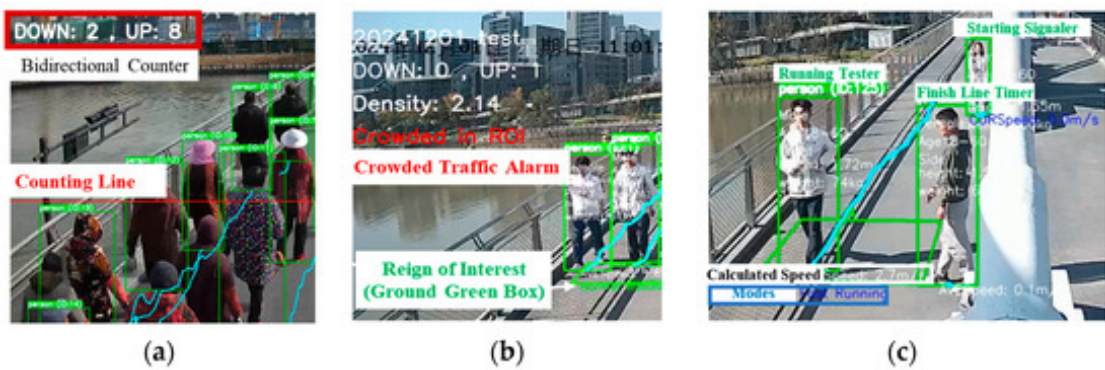


Figura 4.4: Implementazione di un sistema automatizzato di *Computer Vision* per il monitoraggio e la gestione dei flussi pedonali in contesti urbani. L'architettura integra algoritmi di *object detection* e *multi-target tracking* per stimare in tempo reale densità, velocità e traiettorie dei soggetti. (fonte: mdpi.com)

- *Gestione e moderazione dei social media*: i social network integrano algoritmi di riconoscimento visivo per automatizzare l'applicazione di etichette fotografiche, identificando i tratti caratteristici degli utenti. Parallelamente, tale tecnologia supporta i filtri di Realtà Aumentata (AR), i quali mappano i punti di riferimento facciali per calibrare il posizionamento di elementi virtuali in base ai movimenti del volto. Oltre a queste funzionalità, l'analisi automatizzata consente di intercettare e filtrare elementi inappropriati o non conformi alle linee guida, garantendo la sicurezza della piattaforma e ottimizzando l'esperienza d'uso dell'utente.
- *Occhiali smart e sistemi di realtà aumentata*: gli occhiali intelligenti, dotati di software di *Computer Vision*, consentono di sovrapporre flussi informativi in tempo reale direttamente sul campo visivo dell'utente fornendo dati contestuali sugli oggetti e sui luoghi inquadrati.
- *Elettrodomestici intelligenti e domotica*: l'integrazione del riconoscimento delle immagini nell'ecosistema degli elettrodomestici abilitati all'IoT³ permette funzionalità avanzate di monitoraggio e automazione domestica. Nei frigoriferi intelligenti, ad esempio, consente la gestione automatizzata dell'inventario e delle scadenze, mentre nei sistemi di aspirazione robotica guida il rilevamento e l'aggiramento dinamico degli ostacoli. La medesima tecnologia ottimizza i cicli di lavaggio nelle lavatrici mediante il riconoscimento del tipo di tessuto e, supporta la cottura assistita nei forni attraverso l'identificazione degli alimenti.
- *Robotica autonoma e logistica industriale*: nel settore della logistica, i robot adibiti alle consegne si affidano alla *Computer Vision* per la rilevazione degli ostacoli durante la navigazione in contesti urbani. In modo analogo, all'interno dei magazzini e degli

³Internet of Things

stabilimenti industriali, i veicoli a guida automatizzata e i bracci robotici sfruttano la medesima tecnologia per la scansione e il recupero degli oggetti.

- *Diagnostica per immagini e ambito medicale:* l'analisi computazionale delle immagini mediche assiste il personale sanitario nell'interpretazione di radiografie e risonanze magnetiche. I sistemi basati sulla *Computer Vision* sono in grado di isolare anomalie microscopiche che potrebbero sfuggire all'esame visivo umano favorendo una diagnosi precoce. In Figura 4.5 si presenta un esempio di questa casistica di applicazione.

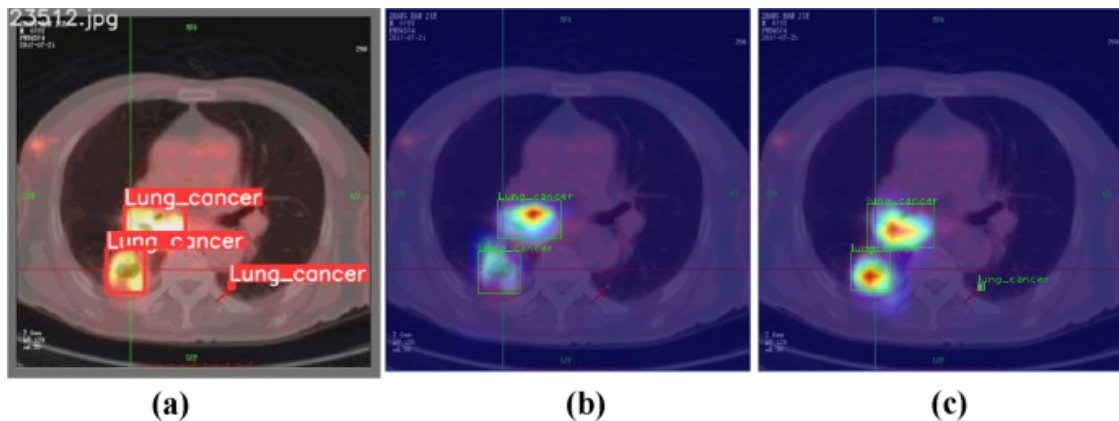


Figura 4.5: Esempio di localizzazione di un tumore polmonare mediante l'architettura multimodale Cross Modal YOLOv5 (CMYOLOv5). Il sistema integra le informazioni anatomiche ad alta risoluzione della TC⁴ con quelle fisiopatologiche e metaboliche della PET⁵ ottimizzando il rilevamento di lesioni geometricamente irregolari (fonte: nature.com)

- *Elaborazione automatizzata dei documenti in ambito finanziario:* gli istituti bancari e le società finanziarie impiegano il riconoscimento delle immagini per automatizzare la verifica dei documenti d'identità, degli assegni e delle pratiche contrattuali, snellendo le procedure di onboarding e di Know Your Customer (KYC).

4.2 Contesto e metodologia per attuare l'Image Recognition

In merito all'implementazione delle tecniche di *Computer Vision*, l'analisi si è concentrata sull'impiego di Amazon Rekognition, una soluzione AWS già introdotta nella Sezione 2.5.2. Tale piattaforma si basa sull'utilizzo di reti neurali profonde per l'estrazione di dati dai contenuti visivi catturati dai dispositivi ottici, integrando le funzionalità di *object detection* e segmentazione delle immagini. Lo strumento è stato applicato ad una varietà di situazioni con l'obiettivo di testarne l'efficacia operativa dinanzi a casistiche aziendali concrete. In particolare, lo studio spazia dal rilevamento di oggetti e dall'analisi biometrica dei volti, fino alla verifica dell'identità per il controllo degli accessi e al tracciamento temporale di flussi video per la videosorveglianza intelligente.

Il percorso che verrà successivamente presentato segue la stessa struttura logica utilizzata nel precedente caso di studio, articolandosi nelle seguenti fasi:

1. identificazione degli scenari operativi e dei requisiti applicativi;
2. selezione, ottimizzazione tecnica e caratterizzazione del dataset visivo;

⁴Tomografia Computerizzata

⁵Tomografia a Emissioni di Positroni

3. estrazione e filtraggio dei dati strutturati;
4. analisi qualitativa e valutazione dell'accuratezza delle rilevazioni;
5. bilancio dei punti di forza e dei vincoli tecnologici riscontrati.

Nella sezione successiva verrà illustrata l'implementazione di Amazon Rekognition descrivendo le procedure necessarie a ottenere un'analisi ottimale ed accurata dei dati raccolti.

4.2.1 Implementazione con Amazon Rekognition

Al fine di analizzare in modo ottimale i dati con Amazon Rekognition, anche per questo servizio verrà effettuato l'accesso alla AWS Management Console, verrà scelto il servizio "Amazon Rekognition" e, infine, verrà selezionato "Prova la demo".

Nonostante anche questo servizio offra la possibilità di operare su larga scala tramite chiamate API o importando i file multimediali direttamente da bucket Amazon S3, si è scelto di procedere attraverso l'interfaccia grafica predisposta da AWS. Tale modalità consente di caricare direttamente i file d'esempio (in formato JPEG, PNG o MP4) nell'apposita area di upload per visualizzare istantaneamente l'output grafico e i metadati associati.

Nella Figura 4.6 viene illustrata la situazione in cui l'utente si dovrebbe trovare avendo seguito tutti i passaggi sopra descritti.

The screenshot displays the Amazon Rekognition console interface. On the left, a sidebar lists various features under 'Moderazione personalizzata' and 'Demo'. The 'Demo' section is expanded, showing 'Rilevamento delle etichette' (Label Detection) as the selected option. The main area features a video player showing a street scene with a person crossing the road, overlaid with bounding boxes and labels. Below the video, there are options to 'Scegliere un'immagine di esempio' (Choose an example image) and 'Utilizza la tua immagine' (Use your image), with buttons for uploading or dragging and dropping an image. The right panel, titled 'Cerca tutte le etichette' (Search for all labels), contains a search bar and a table of results. The table lists labels such as 'Neighborhood', 'City', 'Road', 'Street', 'Urban', and 'Person' with their respective confidence scores. The 'Person' label has a confidence score of 98.7% and a 'Mostra di più' (Show more) link.

Cerca tutte le etichette	
Verifica se supportiamo la tua etichetta	
Inserisci un nome per l'etichetta	
Risultati	
Neighborhood	99.9 %
City	99.9 %
Road	99.9 %
Street	99.9 %
Urban	99.9 %
Person	98.7 %
Mostra di più	
Richiesta	
Risposta	

Figura 4.6: Illustrazione della configurazione di Rekognition

È possibile, inoltre, navigare nel menù laterale per selezionare la specifica funzionalità di Computer Vision da testare di volta in volta. Ciascuno strumento prevede l'utilizzo di modelli pre-addestrati e gestiti nativamente da AWS, i quali applicano automaticamente riquadri di delimitazione e calcolano i punteggi di confidenza per ogni elemento rilevato. I microservizi che mette a disposizione Rekognition comprendono:

- *Microservizi con possibilità di addestrare i modelli:*

- *Moderazione personalizzata*: al fine di ottenere una maggiore precisione delle previsioni è possibile adattare il modello di moderazione pre-addestrato di Amazon Rekognition.
- *Utilizzo di etichette personalizzate*: è utilizzabile qualora si voglia creare un modello di *Machine Learning* personalizzato per trovare oggetti e scene unici per una determinata organizzazione; viene utilizzato, ad esempio, per il controllo degli sponsor, il rilevamento dei momenti salienti all'interno di una molteplicità di immagini e l'automazione in fabbrica o in magazzino. Nella Figura 4.7 vengono mostrati i passaggi da seguire per utilizzare questo microservizio.

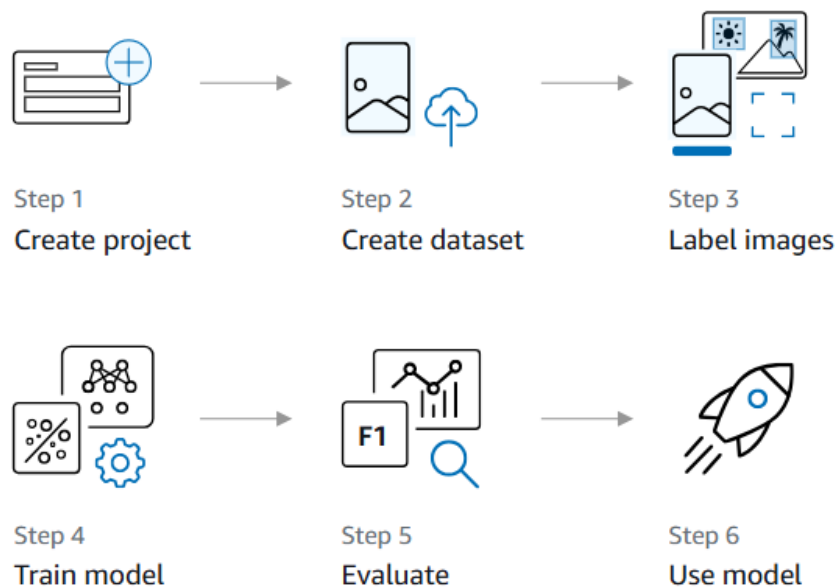


Figura 4.7: Illustrazione dei passaggi da seguire per utilizzare il microservizio di etichette personalizzate di Rekognition (fonte: aws.amazon.com)

- *Microservizi che utilizzano modelli built-in di AWS:*
 - *Analisi di massa*: consente di analizzare fino a 10.000 immagini contemporaneamente con una funzionalità di riconoscimento immagini supportata. Nella Figura 4.8 vengono mostrati i passaggi da seguire per utilizzare questa piattaforma.

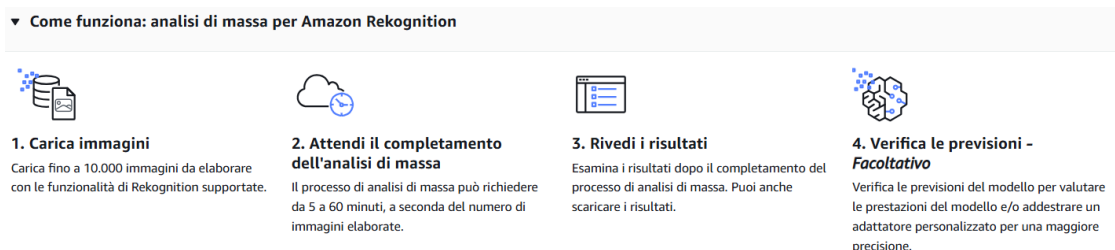


Figura 4.8: Illustrazione dei passaggi da seguire per utilizzare il microservizio di analisi di massa di Rekognition (fonte: aws.amazon.com)

- *Rilevamento delle etichette*: Rekognition etichetta automaticamente oggetti, concetti e scene nelle immagini e fornisce un punteggio di affidabilità da 0% a 100%.

- *Proprietà dell'immagine*: attraverso tale modalità Rekognition riesce a rilevare varie proprietà di un'immagine, come i colori dominanti, la luminosità, le ombre e i contrasti.
- *Moderazione delle immagini*: Rekognition rileva automaticamente contenuti espliciti nelle immagini e fornisce punteggi di affidabilità da 0% a 100%.
- *Analisi facciale*: questo microservizio fornisce la possibilità di ottenere un'analisi completa degli attributi facciali delle persone rappresentate in un'immagine, inclusi i punteggi di affidabilità da 0% a 100%.
- *Confronto facciale*: Rekognition permette il confronto dei volti per analizzarne la corrispondenza in base a una percentuale di somiglianza.
- *Vivacità del viso*: questa sezione permette di verificare la vivacità di un viso in tempo reale con un punteggio di affidabilità che spazia da 0% a 100%.
- *Riconoscimento di volti celebri*: in questo caso Rekognition riconosce automaticamente i volti celebri associando ad essi dei punteggi di affidabilità da 0% a 100%.
- *Testo nell'immagine*: attraverso questa modalità è possibile estrarre automaticamente il testo dalle immagini.
- *Rilevamento dei PPE*: questo microservizio è utilizzato per rilevare automaticamente i dispositivi di protezione individuale (PPE), come coperture per il viso, caschetti da lavoro e guanti sulle persone.
- *Analisi dei video archiviati*: Rekognition, attraverso questo servizio, analizza i video archiviati all'interno di Amazon S3 e rileva oggetti, attività, persone e celebrità.
- *Eventi video in streaming*: tramite questo strumento viene fornita un'analisi in tempo reale, basata sul *Machine Learning*, di streaming video in diretta per rilevare etichette come persone, animali domestici e pacchi.

A causa dell'elevato numero di immagini necessarie a illustrare i diversi risultati, in questo caso contrariamente a quanto fatto nel caso studio precedente, è stato scelto di includere i riscontri visivi e il formato dei risultati direttamente all'interno delle singole sezioni sperimentali. Si noti, comunque, che tutti i microservizi presentati forniscono i propri output sia tramite interfaccia grafica sia tramite API.

4.3 Case study: Analisi sperimentale dei servizi di Amazon Rekognition

Il case study riportato in questo capitolo si concentra sul settore della gestione dei flussi multimediali e della sicurezza. Tali ambiti, in questa sezione, si fondano su due principi cardine, quali la comprensione e l'elaborazione automatizzata dei contenuti visivi. L'ottimizzazione di questi processi rappresenta, infatti, un acceleratore e un vero e proprio vantaggio competitivo per le aziende che mettono in atto questo tipo di automatismi. Applicare l'*Image Recognition* in questo contesto significa trovare una soluzione al monitoraggio manuale dei file multimediali e dei contenuti video, riducendo i tempi di visualizzazione da parte dei dipendenti e, al contempo, permettendo l'identificazione tempestiva di oggetti d'interesse, dati biometrici o fattori di rischio per la sicurezza pubblica e privata.

L'esperimento mira a testare la capacità di Amazon Rekognition di interpretare la presenza di oggetti di qualsiasi tipo, di elementi commerciali e di scenari (Label Detection), di estrarre i tratti somatici e le emozioni più significative (Face Detection and Analysis) e di identificare

figure pubbliche (Celebrity Recognition). Inoltre, il lavoro esplora l'efficacia della piattaforma nell'ambito del tracciamento temporale di soggetti in flussi video di videosorveglianza smart, delineando il passaggio dall'analisi dei singoli fotogrammi statici all'elaborazione di contenuti video. Di seguito, per ogni microservizio di Rekognition, vengono analizzate nel dettaglio le fasi di elaborazione dei campioni utilizzati, i parametri metrici emersi dall'elaborazione e una successiva valutazione qualitativa degli output restituiti dalla piattaforma cloud.

4.3.1 Rilevamento e classificazione degli oggetti (Labels Detection)

Il primo microservizio analizzato si posiziona nel contesto dell'ottimizzazione logistica, della gestione automatizzata degli asset visivi aziendali alla sicurezza sia pubblica che privata, riconoscendo persone, contesti e oggetti pericolosi. All'interno di scenari aziendali complessi, come la catalogazione dei prodotti in un magazzino o l'etichettatura dei contenuti per un catalogo digitale, l'identificazione manuale degli oggetti richiede un dispendio notevole di tempo e risorse introducendo, inoltre, il rischio di errori o omissioni umane. Relativamente agli scenari critici, come il monitoraggio di luoghi pubblici o privati, il riconoscimento di situazioni di pericolo richiede un'elevata concentrazione da parte di coloro che sono responsabili della sicurezza e, inoltre, è necessaria una costante sorveglianza umana.

L'esperimento pratico è stato condotto fornendo in input alla piattaforma cloud molteplici immagini campione rappresentative di un contesto commerciale e di sicurezza pubblica. L'obiettivo del test è valutare l'accuratezza dell'algoritmo nel riconoscere sia gli oggetti specifici in primo piano, sia gli elementi di sfondo, testando l'affidabilità della classificazione offerta dai modelli *built-in* di AWS.

La pianificazione dei test è stata strutturata partendo da inquadrature panoramiche estese, caratterizzate da un'elevata densità di oggetti, per poi passare progressivamente verso riprese più ravvicinate e strutturalmente semplici.

Per ciascun test verranno mostrate insieme l'immagine iniziale priva di elaborazioni, l'immagine elaborata e l'estratto del *confidence score*. Di seguito vengono riportati i test svolti suddivisi per categoria e i contestuali risultati in cui sono identificate le etichette di catalogazione.

- *Catalogazione di prodotti all'interno di un magazzino (parte 1)*: per approfondire questa casistica ed esaminare l'efficacia del servizio in scenari commerciali complessi, sono stati scelti come campioni di analisi gli scaffali di alcuni supermercati. Nello specifico, la Figura 4.9 mostra come si è comportato l'algoritmo relativamente al primo test per questa categoria.

Il primo test sottopone ad Amazon Rekognition un'inquadratura ad ampio raggio di uno scaffale saturo di prodotti. In questa condizione di forte affollamento visivo e distanza ottimale ridotta, l'algoritmo *built-in* attua un approccio di classificazione macroscopico. Come evidenziato dalla tabella dei risultati, i punteggi di confidenza più elevati si concentrano su etichette contestuali generali come *Indoors* (99.9%), *Shop* (99.4%), *Market* (99.6%), *Grocery store* (95.7%) o *Supermarket* (95.7%). Il sistema riconosce perfettamente l'ambiente e la natura merceologica generale, ma fallisce nell'isolare i singoli pacchi o i brand dei prodotti dimostrando che la distanza e l'elevata densità di oggetti riducono la capacità di Amazon Rekognition di attuare la *label detection*.

- *Catalogazione di prodotti all'interno di un magazzino (parte 2)*: nel secondo test (Figura 4.10), l'inquadratura viene ravvicinata concentrandosi su un numero inferiore di prodotti distribuiti su pochi ripiani. Questo cambio di prospettiva si riflette immediatamente sull'output dell'algoritmo: sebbene rimangano indici di confidenza elevati per le categorie

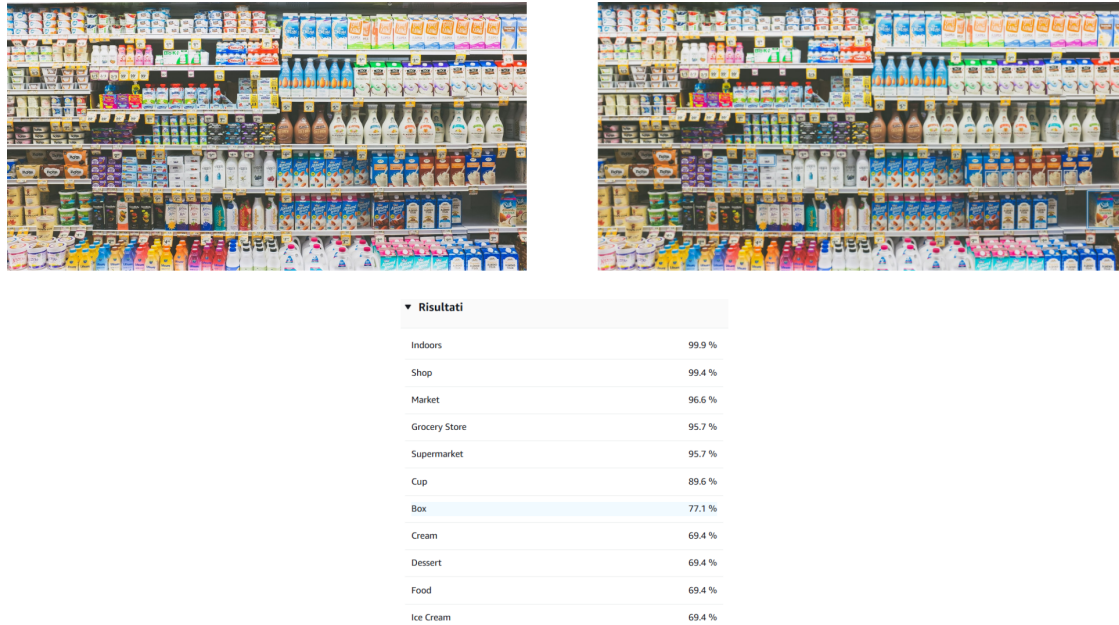


Figura 4.9: Illustrazione del comportamento di Rekognition relativamente a degli scaffali di un supermercato (parte 1)

macro (come *Shelf* (99.6%) o *Shop* (99.2%)), la riduzione della densità visiva permette a Rekognition di tentare una classificazione più specifica. Emergono, infatti, etichette legate alle geometrie dei contenitori (come *Cup* (90.2%) o *Tin* (63.9%)). Tuttavia, anche in questo scenario più favorevole, l'algoritmo non è in grado di distinguere analiticamente ogni singola confezione presente, confermando la difficoltà nel tracciare confini netti tra prodotti molto vicini con pattern grafici complessi; tale complessità produce anche errori relativi alle illustrazioni stampate sulle confezioni che il sistema talvolta scambia erroneamente per persone reali, come dimostrato dal rilevamento della label *Person* (83.9%) tra i risultati.

- *Catalogazione di prodotti all'interno di un magazzino (parte 3):* nel terzo test (Figura 4.11), l'inquadratura viene ravvicinata ulteriormente concentrandosi, questa volta, su un numero inferiore di prodotti distribuiti su due ripiani. Questo cambio di prospettiva si riflette in modo parziale sull'output dell'algoritmo: nonostante permangano indici di confidenza massimi per le macro-categorie corrette, come "*Food*" e "*Sweets*" (entrambe al 99.7%), l'algoritmo continua a mostrare forti difficoltà nel distinguere le singole confezioni. La criticità più evidente di questo test risiede, però, nell'attivazione di un falso positivo: il sistema rileva l'etichetta "*First Aid*" (kit di primo Soccorso) con un fattore di confidenza estremamente elevato pari al 95.5%. Questo errore di classificazione deriva, con ogni probabilità, da una distorsione di forme e di colore in cui i pattern grafici delle confezioni inferiori, caratterizzati da scritte con dettagli rossi e blu su sfondo bianco, vengono erroneamente associati dall'algoritmo a kit di pronto soccorso, evidenziando, ancora una volta, l'inaffidabilità dei modelli in presenza di situazioni dense e complesse.
- *Gestione logistica e sicurezza in ambito aeroportuale (caso diurno):* al fine di testare il comportamento del modello *built-in* in un ambiente ad alta complessità dinamica, la seconda

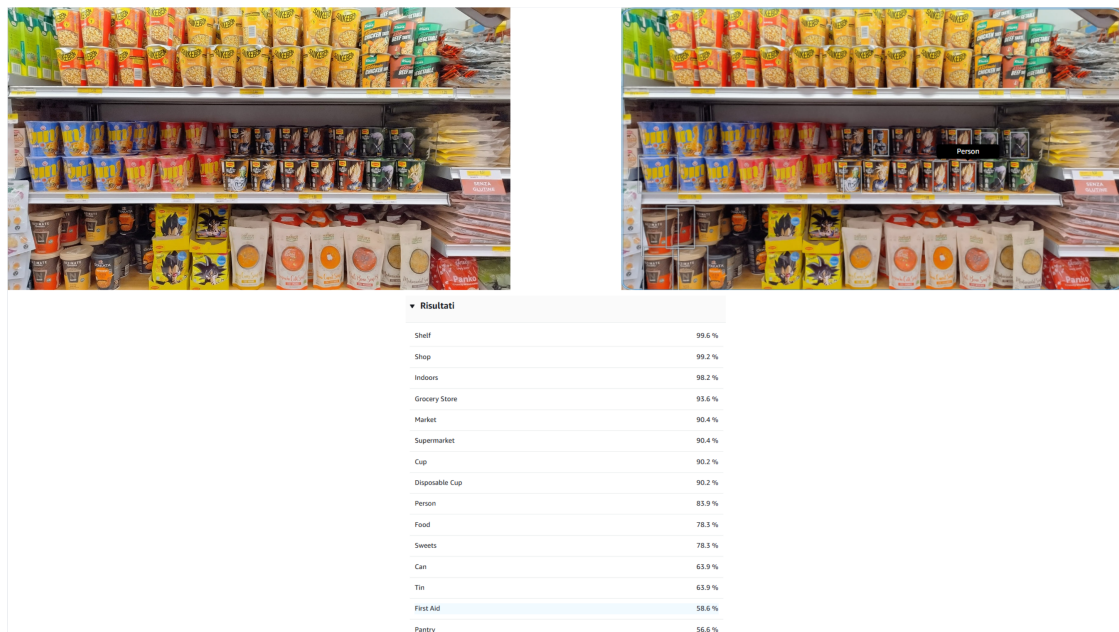


Figura 4.10: Illustrazione del comportamento di Amazon Rekognition relativamente a degli scaffali di un supermercato (parte 2)

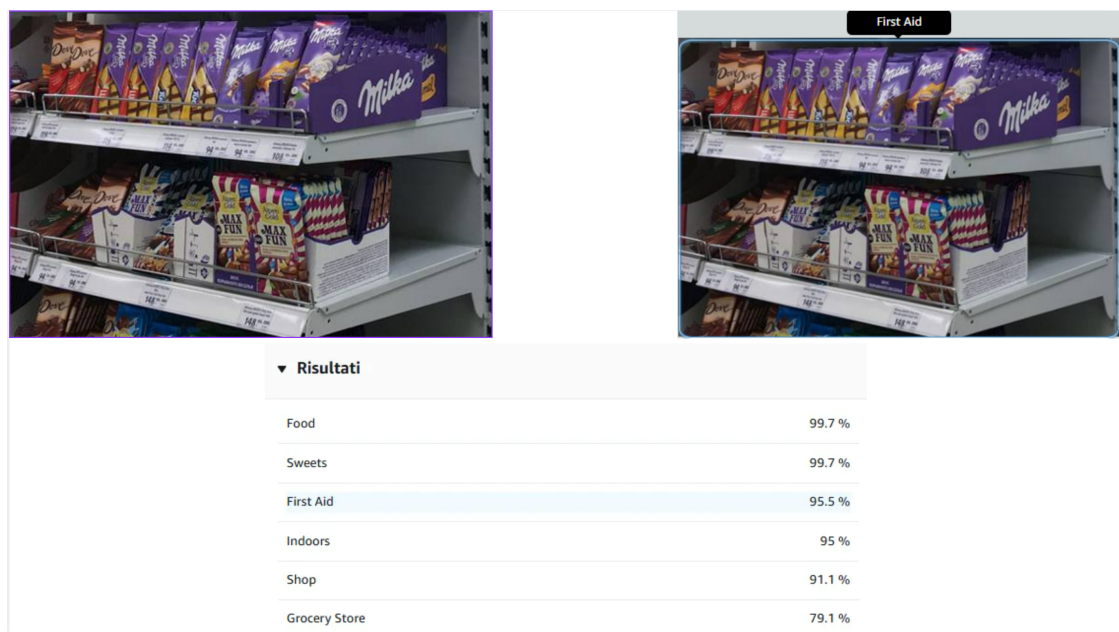


Figura 4.11: Illustrazione del comportamento di Amazon Rekognition relativamente a degli scaffali di un supermercato (parte 3)

sessione sperimentale in ambito *Labels Detection* si sposta sul monitoraggio delle infrastrutture aeroportuali. In questo contesto, l'*Image Recognition* non serve solo alla catalogazione statica, ma diventa uno strumento di supporto per la logistica di terra e per la sicurezza perimetrale. Le immagini sono state ottenute da scatti derivanti dalle dirette streaming dell'aeroporto di Malta (fonte: <https://maltairport.com/webcam/>). Nelle Figura 4.12 viene mostrato come si è comportato l'algoritmo relativamente al primo test di questa categoria.

Il primo test di questa serie analizza un'inquadratura panoramica grandangolare del

piazzale di sosta e delle infrastrutture di volo. Come evidenziato dal riscontro visivo, l'algoritmo dimostra un'elevata precisione geometrica riuscendo a mappare e isolare tramite *bounding box* (i riquadri colorati che isolano gli elementi nella foto inferiore) sia gli aeromobili posizionati a grande distanza sullo sfondo, sia i mezzi logistici in movimento in primo piano.

Il sistema identifica lo scenario generale con un livello di accuratezza quasi perfetto registrando, infatti, la label "Airport" al 99.8%. La stabilità del modello *built-in* emerge chiaramente nel riconoscimento simultaneo delle classi "Aircraft", "Airplane", "Transportation" e "Vehicle", tutte associate a una confidenza del 94.5%. Scendendo nel dettaglio degli elementi di terra, l'algoritmo rileva correttamente le classi "Car" (89.2%), "Runway" (85.6%) e "Truck" (80.9%). Questo test, quindi, dimostra una buona capacità di classificazione anche in presenza di inquadrature dall'alto, tipiche dei sistemi di videosorveglianza.

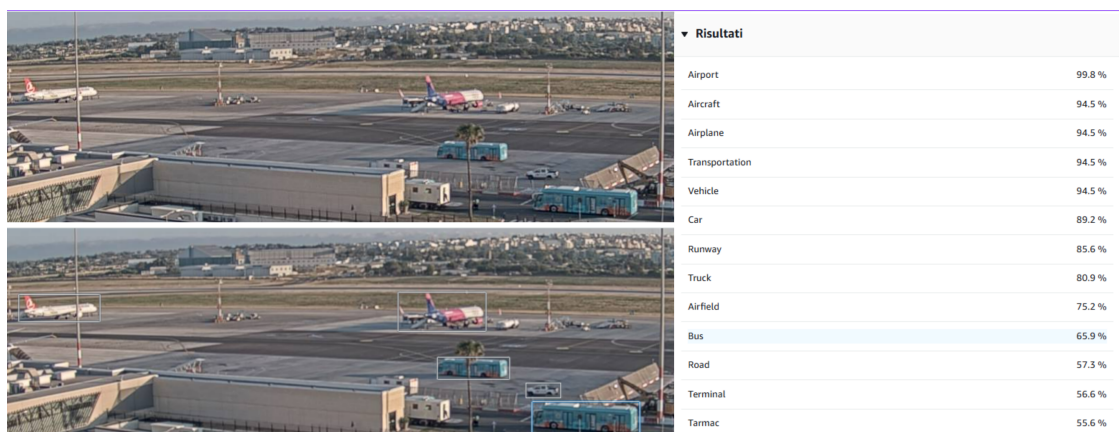


Figura 4.12: Illustrazione del comportamento di Amazon Rekognition relativamente a un aeroporto (caso diurno) (fonte: maltairport.com)

- *Gestione logistica e sicurezza in ambito aeroportuale (caso notturno):* il secondo test in ambito aeroportuale (Figura 4.13) introduce un importante fattore di stress per i modelli di *Computer Vision*: la scarsa visibilità dovuta all'illuminazione notturna. La camera, in questo caso, riprende la stessa posizione di quella del test precedente, ma in condizioni ambientali critiche, caratterizzate da forti contrasti di luce artificiale, riflessi sui sensori della telecamera e una generale perdita di dettaglio sullo sfondo.

Nonostante le condizioni di illuminazione, l'analisi evidenzia una sorprendente stabilità e robustezza dell'algoritmo *built-in*. Amazon Rekognition identifica correttamente il contesto generale registrando la label "Airport" con un punteggio quasi perfetto del 99.9%. L'inquadratura panoramica permette di isolare con precisione le sagome degli aeromobili in sosta sulle piazzole, attivando le classi "Aircraft", "Airplane", "Transportation" e "Vehicle" con una confidenza del 95.5%. Anche le componenti strutturali dell'area vengono rilevate con successo, come dimostrano le etichette "Airfield" (95.1%) e "Terminal" (88.0%).

L'impatto della mancanza di luce si riflette, invece, sulla categorizzazione degli elementi a terra. L'asfalto delle piste, privo di una luce omogenea, perde il contrasto necessario a una distinzione netta; infatti, le label "Road" (71.2%), "Runway" (57.2%) e "Tarmac" (56.9%) subiscono una sostanziale riduzione negli indici di confidenza rispetto al caso diurno. Questo comportamento sperimentale dimostra che, se da un lato l'algoritmo

di AWS possiede un'eccellente resilienza nel tracciamento e nella classificazione di elementi fondamentali anche in condizioni di scarsa visibilità, dall'altro le performance di dettaglio dipendono fortemente dalla qualità e dall'omogeneità dell'illuminazione ambientale, un fattore fondamentale da tenere in considerazione nella progettazione di sistemi di sorveglianza e logistica automatizzata 24 ore su 24.

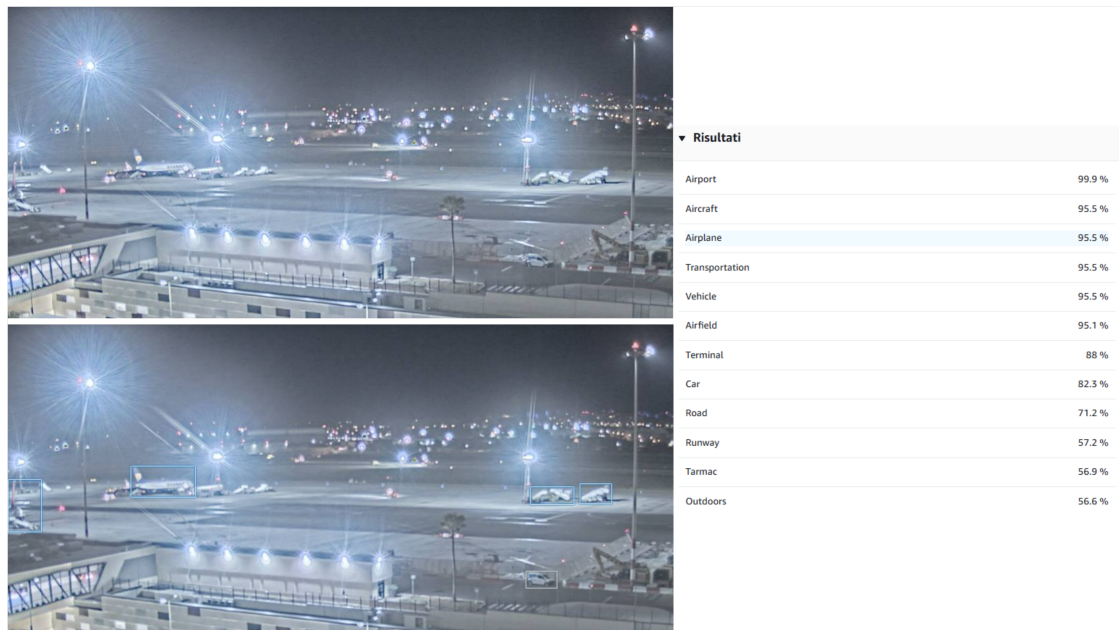


Figura 4.13: Illustrazione del comportamento di Amazon Rekognition relativamente a un aeroporto in condizioni di scarsa visibilità (caso notturno) (fonte: maltairport.com)

- *Gestione logistica e sicurezza in ambito aeroportuale (dettaglio ravvicinato):* il terzo ed ultimo test (Figura 4.14), a distanza ravvicinata, si concentra sulle attività di movimentazione merci e supporto a terra. Questo scenario introduce una complessità notevole dovuta alla forte vicinanza e parziale sovrapposizione tra i diversi elementi logistici.

Come si comprende dai risultati del terzo test, il modello *built-in* si dimostra straordinariamente robusto nel distinguere i trasporti di terra: le classi "Transportation", "Truck" e "Vehicle" registrano tutte un indice di confidenza quasi perfetta, pari al 99.9%. L'algoritmo rileva accuratamente sia il camion telato adibito ai servizi di bordo a fianco all'ala, sia il trattore logistico in primo piano. Inoltre, seppur con una confidenza minore, il sistema recepisce la presenza umana a bordo del trattore porta bagagli rilevando la classe "Person" al 62.1%.

Tuttavia, si evidenziano due limiti significativi dell'algoritmo. Il primo risiede nella mancata rilevazione dell'aeromobile in sosta: la parziale inquadratura della fusoliera, il taglio netto della coda e l'estrema vicinanza al camion telato ingannano i modelli predittivi. Il secondo limite è rappresentato da un'evidente distorsione di classificazione delle infrastrutture: il pontile d'imbarco e le strutture tecniche di sfondo vengono interpretate erroneamente attivando classi errate quali "Architecture" (81.0%), "Turret" (81.0%), "Factory" (57.1%) e "Manufacturing" (55.7%). Questo test conferma che le forti variazioni di prospettiva e i tagli degli oggetti influenzano negativamente le capacità del servizio.

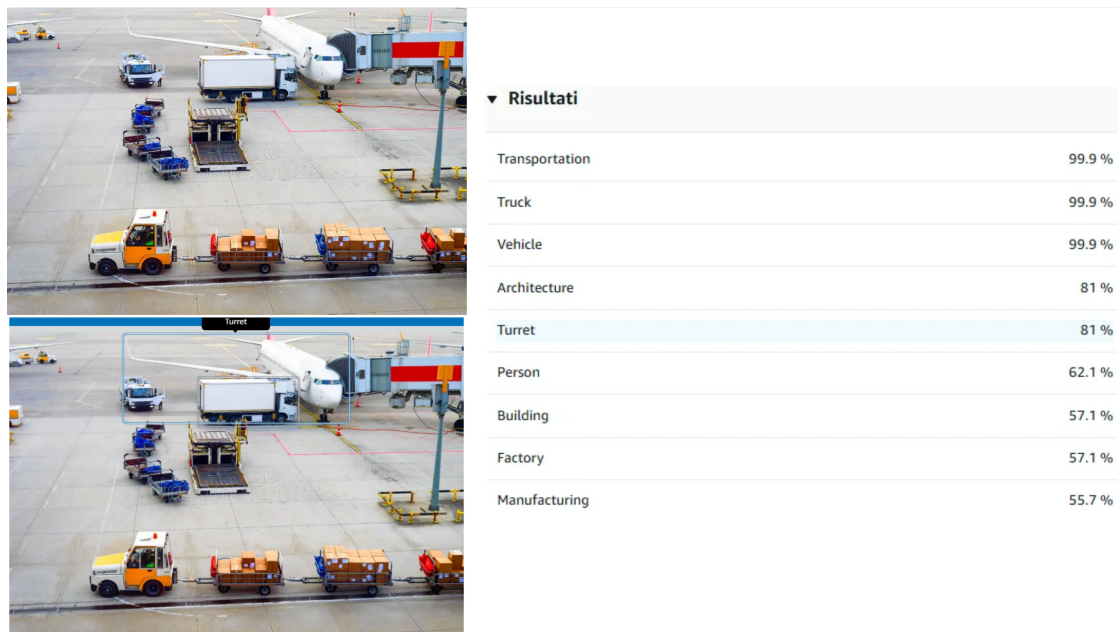


Figura 4.14: Illustrazione del comportamento di Rekognition relativamente a un aeroporto nel dettaglio (dettaglio ravvicinato)

4.3.2 Analisi facciale e riconoscimento delle emozioni (Face Detection and Analysis)

Il secondo microservizio analizzato è collocabile all'interno del contesto del marketing, delle risorse umane e dell'automotive in relazione alla sicurezza. In quest'ultimo ambito, in particolare, l'implementazione dell'analisi facciale a bordo dei veicoli ricopre un ruolo strategico, consentendo il monitoraggio in tempo reale dello stato psicofisico del conducente al fine di prevenirne il colpo di sonno o la distrazione. All'interno di scenari quotidiani complessi, come la guida di un veicolo, lo svolgimento di colloqui di lavoro o la valutazione dell'efficacia degli spot pubblicitari, l'applicazione del riconoscimento facciale permette la rilevazione di tratti biometrici che caratterizzano determinate reazioni emotive degli individui. In assenza di tali tecnologie, risulterebbe notevolmente complicato quantificare il livello di gradimento di una persona alla visione di un messaggio pubblicitario, così come monitorare in tempo reale lo stato di vigilanza di un individuo alla guida. L'utilizzo della *Face Detection and Analysis*, per quanto descritto, ottimizza non solo i processi aziendali, ma assume un posto fondamentale nella sicurezza stradale consentendo il salvataggio di vite umane.

In linea con l'approccio adottato nei casi precedenti, l'obiettivo di questo esperimento è testare le capacità di Amazon Rekognition; in questo specifico caso, si intende misurare l'accuratezza del servizio nell'identificazione dei visi, nello stimare genere ed età, nella classificazione delle emozioni dominanti e delle caratteristiche degli individui. A tale scopo, il dataset presentato è stato appositamente selezionato per includere campioni catturati da varie prospettive e caratterizzati da alcune tipologie di occlusioni fisiche, elementi ideali per valutare la resilienza del modello.

In questo caso, per ciascun test, verranno presentati insieme l'output grafico di Rekognition e i punteggi di confidenza. Di seguito vengono riportati i test condotti suddivisi per categoria aziendale e i contestuali risultati.

- *Analisi facciale di individui alla guida - frontale:* al fine di garantire il rigore scientifico e il rispetto dei diritti d'autore, i campioni esaminati in questa sezione provengono dal

dataset accademico internazionale *Driving Monitoring Dataset (DMD)* Ortega *et al.* [2020]. L'inquadratura utilizzata simula la prospettiva di una telecamera di bordo installata sul cruscotto o in corrispondenza dello specchietto retrovisore del veicolo. La Figura 4.15 rappresenta la prima casistica.

Come evidenziato dal riscontro visivo in Figura 4.15, Amazon Rekognition isola il volto del conducente applicando un *bounding box* azzurro con un indice di accuratezza quasi perfetto (*"looks like a face"* al 99.9%). Questo strumento, inoltre, evidenzia come riesce ad individuare i punti di ancoraggio biometrici (occhi, naso e bocca). Dal punto di vista della profilazione, i modelli predittivi *built in* di Rekognition stimano il genere del soggetto come maschile con una confidenza del 99.4% e ne calcolano una fascia d'età compresa tra i 39 e i 47 anni.

Più nello specifico, i parametri comportamentali, critici per un sistema di monitoraggio alla guida, dimostrano un'efficacia eccellente. Infatti, i dati emersi mostrano le classi *"eyes are open"* al 99.9% e *"mouth is closed"* al 99.6%. La combinazione di queste due caratteristiche dell'individuo potrebbe portare all'esclusione di scenari di rischio legati alla sonnolenza, confermando lo stato di vigilanza dell'autista.

Un ulteriore riscontro è visibile nello stato emotivo in cui emergono le etichette *"not smiling"* (99.5%) e *"appears to be calm"* (95.8%), indici tipici di una guida concentrata e priva di alterazioni. Esaminando nel dettaglio l'array delle emozioni (*"Emotions"*), visibile in Figura 4.16, il sistema attribuisce una confidenza maggiore alla classe *"CALM"* (95.81%), affiancata da valori statisticamente irrilevanti per le altre categorie come *"SAD"* (2.33%) e *"CONFUSED"* (0.41%), mentre emozioni più critiche, come *"FEAR"* (0.015%) o *"ANGRY"* (0.012%), risultano prossime allo zero. Infine, il sistema rileva con successo la presenza di occhiali da vista (*"wearing glasses"* al 99.8%) escludendo la presenza di lenti da sole (*"not wearing sunglasses"* al 99.7%).

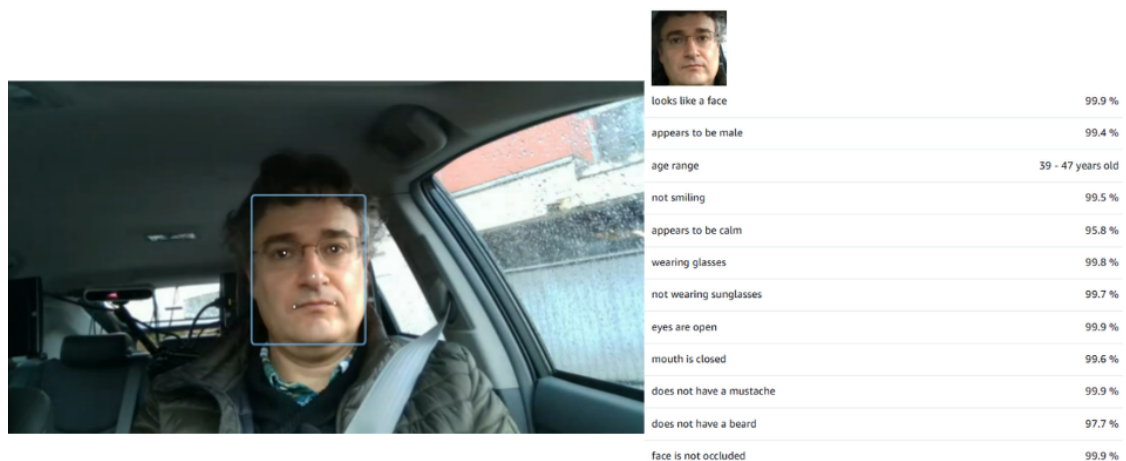


Figura 4.15: Illustrazione del comportamento di Rekognition tramite la *Face analysis* relativamente a un individuo alla guida (posizione frontale) (fonte: DMD - Driver Monitoring Dataset)

- *Analisi facciale di individui alla guida - laterale:* il secondo esperimento (Figura 4.17) viene condotto sullo stesso soggetto ma variando la posizione della ripresa; l'inquadratura si sposta lateralmente mostrando una prospettiva di profilo. Questo test mira a simulare il posizionamento della telecamera sul montante anteriore dell'abitacolo valutando la resilienza del sistema in presenza di occlusioni parziali del volto.

```

"Emotions": [
  {
    "Type": "CALM",
    "Confidence": 95.81473541259766
  },
  {
    "Type": "SAD",
    "Confidence": 2.33306884765625
  },
  {
    "Type": "CONFUSED",
    "Confidence": 0.41484832763671875
  },
  {
    "Type": "DISGUSTED",
    "Confidence": 0.05717277526855469
  },
  {
    "Type": "SURPRISED",
    "Confidence": 0.023379921913146973
  },
  {
    "Type": "FEAR",
    "Confidence": 0.01590251922607422
  },
  {
    "Type": "ANGRY",
    "Confidence": 0.012171268463134766
  },
  {
    "Type": "HAPPY",
    "Confidence": 0.00387032818980515
  }
],

```

Figura 4.16: Rappresentazione dei risultati dell'array delle emozioni rispetto al comportamento di un guidatore (inquadratura frontale), valutati tramite *Face analysis* di Amazon Rekognition

In questo test, come mostrato nei risultati in Figura 4.17, l'algoritmo di Rekognition si dimostra eccellente nell'identificazione biometrica primaria mantenendo il dato *"looks like a face"* al valore massimo del 99.9%. Tuttavia, il cambio di prospettiva si ripercuote, seppur in minima parte, sulle stime biometriche secondarie. Infatti, la confidenza nell'identificazione del genere subisce una riduzione arrivando al 95%; anche la fascia d'età stimata riceve una variazione venendo calcolata tra i 44 e i 52 anni (rispetto ai 39-47 anni del test frontale).

L'impatto più critico della variazione di prospettiva si nota nel rilevamento degli accessori: la label *"wearing glasses"* crolla al 65.3%. Per quanto riguarda lo stato di sicurezza del conducente, il microservizio conferma la sua affidabilità: le classi relative al monitoraggio dello stato di veglia, ovvero *"eyes are open"* (98.8%) e *"mouth is closed"* (95.6%) mantengono punteggi di confidenza elevati.

Relativamente all'analisi delle emozioni, come visibile in Figura 4.18, l'inquadratura laterale applica una variazione degli stati d'animo. Sebbene la classe *"CALM"* rimanga la principale, con una confidenza del 76.14%, si registra una riduzione rispetto al 95.81% del test frontale. Questo calo è affiancato ad un incremento anomalo della label *"SAD"*, che sale fino al 14.83%. Tale caso evidenzia come il cambio di inquadratura e la parziale copertura della faccia possano trarre in inganno l'algoritmo. Le restanti classi, come *"CONFUSED"* (0.48%) o *"ANGRY"* (0.18%), si confermano su valori irrilevanti. Questo risultato dimostra che, sebbene le variazioni di prospettiva riducano la precisione descrittiva dei dettagli accessori, la piattaforma riesce, comunque, a conservare una buona

accuratezza, necessaria a determinare lo stato di attenzione e d'animo del conducente anche da posizioni di ripresa non ottimali.

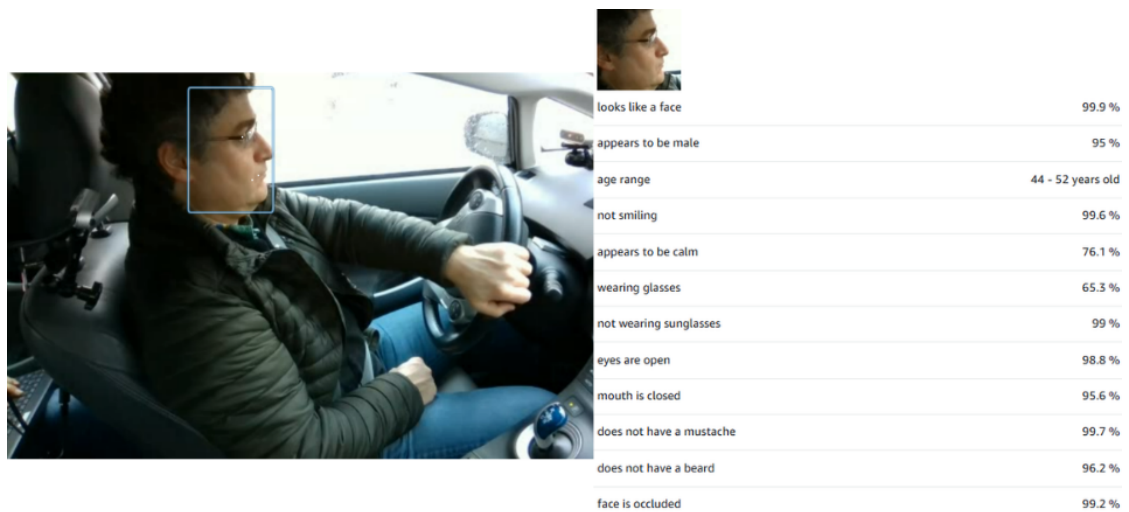


Figura 4.17: Illustrazione del comportamento di Amazon Rekognition tramite *Face analysis* + relativamente a un individuo alla guida (posizione laterale) (fonte: DMD - Driver Monitoring Dataset)

- *Analisi facciale di individui alla guida - sonnolenza:* il terzo esperimento introduce lo scenario più critico e strategico per i sistemi di sicurezza attiva in ambito automotive, ovvero il rilevamento della sonnolenza del conducente. Nell'immagine analizzata, il soggetto viene ritratto in un momento di estremo calo dell'attenzione, caratterizzato da palpebre completamente abbassate e dall'atto di sbadigliare, coprendo parzialmente il volto con la mano, come si vede in Figura 4.19.

In questo caso, l'algoritmo di Amazon Rekognition si dimostra molto efficiente nel tracciamento oculare, registrando la classe *"eyes are closed"* con un indice di confidenza quasi perfetto, pari al 99.9%. Questo dato rappresenta il parametro fondamentale per l'attivazione in tempo reale di segnali acustici o vibrazioni sul volante all'interno dell'abitacolo per contrastare un colpo di sonno imminente.⁶

Tuttavia, l'analisi degli altri parametri biometrici mette in luce alcune limitazioni algoritmiche causate dall'occlusione fisica. Il sistema fallisce nel riconoscere l'apertura della bocca a causa della presenza della mano, restituendo erroneamente la label *"mouth is closed"* con un'elevata confidenza del 95.8%.

Relativamente all'analisi delle emozioni, il risultato dell'array emotivo visibile in Figura 4.20 conferma l'impatto critico dell'occlusione del volto nei modelli *built in* di AWS. L'algoritmo classifica l'espressione del conducente prevalentemente sotto la voce *"DISGUSTED"*, con una confidenza elevata dell'86.13%. Questo evidente falso positivo deriva dall'assenza di una classe specifica dedicata alla stanchezza o alla sonnolenza nei modelli standard; infatti, il sistema interpreta erroneamente tale espressione come una manifestazione di disgusto. Le sotto-categorie secondarie evidenziano, inoltre, una lieve incertezza con la label *"CONFUSED"* al 6.06%.

⁶Questa tecnologia di sicurezza attiva è già ampiamente implementata all'interno di alcune automobili moderne, tra cui spiccano l'Audi Q8 Sportback e-tron, le Mercedes-Benz Classe C e Classe S, la BMW i5 e la Serie 7.


```

"Emotions": [
  {
    "Type": "CALM",
    "Confidence": 76.142578125
  },
  {
    "Type": "SAD",
    "Confidence": 14.83154296875
  },
  {
    "Type": "CONFUSED",
    "Confidence": 0.48224130272865295
  },
  {
    "Type": "ANGRY",
    "Confidence": 0.18568038940429688
  },
  {
    "Type": "DISGUSTED",
    "Confidence": 0.05154609680175781
  },
  {
    "Type": "SURPRISED",
    "Confidence": 0.031888484954833984
  },
  {
    "Type": "HAPPY",
    "Confidence": 0.021902721375226974
  },
  {
    "Type": "FEAR",
    "Confidence": 0.005888938903808594
  }
],

```

Figura 4.18: Rappresentazione dei risultati dell'array delle emozioni rispetto al comportamento di un guidatore (inquadratura laterale), valutati tramite *Face Analysis* di Amazon Rekognition

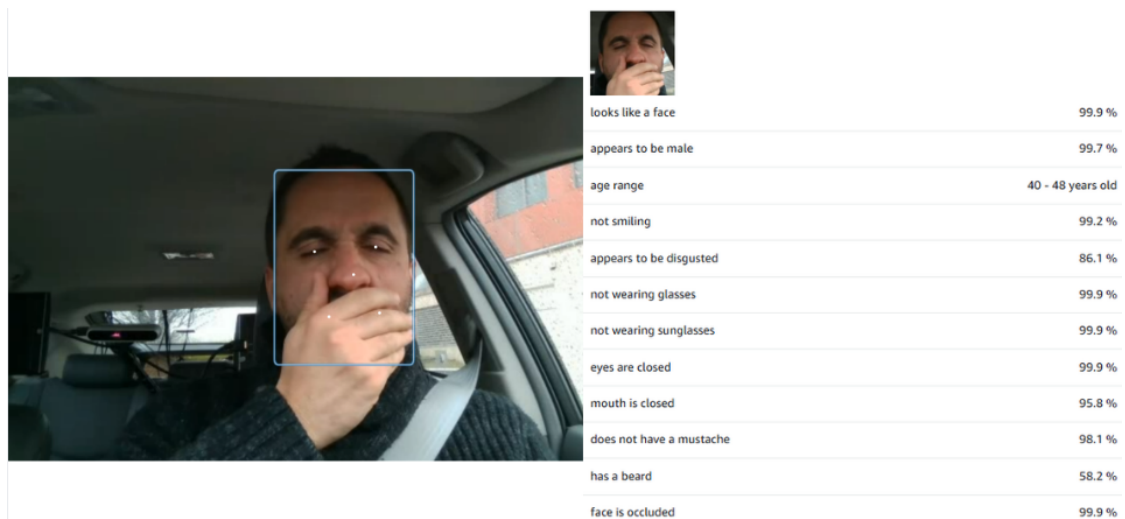


Figura 4.19: Illustrazione del comportamento di Amazon Rekognition tramite *Face analysis* relativamente a un individuo alla guida (sonnolenza) (fonte: DMD - Driver Monitoring Dataset)

4.3.3 Riconoscimento di personaggi famosi (Celebrity Recognition)

La *Celebrity Recognition* rappresenta il terzo microservizio preso in esame. Tale funzionalità si posiziona all'interno di settori di Digital Asset Management, di Influencer Marketing e del

```

"Emotions": [
  {
    "Type": "DISGUSTED",
    "Confidence": 86.1328125
  },
  {
    "Type": "CONFUSED",
    "Confidence": 6.0679121017456055
  },
  {
    "Type": "CALM",
    "Confidence": 5.1513671875
  },
  {
    "Type": "SAD",
    "Confidence": 0.7965087890625
  },
  {
    "Type": "ANGRY",
    "Confidence": 0.473785400390625
  },
  {
    "Type": "SURPRISED",
    "Confidence": 0.14972686767578125
  },
  {
    "Type": "FEAR",
    "Confidence": 0.022542476654052734
  },
  {
    "Type": "HAPPY",
    "Confidence": 0.003029902931302786
  }
],

```

Figura 4.20: Rappresentazione dei risultati dell'array delle emozioni rispetto al comportamento di un guidatore (faccia occlusa), valutati tramite *Face Analysis* di Amazon Rekognition

controllo del diritto d'autore nei media.

All'interno dell'attuale scena digitale e dei social media, monitorare manualmente l'immensa quantità di contenuti pubblicati per verificare dove e come compaia il volto di un influencer o di un personaggio famoso è un compito quasi impossibile. L'integrazione di un sistema automatizzato risponde, quindi, all'esigenza aziendale di monitorare l'efficacia delle campagne di marketing in tempo reale e, allo stesso tempo, offre uno strumento fondamentale per la tutela del brand, permettendo, tra le altre cose, di intercettare tempestivamente l'uso non autorizzato dell'immagine di figure pubbliche all'interno di profili falsi, inserzioni abusive o contenuti ingannevoli.

L'esperimento, anche in questo caso, mira a testare l'accuratezza di Amazon Rekognition, nello specifico, nel confrontare i tratti caratteristici delle immagini fornite in input con un database globale di celebrità gestito direttamente dal provider AWS. In particolare, si vuole valutare la precisione del servizio nell'estrazione del nome e cognome del soggetto, un link alla pagina biografica di riferimento e dei relativi parametri metrici di confidenza. Per esaminare la robustezza del sistema, la pianificazione delle prove è stata strutturata incrementando progressivamente il livello di difficoltà: si è partiti dall'analisi lineare di un singolo soggetto isolato per poi aumentare pian piano il numero di personaggi pubblici presenti contemporaneamente all'interno dello stesso fotogramma. Come per le sezioni precedenti, ciascun test integrerà l'output visivo elaborato assieme al risultato metrico restituito. Di seguito vengono riportati i casi studio presi in esame.

- *Popstar americana - Ariana Grande*: il primo test sperimentale della sezione analizza il

volto di Ariana Grande, popstar e influencer. Questa scelta simula lo scenario ideale per un'azienda che gestisce campagne di marketing e necessita di tracciare in tempo reale la presenza dei propri testimonial all'interno dei flussi multimediali e dei social media. Nella Figura 4.21 viene riportato l'output grafico restituito dalla piattaforma cloud, mentre, al fine di fornire un ulteriore dettaglio, nella Figura 4.22 viene riportata la parte dell'output relativa alle emozioni rilevate.

Come mostrato dall'interfaccia grafica in Figura 4.21, Amazon Rekognition isola il volto del soggetto tramite un *bounding box* e valida l'identità della celebrità con un indice di fiducia del 99.9%. Inoltre, l'array delle emozioni ("*Emotions*", Figura 4.22) offre un interessante spunto degno di nota sul volto. Di fronte a una posa non convenzionale, il sistema esclude la presenza di un sorriso ("*Value*": *false* al 93.3%) e categorizza l'espressione prevalentemente come "*SURPRISED*" (67.1%), seguita da percentuali minori di "*HAPPY*" (11.2%) e "*CALM*" (10.4%). I modelli pre-addestrati di AWS, quindi, anche in ambito *Celebrity Recognition*, forniscono dati utili per la valutazione delle emozioni.

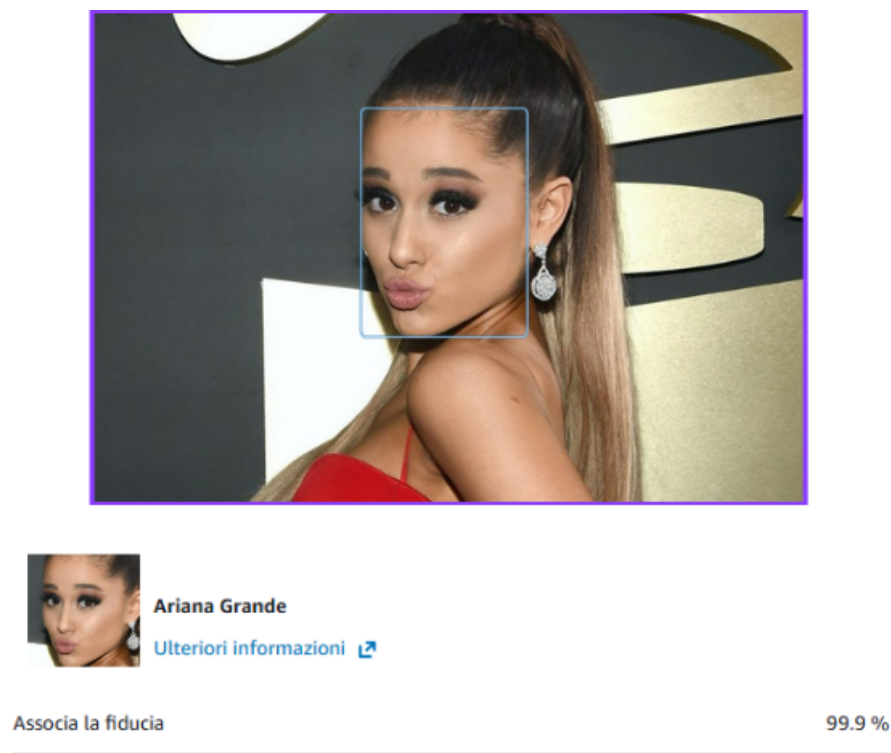


Figura 4.21: Illustrazione del comportamento di Rekognition tramite la *Celebrity Recognition* relativamente alla popstar Ariana Grande

- *Influencer - Khaby Lame e Chiara Ferragni*: il secondo test della sezione si configura tramite due esponenti dell'influencer marketing Americano e Italiano. Gli obiettivi di questa analisi sono, principalmente, due: valutare la capacità della piattaforma cloud di gestire ed elaborare due volti all'interno della stessa immagine, e allo stesso tempo, verificare il livello di aggiornamento del database di AWS rispetto a celebrità nate nei social network (TikTok e Instagram). Nella Figura 4.23 viene riportato l'output grafico generale restituito da Amazon Rekognition, mentre la Figura 4.24 mostra i dettagli dei risultati API relativi alle emozioni dei due soggetti.


```

"Emotions": [
  {
    "Type": "SURPRISED",
    "Confidence": 67.12225341796875
  },
  {
    "Type": "HAPPY",
    "Confidence": 11.263951301574707
  },
  {
    "Type": "CALM",
    "Confidence": 10.498860359191895
  },
  {
    "Type": "CONFUSED",
    "Confidence": 8.199056625366211
  },
  {
    "Type": "DISGUSTED",
    "Confidence": 0.6221771240234375
  },
  {
    "Type": "FEAR",
    "Confidence": 0.26397705078125
  },
  {
    "Type": "SAD",
    "Confidence": 0.10156631469726562
  },
  {
    "Type": "ANGRY",
    "Confidence": 0.07696151733398438
  }
],
"Smile": {
  "Value": false,
  "Confidence": 93.37238311767578
}

```

Figura 4.22: Illustrazione delle emozioni rilevate da *Celebrity Recognition* nell'immagine precedente

In questo caso, come evidenziato dal riscontro visivo, il sistema dimostra un'ottima capacità di localizzazione primaria isolando entrambi i volti tramite i rispettivi *bounding box*. Tuttavia, l'algoritmo, in entrambe le identificazioni degli influencer, fallisce producendo due falsi positivi. Il microservizio scambia infatti Khaby Lamé per il calciatore Enock Kwateng associando al match una fiducia del 75.7%, mentre Chiara Ferragni viene confusa con l'attrice francese Alice Isaaz con una fiducia del 79.4%. Questa difficoltà evidenzia un limite dei modelli built-in di AWS probabilmente riconducibile a una preferenza per figure dello sport e del cinema rispetto alle celebrità digitali, nonostante l'ampia popolarità di queste ultime.

Al contrario, l'analisi dell'espressione facciale estratta dalle API restituisce parametri coerenti con le espressioni reali dei soggetti. Per quanto riguarda Chiara Ferragni, l'algoritmo rileva come stato dominante la categoria "CALM" (44.1%), affiancata da un 12.7% di "HAPPY"; l'elemento di maggior confidenza è l'attivazione del flag "Smile", impostato su *true* con una confidenza del 74.3%, che interpreta correttamente il leggero accenno di sorriso sul volto dell'influencer.

Nel caso di Khaby Lamé, il microservizio rileva in modo ottimale la sua celebre e ironica espressione impassibile. La classe "CALM" registra una confidenza solida del 79.5% supportata dall'esclusione del sorriso ("Smile": *false* all'81.9%). È interessante notare anche l'attivazione della label "CONFUSED" all'11.2%: la contrazione delle sopracciglia tipica dell'espressione comunicativa celebre del creator viene tradotta dal sistema come uno stato di confusione, confermando la sensibilità dell'algoritmo nel cogliere anche le più piccole espressioni.

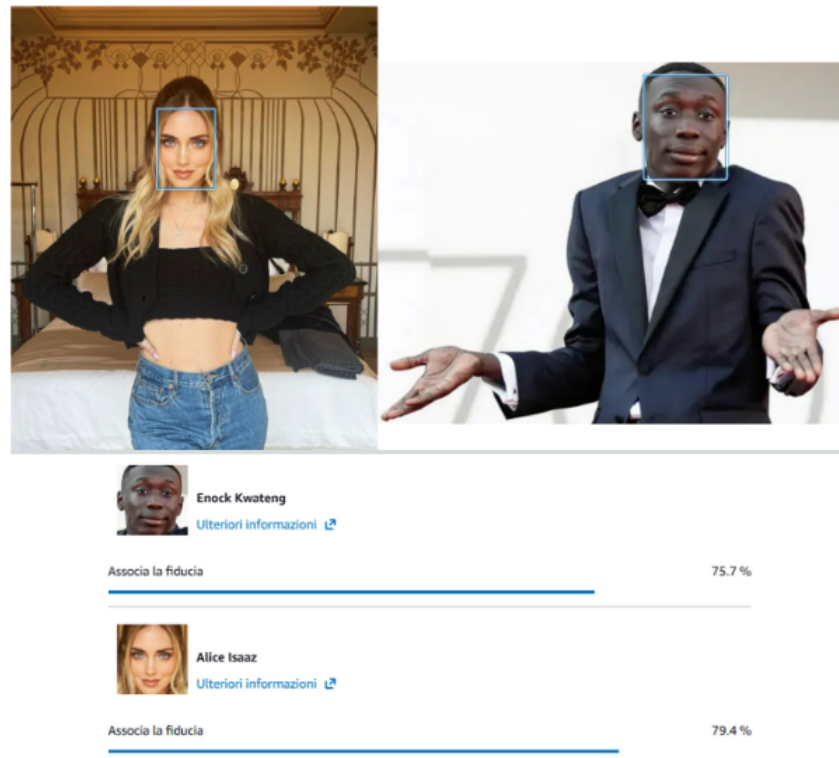


Figura 4.23: Illustrazione del comportamento di Amazon Rekognition tramite *Celebrity Recognition* relativamente agli influencer Chiara Ferragni e Khaby Lame

Chiara Ferragni

```
{
  "Type": "CALM",
  "Confidence": 44.18526848289961
},
{
  "Type": "HAPPY",
  "Confidence": 12.742746353149414
},
{
  "Type": "SURPRISED",
  "Confidence": 0.6532669067382812
},
{
  "Type": "SAD",
  "Confidence": 0.5970801220703125
},
{
  "Type": "DISGUSTED",
  "Confidence": 0.171661376953125
},
{
  "Type": "CONFUSED",
  "Confidence": 0.14646849036216736
},
{
  "Type": "ANGRY",
  "Confidence": 0.06413459777832031
},
{
  "Type": "FEAR",
  "Confidence": 0.029897689819335938
}
},
"Smile": {
  "Value": true,
  "Confidence": 74.31609344482422
}
```

Khaby Lame

```
{
  "Type": "CALM",
  "Confidence": 79.568546875
},
{
  "Type": "CONFUSED",
  "Confidence": 11.2088078125
},
{
  "Type": "SAD",
  "Confidence": 0.37212371826171875
},
{
  "Type": "HAPPY",
  "Confidence": 0.3194172978401184
},
{
  "Type": "SURPRISED",
  "Confidence": 0.28228759765625
},
{
  "Type": "DISGUSTED",
  "Confidence": 0.24127968205078125
},
{
  "Type": "FEAR",
  "Confidence": 0.04818439483642578
},
{
  "Type": "ANGRY",
  "Confidence": 0.0088523464202888859
}
},
"Smile": {
  "Value": false,
  "Confidence": 81.93734741218938
}
```

Figura 4.24: Illustrazione delle emozioni rilevate da *Celebrity Recognition* nell'immagine precedente

- *Confronto collettivo - gruppo di attori internazionali e nazionali*: il terzo esperimento rappresenta un'analisi collettiva con un numero elevato di soggetti. In questo specifico caso, per preservare la fluidità del testo, si è scelto di omettere i risultati delle API. Tale decisione deriva dal fatto che l'obiettivo primario di questo microservizio è rappresentato dalla validazione dell'identità delle celebrità e non nell'analisi emozionale dei soggetti. Inoltre, l'inclusione di stringhe di codice ripetute per un numero così elevato di individui avrebbe appesantito la trattazione senza valorizzarla. Nella Figura 4.25 vengono rappresentati i soggetti con i relativi *bounding box*, mentre nella Figura 4.26 vengono riportati i rilevamenti della piattaforma cloud.



Figura 4.25: Illustrazione di un gruppo di attori internazionali e nazionali sottoposti all'analisi di *Celebrity Recognition*

Nel primo scenario, relativo al cast della serie statunitense *"Friends"* (Figura 4.25 gruppo superiore), il sistema dimostra, ancora una volta, un'eccellente capacità di isolamento geometrico primario applicando correttamente i *bounding box* a tutti i volti presenti nell'immagine. Per quanto riguarda l'identificazione, l'algoritmo si rivela quasi infallibile per le star hollywoodiane di tale portata, confermando, come visibile nella parte sinistra della Figura 4.26, l'identità di Matthew Perry, David Schwimmer, Jennifer Aniston e Matt LeBlanc con un indice di fiducia assoluto del 99.9%, e Courteney Cox al 99.6%. Si registra, tuttavia, un unico falso positivo nell'ultimo soggetto a destra: l'attrice Lisa Kudrow viene scambiata per l'artista francese Ophélie Winter con una confidenza comunque elevata dell'89.5%. Questo errore evidenzia come la parziale occlusione dei tratti caratteristici del volto possano indurre i modelli biometrici pre-addestrati a generare collegamenti errati basati su piccole somiglianze presenti nel database di AWS.

Il secondo scenario, relativo al gruppo di attori italiani (Figura 4.25 gruppo inferiore), offre un riscontro sperimentale che contraddice le limitazioni riscontrate nella precedente casistica degli influencer. Amazon Rekognition riesce non solo a mappare simultaneamente tutti e sette i volti allineati nell'immagine, ma identifica anche ciascun attore restituendo per ognuno di essi un indice di fiducia costante e massimo pari al 99.9%, come risulta visibile nella parte destra della Figura 4.26

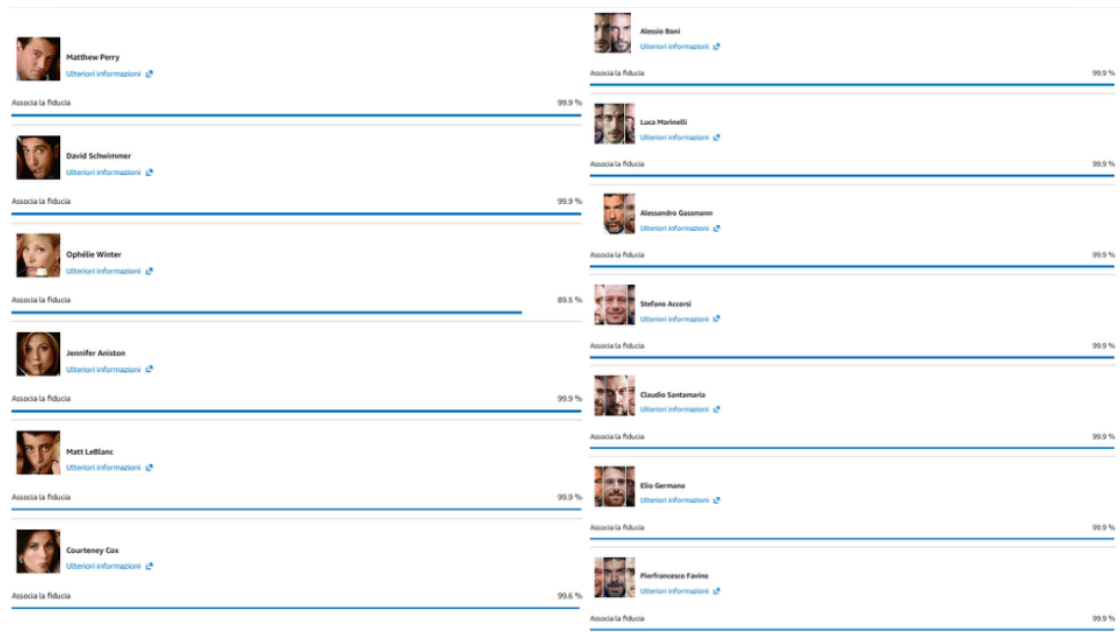


Figura 4.26: Illustrazione del comportamento del microservizio *Celebrity Analysis* in merito a un gruppo di attori nazionali e a un gruppo di attori internazionali

4.3.4 Videosorveglianza smart: Analisi temporale e tracciamento di intrusi

L'ultimo microservizio analizzato sposta il focus dell'analisi dall'elaborazione di immagini all'analisi di flussi video, un ambito che assume un'importanza elevata per la sicurezza fisica aziendale e per la tutela dei lavoratori. All'interno di contesti aziendali o logistici, il monitoraggio tradizionale delle telecamere, affidato esclusivamente al personale di vigilanza, mostra forti limiti legati all'inevitabile calo di attenzione al cospetto di decine di schermi attivi 24 ore su 24. Applicare le funzionalità video di Amazon Rekognition in questo settore significa automatizzare la sorveglianza perimetrale antifurto e quella interna per la sicurezza dei lavoratori riducendo i tempi di reazione del personale adibito al monitoraggio, e permettendo l'identificazione di accessi non autorizzati o situazioni di pericolo.

L'esperimento relativo a questa sezione introduce la variabile dell'analisi temporale mirando a misurare la capacità dell'algorithmo di effettuare il tracciamento continuo di un soggetto in movimento. In particolare, l'obiettivo è verificare la precisione dell'algorithmo nell'assegnare un'etichetta al soggetto in moto e nel garantirne la persistenza temporale per l'intera durata della sua permanenza all'interno dell'inquadratura.

In quanto meno immediata e user friendly rispetto a quelle degli altri microservizi basati su singole immagini, risulta necessaria una descrizione dell'interfaccia utente dedicata all'analisi dei flussi video. Come mostrato in Figura 4.28, la piattaforma cloud adotta un approccio differente per l'elaborazione dei contenuti video: non vengono applicati riquadri grafici in sovrapposizione al video in tempo reale, ma i risultati estratti lungo l'intera timeline vengono uniti all'interno di un pannello di controllo laterale.

L'interfaccia organizza i dati in categorie: nella parte superiore sinistra, la sezione "*Celebrità*" contiene i volti delle personalità note identificate durante la riproduzione (nel caso di esempio in Figura 4.28, il fondatore di Amazon Jeff Bezos e il suo interlocutore). Subito sotto, il blocco "*Oggetti e attività*" raccoglie ciò che risulta visibile nel contenuto multimediale tramite etichette. L'algorithmo non si limita a identificare i singoli elementi fisici presenti (come "*Suit*", "*Coat*" o "*Monitor*"), ma riconosce la natura dell'evento in corso estraendo concetti astratti quali "*Press Conference*" e "*Photography*". L'interfaccia consente, inoltre, di

selezionare le singole etichette per esaminare, attraverso una barra cronologica arancione (dettagliata in Figura 4.27), la durata esatta e i precisi intervalli temporali in cui le specifiche entità compaiono all'interno del flusso video.



Figura 4.27: Illustrazione del dettaglio della comparsa delle specifiche entità nel video

Inoltre, il microservizio integra un sistema di filtraggio automatico volto a segnalare l'eventuale presenza di *"Etichette moderate"*, informando l'utente qualora il flusso video contenga scene esplicite, violente o non conformi alle policy aziendali. Poiché l'interfaccia grafica offre esclusivamente un riassunto visuale semplificato, i dati dettagliati risiedono nel pulsante *"Scarica la risposta come file JSON"*. È proprio all'interno di questo insieme di dati testuale che il sistema mappa analiticamente i *timestamp* di ogni frazione di secondo consentendo agli sviluppatori di registrare l'esatto momento di ingresso, movimento e uscita di un'entità dalla scena.

Tuttavia, data l'elevata quantità e densità di informazioni contenute nel risultato JSON completo, in questa trattazione verranno commentati i risultati strutturati tramite l'interfaccia grafica della console; tale scelta permette di preservare la leggibilità del capitolo, offrendo comunque, una panoramica esaustiva dei parametri estratti dalla piattaforma cloud.

Esposto il funzionamento della console, il test sperimentale vero e proprio è stato condotto su una clip video reale estratta dalla piattaforma YouTube raffigurante scenari critici di tentativi di furto e intrusioni perimetrali. Al fine di garantire il rispetto delle normative relative alla protezione dei dati personali e tutela della privacy, è stato opportunamente selezionato un filmato in cui i volti dei soggetti coinvolti non risultano in alcun modo visibili o identificabili. Questa scelta permette di testare l'algoritmo cloud su un flusso video reale caratterizzato dalla tipica riduzione di qualità dovuta alla trasmissione in rete rimanendo, al contempo, conformi con i requisiti etici della ricerca.

Al fine di ottimizzare la fluidità della trattazione, i risultati del test vengono esposti attraverso una serie di fotogrammi estratti in ordine temporale (corrispondenti all'inizio, a metà e al termine della clip).

Il primo fotogramma estratto dall'esperimento (Figura 4.29) si colloca al timestamp 0:07 della clip e introduce uno scenario di estrema complessità visiva per i modelli di Computer Vision. Risulta opportuno precisare che l'analisi inizia solo a partire dal settimo secondo poiché la parte iniziale del filmato costituisce un'introduzione priva di frame raffiguranti il

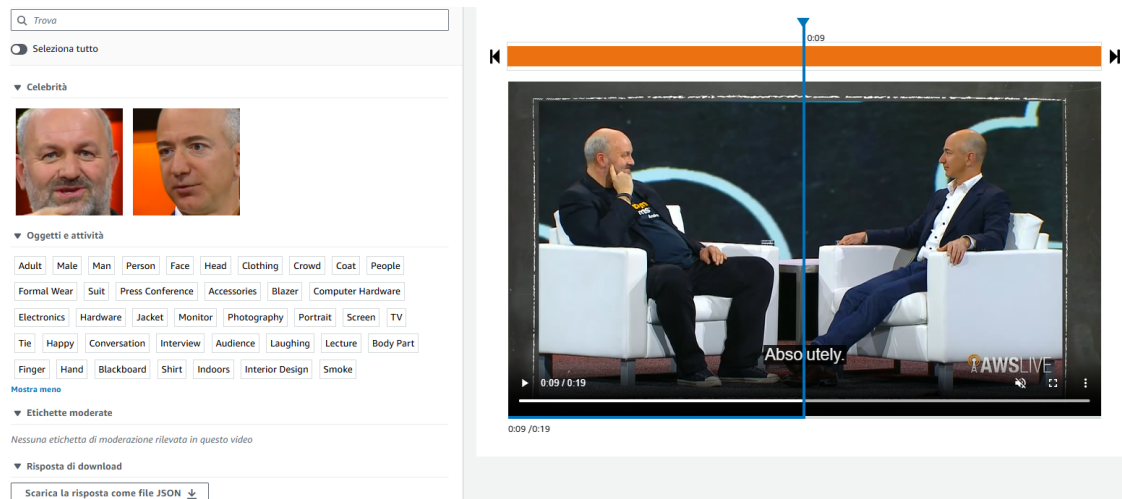


Figura 4.28: Illustrazione dell'interfaccia grafica del microservizio *Analisi dei video archiviati* di Amazon Rekognition

luogo. La scena è ripresa da una telecamera di sicurezza in modalità notturna caratterizzata da una scarsa illuminazione ambientale, da una bassa risoluzione e da elementi di disturbo dati dall'ambiente circostante. In questo contesto, l'intruso cerchiato di rosso in foto si trova posizionato sul margine destro del frame parzialmente fuso con l'oscurità dello sfondo e con la tettoia presente nel determinato luogo risultando quasi invisibile a uno sguardo superficiale da parte di un operatore umano. Al fine di testare la reattività del sistema in questa frazione temporale critica, sono state selezionate le etichette "Person", "Head", "Adult", "Male", "Man", "Female" e "Woman", impostando, così, un controllo volto alla rilevazione di intrusi per verificare esattamente in quale momento venisse rilevata la persona.

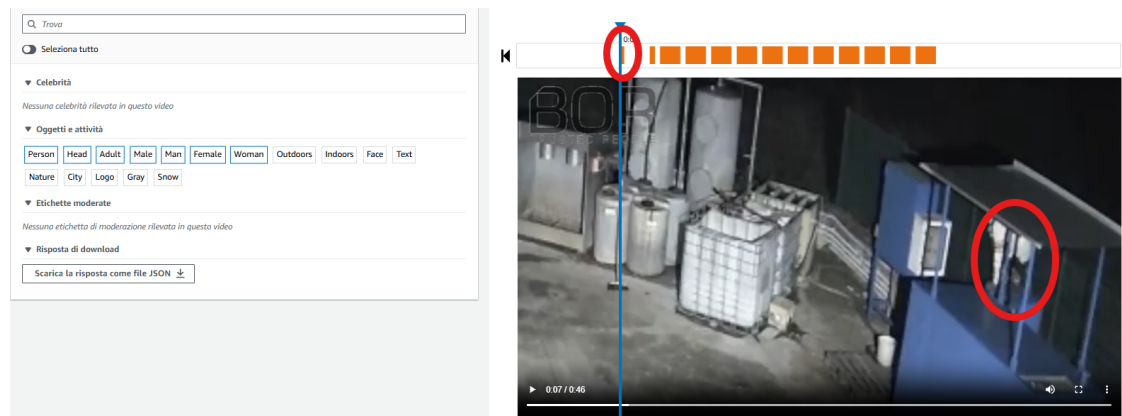


Figura 4.29: Illustrazione delle rilevazioni del microservizio *Analisi dei video archiviati* di Amazon Rekognition in merito al primo frame video fornito ad esso (fonte: youtube.com - canale: Bor Security)

Il secondo fotogramma analizzato (Figura 4.30) è preso dal secondo 0:21; esso rappresenta la fase intermedia dell'esperimento e mostra l'evoluzione dell'azione. In questa frazione temporale il soggetto si è spostato dal margine perimetrale destro verso il centro dell'area monitorata posizionandosi in prossimità del contenitore bianco. Questo spostamento offre un modo per valutare la stabilità del tracciamento video in condizioni di movimento attivo. Come si vede dall'interfaccia grafica la barra cronologica superiore presenta una successione continua di blocchi arancioni che precedono e seguono il cursore blu. Tale evidenza dimostra

la rilevazione persistente dell’algoritmo: Amazon Rekognition non frammenta l’analisi in singoli rilevamenti isolati, ma riconosce l’identità e la continuità dell’entità che si muove nello spazio.

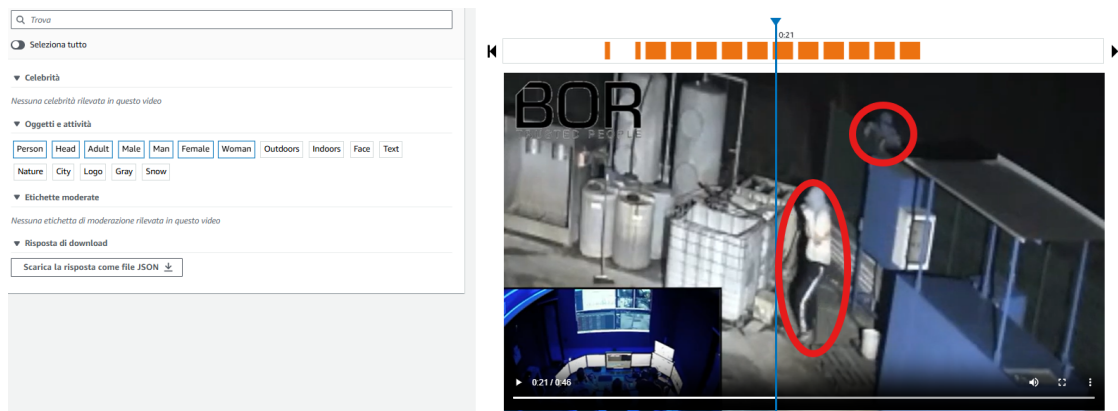


Figura 4.30: Illustrazione delle rilevazioni del microservizio *Analisi dei video archiviati* di Amazon Rekognition in merito al secondo frame video fornito ad esso (fonte: youtube.com - canale: Bor Security)

Il terzo e ultimo fotogramma (Figura 4.31) si colloca al secondo 0:32 e illustra la conclusione dell’evento critico mostrando il momento esatto in cui i soggetti intrusi hanno abbandonato l’area monitorata. In questa frazione temporale, la scena si presenta completamente vuota e priva di movimento.

Questo scatto finale offre un riscontro di fondamentale importanza per valutare la reattività e la precisione di Amazon Rekognition Video nel determinare la fine di un’anomalia perimetrale. Come si osserva chiaramente dall’interfaccia grafica, in corrispondenza con il posizionamento del cursore blu al secondo 0:32, la successione dei blocchi arancione sulla barra cronologica si interrompe lasciando lo spazio successivo della timeline totalmente bianco.

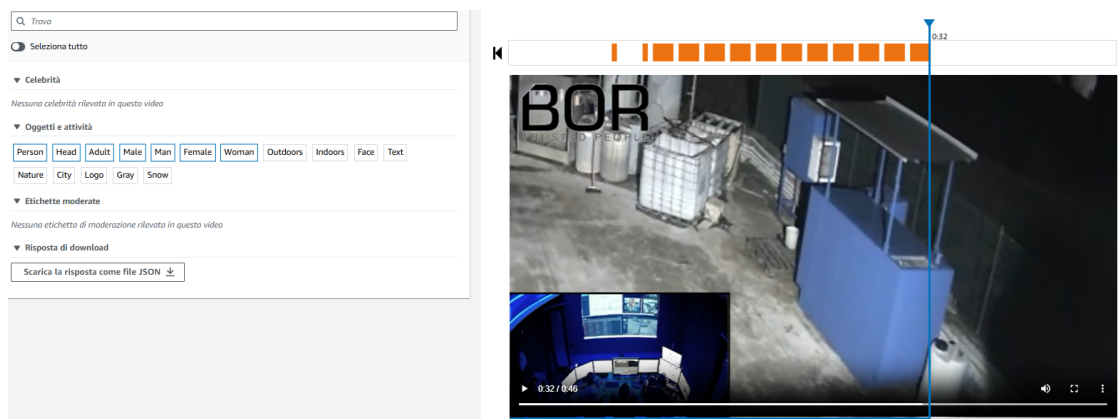


Figura 4.31: Illustrazione delle rilevazioni del microservizio *Analisi dei video archiviati* di Amazon Rekognition in merito al terzo frame video fornito ad esso (fonte: youtube.com - canale: Bor Security)

4.3.5 Analisi qualitativa

L’analisi qualitativa dei test eseguiti su Amazon Rekognition ha evidenziato la risposta dei modelli *built-in* in base a geometrie visive, affollamento e condizioni ambientali.

Per quanto riguarda la *Label Detection*, l'accuratezza dipende dalla densità e dalle geometrie della scena. Nei supermercati l'alta densità visiva causa inefficienze e falsi positivi critici, come lo scambio di snack per attrezzatura medica ("*First Aid*" al 95.5%). Al contrario, in ambito aeroportuale, la presenza di entità grandi e isolate garantisce risultati soddisfacenti (superiori al 95%), dimostrando resilienza anche nell'oscurità notturna.

Relativamente alla *Face Detection and Analysis* (dataset DMD), lo strumento è efficiente nel tracciamento dello stato degli occhi e della bocca (99.9%) come supporto salvavita. Tuttavia, risente di cambi di prospettiva e degli ostacoli: nei test di profilo il rilevamento degli occhiali crolla al 65.3%, mentre la mano davanti alla bocca genera il falso positivo "*mouth is closed*" (95.8%) e la classificazione errata dell'emozione "*DISGUSTED*" (86.13%).

Nell'ambito della *Celebrity Recognition*, viene rilevata un'asimmetria nel database globale di AWS. Il sistema è perfetto (99.9%) con star globali (Ariana Grande) e scene di cast cinematografici con molteplici soggetti. Esso fallisce invece, completamente con le figure nate sui social media: Chiara Ferragni e Khaby Lame generano falsi positivi venendo confusi, rispettivamente, con un'attrice francese e un calciatore.

Infine l'*analisi dei video*, sfrutta la barra cronologica arancione per registrare gli intervalli di comparsa delle entità. Nel test, il perfetto sincronismo tra l'ingresso del soggetto (0:07) e la sua uscita dalla scena (0:32) dimostra la stabilità del tracciamento temporale nel rilevare la durata reale di un'anomalia perimetrale anche in notturna.

4.4 Punti di forza e limitazioni riscontrate

Eseguita l'analisi delle funzionalità e dei risultati dei quattro microservizi, è possibile riassumere i vantaggi e le limitazioni riscontrate nell'utilizzo di Amazon Rekognition.

4.4.1 Punti di forza riscontrati

Amazon Rekognition ha dimostrato una serie di funzionalità altamente efficienti per l'integrazione della Computer Vision in contesti applicativi reali. I principali punti di forza riscontrati includono:

- *Interfaccia user friendly*: tutti i microservizi SaaS offerti da Amazon Rekognition, in particolare i quattro moduli analizzati, mettono a disposizione strumenti estremamente intuitivi, utilizzabili efficacemente anche da utenti inesperti nel settore del *Machine Learning*.
- *Estrazione eccellente delle caratteristiche biometriche*: nei test condotti sul monitoraggio del conducente, il modulo di *Face Analysis* ha dimostrato la capacità di estrarre dati fondamentali. Il sistema riesce a evidenziare i punti critici del volto calcolando le possibili emozioni provate dall'autista. Questo approccio permette lo sviluppo di software salvavita in ambito automobilistico aumentando la sicurezza sulle strade.
- *Robustezza e stabilità in condizioni ambientali critiche*: l'analisi sperimentale ha evidenziato un'eccellente resilienza della *Label Detection* e dell'analisi video di fronte a stress ambientali, come il passaggio da riprese diurne a riprese notturne, contrariamente a quanto riportato relativamente all'illuminazione nella Sezione 4.1.2. La capacità dei modelli *built-in* di dimostrare accuratezza nell'isolamento delle etichette anche in condizioni di scarsa illuminazione risulta fondamentale per l'applicazione del servizio in sistemi operativi adibiti alla sicurezza.
- *Output strutturato in formato JSON*: tutti i microservizi esaminati convertono i dati visivi non strutturati in risposte testuali in formato JSON, aggiungendo, tra le altre cose,

metadati ulteriori utili all'analisi (come le emozioni nel caso della *Face Analysis* e nella *Celebrity Recognition*). Questo può essere visto come un punto di forza: la presenza di una tale struttura dati chiara favorisce l'utilizzo della piattaforma cloud attraverso qualsiasi linguaggio di programmazione, semplificando lo sviluppo di software automatizzati.

- *Scalabilità tramite Custom Labels*: l'architettura dei microservizi non vincola l'utente all'utilizzo dei soli modelli *built-in*. Il servizio prevede, infatti, la possibilità di estendere il sistema nativo introducendo etichette personalizzate e specializzate senza dover riprogettare la logica software.

4.4.2 Limitazioni riscontrate

Nonostante gli evidenti vantaggi, sono state messe in luce anche alcune limitazioni di Amazon Rekognition, quali richiedono attenzione in fase di progettazione software. Tali limitazioni sono:

- *Sensibilità alla densità visiva*: confermando quanto detto sulle prospettive nella Sezione 4.1.2, i modelli *built-in* si dimostrano limitati in presenza di un'elevata densità di oggetti. Senza un contesto preciso l'algoritmo opera solo in base alle geometrie e ai colori. Di conseguenza, la vicinanza degli elementi o entità di colore simile può metterlo in difficoltà.
- *Vulnerabilità geometrica ai cambi prospettici e alle occlusioni*: l'efficacia dei modelli di *Face Analysis* risulta dipendente dall'inquadratura del soggetto. Variazioni significative della posizione (come i profili) o la presenza di elementi di disturbo causano una minor efficienza nell'accuratezza del rilevamento. Il sistema risulta progettato per forzare una classificazione anche su dati visivi parziali, introducendo delle inesattezze nei risultati finali.
- *Predilezione del dataset per ambiti convenzionali*: il servizio di *Celebrity Recognition built-in* mostra difficoltà nel riconoscere figure nate sui social media creando una conseguente limitazione relativa all'affidabilità in contesti prevalentemente digitali.

Un case study nel contesto dello Speech-To-Text

Il presente capitolo introduce lo Speech-To-Text come una tecnologia che funge da ponte tra la comunicazione umana e i sistemi digitali, consentendo la conversione automatica del linguaggio parlato in testo scritto. Questa disciplina, pilastro del Natural Language Processing, trova oggi applicazione in vari contesti, dalla trascrizione analitica dei flussi audio nei sistemi aziendali fino ai servizi di documentazione clinica. Il capitolo si pone due principali obiettivi: fornire una panoramica degli aspetti teorici di questa tecnologia e, parallelamente, valutarne l'efficacia attraverso un'analisi sperimentale. Nella prima parte dell'elaborato verranno presentate la definizione, i fondamenti tecnici e i principali casi d'uso attuali. Successivamente, verrà illustrata la metodologia utilizzata per implementare e integrare questa tecnologia attraverso la piattaforma cloud Amazon Transcribe. La sezione principale del capitolo sarà rappresentata da un case study volto a testare le funzionalità chiave del servizio, con l'obiettivo di estrapolarne i principali punti di forza e le relative limitazioni riscontrate.

5.1 Speech-To-Text

Nella società odierna la trascrizione precisa e automatica dei contenuti rappresenta un fattore di supporto sotto molti aspetti. Questa tecnologia è, infatti, oggi indispensabile per poter garantire accessibilità, efficienza, sicurezza e automazione. L'esempio più lampante è il supporto all'inclusione: essa permette, da sempre, a persone con limitazioni motorie o uditive di utilizzare un computer o uno smartphone senza la necessità di avere una tastiera. Al contempo, permette anche a medici, tecnici e operai di compilare rapporti a voce mantenendo le mani libere per operare.

In questo ambito, lo *Speech-To-Text* funge da acceleratore delle attività quotidiane, applicando tecniche di *Natural Language Processing* al fine di raggiungere la sua massima efficienza per recepire e trascrivere analiticamente ciò che da una sorgente input riesce a captare. In questo modo, il sistema ottimizza i flussi di lavoro, riducendo i tempi di inserimento dei dati e azzerando i margini di errore manuale.

A differenza delle altre due tecnologie, lo *Speech-To-Text* è probabilmente la soluzione che offre il supporto più concreto, distinguendosi per un utilizzo immediato e accessibile alla maggior parte della popolazione mondiale. Senza questa tecnologia, la vita quotidiana di numerose persone risulterebbe fortemente penalizzata; ciò ricalca la sua assoluta indispensabilità.

In questo capitolo, lo *Speech-To-Text* rappresenta, quindi, il fondamento teorico necessario da interiorizzare per comprendere a pieno il case study basato su Amazon Transcribe. Tale

introduzione mira a fornire un quadro chiaro del contesto, offrendo al lettore le modalità per valutare sia i punti di forza, sia i limiti della piattaforma.

5.1.1 Definizione di Speech-To-Text

Lo *Speech-To-Text* può essere definito come il processo computazionale di conversione delle parole pronunciate in trascrizione testuale. Tale convenzione viene messa in atto tramite la combinazione del riconoscimento vocale basato sull'Intelligenza Artificiale e la trascrizione. Esso funge, inoltre, da ponte tra l'essere umano e la macchina, consentendo a quest'ultima di elaborare, analizzare e archiviare informazioni.

La Figura 5.1 illustra il principio di funzionamento generale del sistema *Speech-To-Text*.

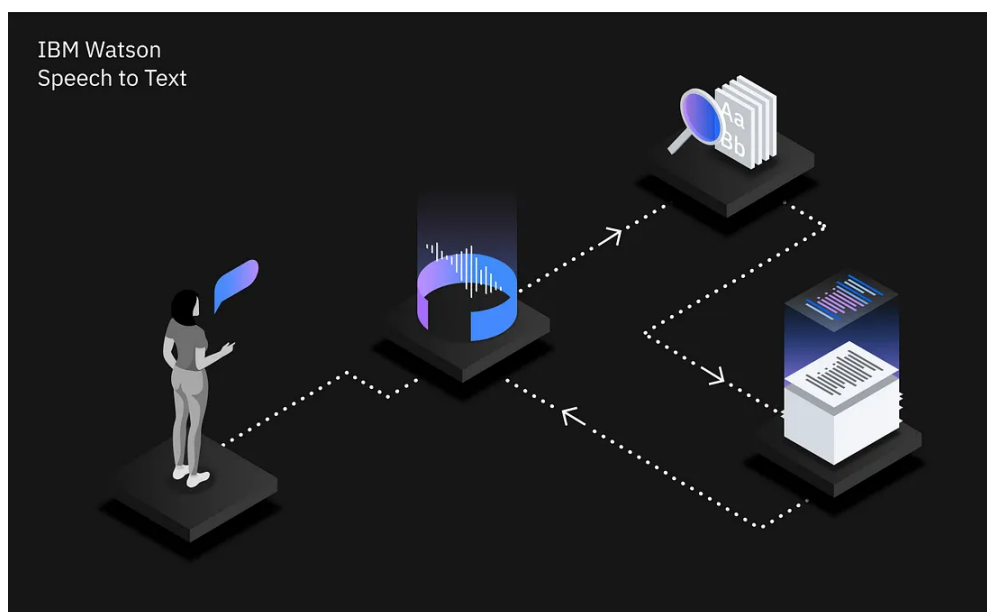


Figura 5.1: Illustrazione del principio generale di funzionamento dello *Speech-To-Text* (fonte: medium.com)

Nella Figura 5.1 è possibile comprendere come il soggetto, parlando attraverso una sorgente audio, permette il passaggio del flusso alla piattaforma centrale (icona con l'onda sonora) che analizza le frequenze. Successivamente, il software confronta l'audio con modelli grammaticali e dizionari (icona della lente d'ingrandimento) per effettuarne la conversione in testo digitale leggibile (icona dei testi digitali); infine, la freccia di ritorno verso la piattaforma centrale evidenzia il processo di apprendimento continuo del sistema, che, correggendo i propri errori, ottimizza le trascrizioni future.

5.1.2 Aspetti tecnici dello Speech-To-Text

Un software che implementa lo *Speech-To-Text* utilizza una serie di componenti:

- *input vocale*: una sorgente audio che cattura la voce e le parole pronunciate;
- *estrazione delle caratteristiche*: un algoritmo che identifica toni e schemi caratteristici all'interno del discorso;
- *decodificatore*: il sistema abbina le caratteristiche del parlato ai caratteri e alle parole attraverso un modello linguistico, sfruttando la propria natura autoregressiva per predire l'elemento successivo basandosi sul contesto precedentemente generato;

- *output*: in questo ultimo componente, il testo finale viene formattato in modo ottimale così che il trascritto risulti leggibile a un essere umano.

La Figura 5.2 illustra i passaggi sopra menzionati per realizzare lo *Speech-To-Text*.

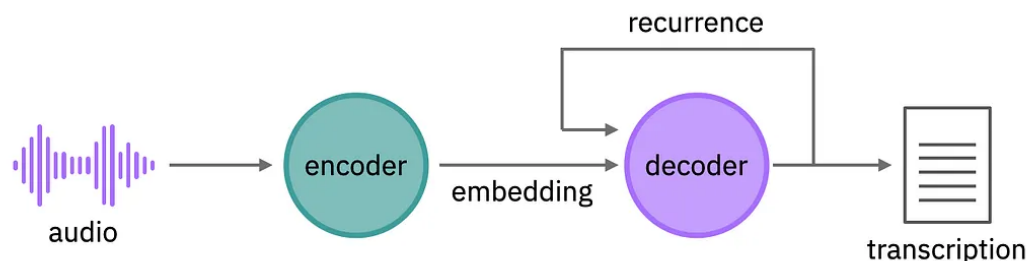


Figura 5.2: Illustrazione del flusso atto a realizzare lo *Speech-To-Text* (fonte: medium.com)

Il pilastro fondamentale dello *Speech-To-Text* è rappresentato dalle reti neurali profonde. Come già illustrato all'interno dei capitoli precedenti, esse non nascono già pronte, ma vengono bensì addestrate su un ampio set di flussi audio, in modo tale da migliorare le loro capacità di riconoscimento nei contesti in cui vengono applicate.

Scendendo nel dettaglio delle tecnologie utilizzate, è necessario nominare i vari algoritmi esistenti in merito: si parte dai sistemi di *riconoscimento automatico del parlato*, chiamati ASR, che sfruttano modelli di linguaggio statistici per identificare e scomporre i fonemi, confrontandoli, allo stesso tempo, con un ampio database linguistico, calcolando la trascrizione più probabile. A questi si affiancano le reti neurali profonde vere e proprie, che, sfruttando strati di neuroni artificiali interconnessi, comprendono in completa autonomia le sfumature della voce e del linguaggio, riuscendo ad offrire risultati più accurati rispetto ai sistemi ASR. Infine, il panorama tecnologico si completa con modelli matematici specifici che gestiscono la voce e il testo. Tra questi si nominano i Modelli di Markov Nascosti, utili per monitorare la sequenza dei suoni riprodotti nel tempo, e i Campi Aleatori Condizionali, che servono, invece, a prevedere la corretta successione delle parole all'interno della frase.

5.1.3 Applicazioni attuali dello Speech-To-Text

Questa tecnologia al giorno d'oggi è integrata nella vita quotidiana e in contesti lavorativi importanti; le applicazioni principali si dividono in tre aree:

- *Generazione di sottotitoli*: lo *Speech-To-Text* è ormai ovunque. Permette di generare sottotitoli automatici sulle piattaforme di streaming e non (aiutando non udenti o chi non parla la lingua), e consente di dettare messaggi ed email sugli smartphone per comunicare più velocemente.
- *Automazione aziendale*: trasforma in testo le registrazioni di riunioni, di interviste e di telefonate dei call center, rendendo le informazioni facili da archiviare e cercare. Inoltre, permette il funzionamento di assistenti vocali e chatbot per l'assistenza ai clienti, facilitando anche lo smistamento automatico delle telefonate in arrivo verso i reparti corretti per ridurre i tempi di attesa.
- *Settore sanitario e medico*: negli ospedali, i medici possono dettare diagnosi e prescrizioni a voce, evitando di perdere tempo a digitare sulla tastiera; in questo modo, possono dedicare tutta la loro attenzione al paziente durante la visita. Il sistema, inoltre, riduce gli errori manuali di trascrizione e crea cartelle cliniche digitali facili da consultare e catalogare.

5.2 Contesto e metodologia per attuare lo Speech-To-Text

In ottica di sviluppo del case study sullo *Speech-To-Text*, la scelta dello strumento è ricaduta su *Amazon Transcribe*. Questo servizio della suite AWS, già introdotto nella Sezione 2.5.3, si basa principalmente sulle Reti Neurali Profonde, inserite all'interno di un sistema complessivo di *Riconoscimento Automatico del Parlato (ASR)*, offrendo l'opportunità di implementare concretamente gli scenari d'uso precedentemente descritti. Per verificare la robustezza dell'algoritmo, il sistema è stato sottoposto a uno *stress test* linguistico mirato alla valutazione di molteplici tipologie di dialetti. Successivamente, al fine di valutarne l'efficacia nella risoluzione di problematiche concrete, la tecnologia è stata applicata all'analisi e alla trascrizione di file audio provenienti dai servizi di assistenza clienti.

La metodologia utilizzata per affrontare le casistiche sopra riportate, analogamente a quanto visto per i due case study precedenti, sarà la seguente:

1. identificazione dei requisiti aziendali;
2. pre-elaborazione e descrizione strutturale dei dati;
3. estrazione e successiva manipolazione dei dati in uscita;
4. valutazione critica e interpretazione qualitativa delle metriche emerse;
5. discussione dei vantaggi competitivi e dei vincoli della tecnologia applicata.

Nella prossima sezione verrà mostrata la configurazione ottimale necessaria per sfruttare al meglio le sue funzionalità all'interno delle applicazioni menzionate.

5.2.1 Implementazione con Amazon Transcribe

In analogia con i due *case study* precedenti, l'accesso al servizio avviene tramite la *AWS Management Console*, digitando, nella barra di ricerca, la voce "*Amazon Transcribe*". Allo stesso modo di Amazon Rekognition, l'interfaccia di questo servizio offre una struttura di navigazione laterale articolata, organizzata in quattro aree che mostrano le varie specializzazioni dei modelli predittivi qui sotto riportati:

- *Amazon Transcribe*: rappresenta l'applicazione standard del servizio. Oltre a consentire la trascrizione in tempo reale generale, introduce tre importanti strumenti di ottimizzazione algoritmica: il *modello linguistico personalizzato*, il *vocabolario personalizzato* e il *filtraggio del vocabolario*. Questi moduli permettono all'utente di addestrare il sistema nel riconoscimento di termini tecnici e acronimi aziendali, eliminando, così, gli errori di trascrizione che potrebbero avvenire durante il processo.
- *Analisi delle chiamate Amazon Transcribe*: una sezione pensata per i contesti di assistenza ai clienti. Si divide in *analisi in tempo reale* e *analisi post-chiamata*, integrando logiche di *Natural Language Processing* che, oltre alla sola conversione audio-testo, estraggono dati sul *sentiment* del cliente, sui tempi di silenzio e sulle *categorie* per la classificazione automatica delle telefonate.
- *Amazon Transcribe Medical*: è un modulo specializzato e pre-addestrato con dati di tipo sanitario. Offre funzionalità simili al servizio standard, quali *trascrizione in tempo reale*, *processi di trascrizione* e *vocabolario personalizzato*, ma studiate appositamente per riconoscere con precisione la terminologia medica, i nomi dei farmaci e le patologie, riducendo i tassi di errore tipici dei modelli in contesti ospedalieri e farmaceutici.

- **AWS HealthScribe**: rappresenta l'evoluzione più recente del servizio basata su Intelligenza Artificiale Generativa. Attraverso le sezioni di *approfondimenti medici* e *processi di trascrizione*, non si limita a trascrivere il dialogo tra medico e paziente, ma genera automaticamente riassunti clinici, cartelle cliniche preliminari e dati pronti per essere inseriti nei sistemi sanitari.

La Figura 5.3 mostra l'interfaccia che l'utente dovrebbe visualizzare avendo seguito i passaggi sopra spiegati.



Figura 5.3: Illustrazione dell'interfaccia di Amazon Transcribe

Al fine di garantire la massima pertinenza scientifica ed evitare di rendere il capitolo eccessivamente prolisso, questa trattazione si concentrerà sull'analisi sperimentale di due delle quattro aree sopra descritte. Le restanti due non riportate, infatti, condividono lo stesso principio di funzionamento, ma si applicano all'ambito medico, un settore per il quale risulta complesso ottenere dati e informazioni che non ledano la riservatezza e la privacy dei soggetti interessati. Per quanto riguarda lo sviluppo dei relativi *case study*, l'attività di test per ciascuno dei due microservizi selezionati inizierà dalle rispettive sezioni principali contrassegnate dalla dicitura *"in tempo reale"*. In ogni caso, i dettagli specifici relativi alle singole interfacce grafiche e ai relativi strumenti di configurazione messi a disposizione da AWS verranno esaminati all'interno delle rispettive sezioni.

Relativamente alla gestione dei risultati, Amazon Transcribe consente di visualizzare l'output testuale direttamente nello schermo e, per specifiche configurazioni, di implementare una serie di ulteriori funzionalità e di memorizzare i dati da esse generate all'interno di *bucket S3* per lo storage. Nonostante l'infrastruttura cloud metta a disposizione, anche per questo servizio, delle API per la programmazione e l'automazione dei servizi all'interno di software di terze parti, ai fini di questa analisi si utilizzeranno esclusivamente le funzionalità offerte dall'interfaccia grafica, approccio ritenuto ottimale per velocizzare le operazioni di visualizzazione dei risultati e di verifica qualitativa delle trascrizioni.

5.3 Case study: Analisi sperimentale dei servizi di Amazon Transcribe

Il case study riportato in questo capitolo si concentra sul settore della gestione dei flussi audio strutturati. Tale ambito, in questa sezione, si sviluppa su due principi fondamentali: l'estrazione dei suoni e l'interpretazione automatizzata del parlato. L'ottimizzazione di questi processi rappresenta, infatti, anche in questo caso, un acceleratore e un vero e proprio vantaggio competitivo per le aziende che necessitano di convertire varie conversazioni in

dati testuali, riducendo i tempi di ascolto e di trascrizione manuale da parte del personale. Applicare lo *Speech-To-Text* in questo contesto significa automatizzare la reportistica o la sottotitolazione, garantendo, allo stesso tempo, un'estrazione di informazioni utili da flussi audio complessi, conversazioni telefoniche o registrazioni ambientali.

L'esperimento mira a testare la capacità di Amazon Transcribe di interpretare e convertire il linguaggio parlato nonostante le diverse variabili linguistiche ed acustiche presenti in un determinato flusso audio. Il lavoro esplora l'efficacia della piattaforma sia nell'elaborazione di file audio pre-registrati di varia natura, sia nell'accuratezza della trascrizione in presenza di fattori di difficoltà, come rumore di fondo, sovrapposizioni vocali o accenti differenti. Di seguito, per i moduli esaminati di Amazon Transcribe, vengono analizzate nel dettaglio le fasi di elaborazione dei campioni utilizzati, i risultati emersi dall'elaborazione e una successiva valutazione qualitativa degli output restituiti dalla piattaforma cloud.

5.3.1 Stress test dialettico con Amazon Transcribe

Il primo microservizio analizzato trova luogo nell'estrazione del testo a partire da flussi multimediali registrati o in tempo reale. All'interno di scenari aziendali moderni, come la trascrizione delle riunioni aziendali o dei contenuti, viene richiesto un dispendio notevole di tempo introducendo, anche, un rischio di omissioni ed errori di battitura. Relativamente a contesti operativi più avanzati, come il monitoraggio dei media o la moderazione automatica dei contenuti audio, la necessità di identificare parole chiave riservate, terminologie sensibili o espressioni non conformi alle piattaforme richiede una sorveglianza e un ascolto costante da parte degli operatori.

L'esperimento pratico è stato eseguito fornendo in input alla piattaforma cloud tre tracce audio al fine di rappresentare vari scenari di complessità linguistica. Dal punto di vista metodologico, sebbene il servizio offra all'utente la possibilità di addestrare modelli linguistici personalizzati, si è scelto di basare i test sui modelli *built-in* forniti da AWS, escludendo quindi il loro addestramento. Con l'obiettivo di testare l'efficacia delle logiche di controllo sui contenuti, si è tuttavia scelto di configurare e implementare uno strumento di ottimizzazione tramite la sezione di *"Filtraggio del vocabolario"*.

In continuità con il precedente *case study* visivo, lo stress test è stato strutturato secondo un approccio incrementale basato su vari gradi di difficoltà linguistica, allo scopo di valutare l'efficacia del riconoscimento di AWS.

Per ciascun test verrà inizialmente presentata la trascrizione testuale restituita dalla piattaforma in assenza di filtri, con l'obiettivo di mostrare l'esatta capacità di ricezione del servizio. Successivamente, verrà mostrato il medesimo output con le logiche di mascheramento applicate, evidenziando così il reale livello di efficacia e accuratezza del filtro sui termini predefiniti e sulle informazioni sensibili.

Prima di poter procedere allo stress test è necessario, per le intenzioni spiegate, configurare il filtraggio del vocabolario. A tale scopo, la Figura 5.4 illustra l'interfaccia di configurazione volta a testare il primo microservizio, mostrandone le impostazioni per valutarne le potenzialità e le capacità.

Come visibile nella Figura 5.4, l'interfaccia consente di configurare varie funzionalità per il primo microservizio: dal rilevamento automatico della lingua, alla suddivisione degli interlocutori, al rilevamento di informazioni sensibili, fino al filtraggio del vocabolario e alla personalizzazione del modello linguistico.

Per l'implementazione del filtraggio del vocabolario, la console mette a disposizione l'opzione *"Crea un nuovo filtro del vocabolario"* (evidenziato in giallo in figura). Questa procedura reindirizza l'utente alla relativa sezione, all'interno della quale è possibile assegnare una denominazione al filtro e procedere al caricamento del file di testo. In questo scenario

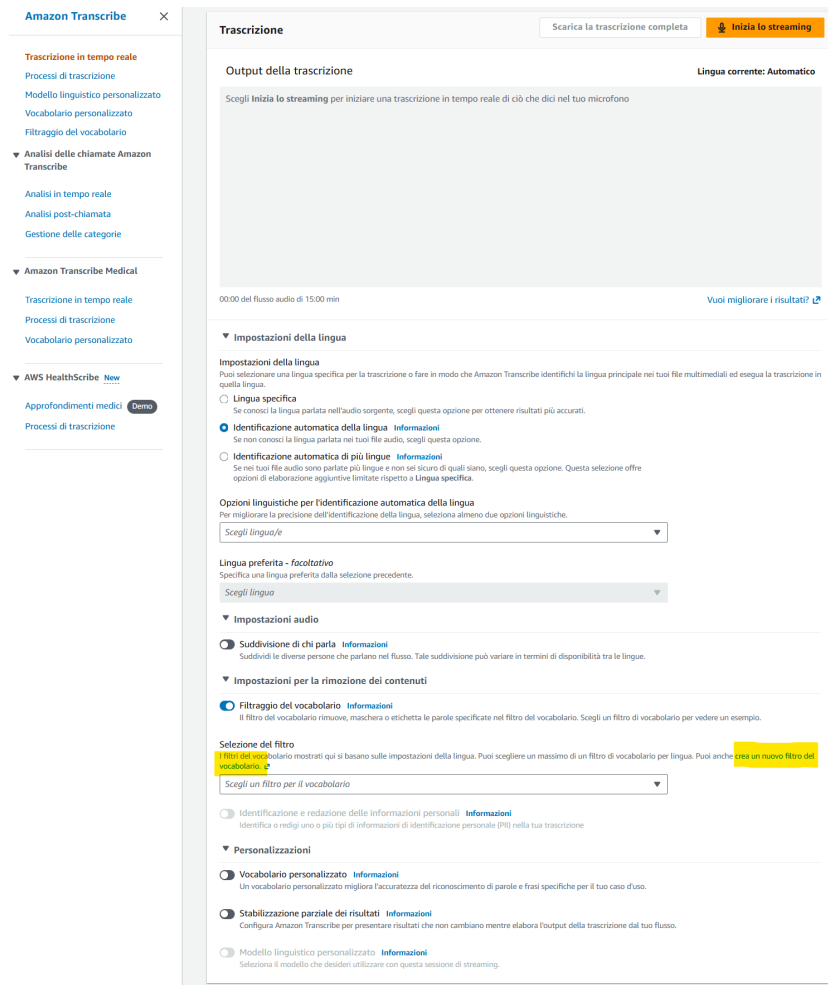


Figura 5.4: Illustrazione dell'interfaccia del primo microservizio di Amazon Transcribe

sperimentale, con finalità puramente illustrative e di verifica, è stato predisposto un file contenente i seguenti vocaboli:

- *king*;
- *war*;
- *empire*;
- *three*;
- *craic*;
- *lad*;
- *development*;
- *technology*;
- *today*;
- *God*;
- *Protestant*;

- *Church*;
- *government*;
- *worship*.

Una volta configurate le impostazioni, è possibile procedere con i test. Di seguito vengono riportati i test suddivisi per categoria e con i relativi risultati emersi dall'elaborazione cloud:

- *Trascrizione (livello base)*: al fine di verificare l'accuratezza del modello *built-in*, il primo campione di analisi presenta il discorso ufficiale di giuramento del Re d'Inghilterra Carlo III. Questo scenario si distingue per una traccia audio pulita, un'assenza pressoché totale di rumore ambientale di fondo e per una dizione chiara, configurandosi come il test a minore complessità dell'intera sessione sperimentale. La Figura 5.5 illustra gli output fatti generare al servizio: nello specifico, viene mostrato il testo in assenza di alterazioni (figura superiore) insieme al risultato elaborato tramite l'applicazione simultanea del filtro del vocabolario personalizzato e del mascheramento delle informazioni sensibili (figura inferiore).

I risultati dell'elaborazione cloud illustrati in Figura 5.5 confermano l'accuratezza del modello *built-in* nel processare un flusso audio chiaro e senza interferenze. L'analisi dei risultati permette di fare alcune importanti considerazioni:

- *PII detection*: l'algoritmo di identificazione delle informazioni sensibili ha dimostrato un'eccellente precisione nel riconoscere le entità critiche. Nello specifico, l'estrazione del nome proprio ha registrato un indice di confidenza pari a 95.70%, mentre l'isolamento degli "indirizzi" ha raggiunto punteggi quasi massimi, pari al 99.93%.
 - *Efficacia del vocabolario personalizzato e mascheramento*: inserita l'impostazione *identificazione e redazione*, il sistema non si è limitato a sostituire i termini del vocabolario personalizzato (evidenziati in blu nell'output originario, come "God", "King", "Church government") applicando la rimozione nel testo finale, ma ha parallelamente rimpiazzato le informazioni sensibili (sottolineate in rosso, come "Charles the 3rd") convertendole nelle entità [NAME] e [ADDRESS].
 - *Contesto linguistico*: l'applicazione automatica dei filtri ha rimosso o alterato intere porzioni di frasi, generando anomalie di punteggiatura (come virgole isolate) e una parziale frammentazione della struttura.
- *Trascrizione in dialetto non standard (Livello intermedio)*: al fine di mettere in difficoltà la tolleranza del modello *built-in*, il secondo campione di analisi presenta un estratto video YouTube intitolato "Irish Farmers Glorious Accent Is So Strong". Questo scenario introduce un livello di stress test intermedio, caratterizzato da un parlato rurale ed estremamente stretto, il quale presenta delle caratteristiche complesse da interpretare persino per un orecchio umano non nativo. Trattandosi, nello specifico, di un'intervista giornalistica televisiva che vede l'alternanza di più voci sul campo, in questa sessione è stata attivata l'opzione che permette la suddivisione automatica degli interlocutori. Inoltre, per ottimizzare l'elaborazione, è stata esplicitamente indicata alla piattaforma la localizzazione linguistica mirata (*Inglese, IE*), mantenendo tuttavia attive le stesse configurazioni di *PII Redaction* e i filtri del vocabolario personalizzato impostati nel test precedente. Per quanto riguarda i risultati relativi a questo test, per chiarezza e maggior possibilità di lettura, è stato scelto di utilizzare le box sottostanti: La prima casella di

Output della trascrizione

Lingua corrente: Inglese, Regno Unito

I, **Charles the 3rd**, by the grace of **God** of the **United Kingdom of Great Britain and Northern Ireland**
 and to my other realms and territories, **King**, defender of the faith.
 do faithfully promise and swear that I shall inviolably maintain and preserve the settlement of the true Protestant religion, as established by the
 laws made in **Scotland**
 in prosecution of the claim of rights
 and particularly by an act in titled an act for securing the Protestant religion and Presbyterian **Church government**
 and by the acts passed in the parliament of both kingdoms for union of the two kingdoms.
 together with the **government**, **worship**, discipline, rights, and privileges of the **Church** of Scotland.
 So help me **God**

Output della trascrizione

Lingua corrente: Inglese, Stati Uniti

I, **[NAME]**, by the grace of of the **[ADDRESS]**.
 and to my other realms and territories, , defender of the faith.
 do faithfully promise and swear that I shall inviolably maintain and preserve the settlement of the true religion, as established by the laws made in
[ADDRESS] in prosecution of the claim of rights
 and particularly by an act in titled an act for securing the religion and Presbyterian .
 and by the acts passed in the parliament of both kingdoms.
 for union of the two kingdoms.
 together with the , .
 discipline, rights and privileges of the of Scotland.
 So help me

Figura 5.5: Illustrazione dell'elaborazione del giuramento di Carlo III senza filtri (figura superiore) e con filtri (figura inferiore)

testo (in giallo) illustra i sottotitoli corretti del video, mentre la seconda casella di testo (in blu) mostra i risultati del servizio senza filtri; infine, la casella di testo (in rosso) mostra le metriche con i filtri applicati.

Sottotitoli corretti del video sottoposto al test

Speaker 01:

Gordian County Kerry are investigating the theft of up to 60 sheep on Mount Brandon. One of the farmers is offering a reward for information leading to the return of the ewes, which are worth thousands of euros.

Speaker 02:

Well, there's a fine dash on them. There's red and blue on them. And anyone can see it, You own them, fine dash on them.

Speaker 03:

Last Saturday Mikey Joe O'Shea brought his flock of Scotch sheep down from the mountain common edge ahead of Lammings. He discovered over 50 were missing, allowing for a number of deaths and strays, Mikey is convinced over 45 sheep had been stolen.

Speaker 02:

Possible at night, there'd be a full moon there up all night and and then sure, it'd be bright out. Now, could anyone go up in the mountains with no lights on? well there's been 35 sheep missing like, and the lambs and ewes, the sheep's count out to a nice bit of money, like. What you do something about it? Nothin',

Speaker 04:

Up there! Don't you see it? They were up there!

Speaker 03:

Mikey's next door neighbour says some of his sheep have also been stolen.

Speaker 04:

Come back, come back, come back. I'm missing about 10 ewes. It's not all that difficult. All we got to do is have a good dog. Have a good and go at night some moonshine night. Just put the dog around them, put them on a trailer or walk em. And then probably somebody else to pick him up. Whoever is doing it knows what he's doing.

Output di AWS (Errori Evidenziati)

Speaker 01:

Gordian County Kerry are investigating the theft of up to 60 sheep on Mount Brandon. One of the farmers is offering a reward for information leading to the return of the ewes, which are worth thousands of euros.

I find that. That's right then. You own them, and you want some see find Last Saturday,

Speaker 02:

Mikey Joe

Speaker 03:

O'Shea brought his flock of Scotch sheep down from the mountain common edge ahead of Lamming. He discovered over 50 were missing, allowing for a number of deaths and strays, Mikey is convinced over 45 sheep had been stolen.

Speaker 01:

What I ate like be full Monday one night and and should it be bright out and could anyone go up in the mountains of our light

Speaker 04:

show.

Speaker 01:

when I was 35 sheep missing like in the lambs and I was in the shape of this sun sun out and nice to what they

Speaker 04:

like.

and

Speaker 01:

be about and laughing.

Speaker 04:

So so

Speaker 03:

Mikey's next door neighbour says some of his sheep have also been stolen.

Speaker 04:

Come back, come back, come back. I ain't missing about 1010 girls.

It's not all that difficult. All we got to do is have a good jog.

Have a good jargon

go at night and moonshine night. Just put the dagger around him, put him on a trailer, walk him.

And then probably somebody else to pick him up.

Whoever is doing it knows what he's doing.

Output di AWS CON filtri (Errori e PII Evidenziati)

Speaker 01:

Gordian County Kerry are investigating the theft of up to 60 sheep on Mount [AD-DRESS]. One of the farmers is offering a reward for information leading to the return of the ewes, which are worth thousands of

Speaker 02:

euros.

Speaker 03:

Well I'll find that. That's what then. You owned them and you want some see find national.

Speaker 04:

Last Saturday,

Speaker 01:

[NAME][NAME] brought his flock of Scotch sheep down from the mountain common edge ahead of

Speaker 02:

Lamming.

Speaker 01:

He discovered over 50 were missing, allowing for a number of deaths and strays, [NAME] is convinced over 45 sheep have been stolen. He discovered over 50 were missing, allowing for a number of deaths and strays, [NAME] is convinced over 45 sheep have been stolen.

Speaker 03:

What I like be Monday one night and and should it be bright out and could anyone up in the mountains where light

Speaker 02:

show.

Speaker 03:

when I was [AGE] sheep missing like in the yams and I was in the shape that sun's gone out and nice to

Speaker 02:

like. and

Speaker 03:

be all about and

Speaker 02:

laughing. So so.

Speaker 03:

Mikey's

Speaker 01:

next door neighbour says some of his sheep have also been stolen. Come

Speaker 03:

back, come back, come

Speaker 02:

back.

Speaker 03:

I'm missing about

Speaker 02:

1010 girls.

Speaker 05:

It's not all that difficult. I think I you have a good

Speaker 02:

dog.

Speaker 03:

have a good dragon Go at night and moonshine night. Just put the

Speaker 05:

dog around him, put him on a trailer or walk him.

Speaker 03:

And then

Speaker 05:

probably somebody else to pick him

Speaker 02:

up.

Speaker 05:

Whoever's doing it knows what he's

Speaker 02:

doing.

Dall'analisi dei risultati emersi all'interno dei tre box, è possibile evidenziare un drastico decadimento delle prestazioni del modello di AWS quando sottoposto a uno stress test con accento non standard. Il primo elemento che merita di essere evidenziato è rappresentato dalle sostituzioni delle parole che causano un cambiamento di significato

del testo. Il sistema mostra evidenti difficoltà nel riconoscimento di alcuni termini: l'espressione *10 ewes* (10 pecore) viene ripetutamente trascritta come *1010 girls*, il termine *lambs* (agnelli) viene confuso con «*yams*» (patate dolci) e la parola *dog* (cane) si trasforma in *jog o*, addirittura, in *dagger* (pugnale). Si riscontra inoltre un particolare fenomeno che mostra la ripetizione della frase: Speaker 01 (*He discovered over 50 were missing...*).

Il problema più grande, tuttavia, riguarda la suddivisione degli interlocutori, che subisce un vero e proprio degrado. Come evidenziato negli ultimi due box, AWS riscontra difficoltà nel comprendere qual'è l'interlocutore che sta effettivamente parlando (ad esempio, assegnando allo Speaker 01 intere frasi appartenenti allo Speaker 02 o allo Speaker 03). A questo problema si aggiunge una frammentazione delle frasi: periodi che erano originariamente fluidi e pronunciati da un unico soggetto diventano interventi di qualche secondo, isolando singoli termini come *Lamming*, *like* o *laughing*, attribuendoli a interlocutori differenti. La confusione dell'algoritmo diventa lampante con la comparsa dello *Speaker 05*, entità totalmente irreale introdotta interpretando i rumori di fondo.

È curioso osservare come, nonostante sia stata impostata la lingua *Irlandese*, l'algoritmo trova difficoltà nel comprendere e trascrivere il termine *them*. Questa parola, infatti, viene più volte confuso con *him*, come visibile nell'ultimo scambio di battute all'interno del riquadro blu.

Infine, la funzione di *PII* introduce ulteriori anomalie dovute alla mancanza di comprensione del contesto da parte del filtro della privacy. Se da un lato l'oscuramento geografico rileva correttamente (*Mount [ADDRESS]*), dall'altro il mascheramento dei nomi propri genera stringhe duplicate (*[NAME][NAME]*). L'errore più evidente si riscontra nell'applicazione del tag *[AGE]*, il quale va a sostituire il numero 35. Nel testo reale, tale cifra indicava il numero di pecore disperse (*35 sheep missing*) e non l'età di un individuo.

- *Trascrizione con cadenza indiana (Livello avanzato)*: al fine di testare il vero e proprio punto di rottura del modello *built-in*, il terzo e ultimo campione di analisi presenta una traccia audio estratta dalla satira "*Phas Gaye Re Obama*" (<https://youtube.com/watch?v=bhR1nJ1oY9M>), caratterizzata da un accento anglo-indiano appositamente estremizzato. Questa casistica rappresenta il massimo livello di stress test poiché introduce sia un dialetto estremamente complesso da interpretare persino per un orecchio madrelingua, sia costanti rumori di fondo generati dalla colonna sonora. A differenza delle prove precedenti, in questo specifico test non è stato implementato alcun filtro; l'algoritmo di AWS, infatti, come visibile nella Figura 5.6, ha riscontrato una difficoltà estrema nell'analisi della traccia, fallendo nella decodifica di oltre il 90% del testo originario e rendendo inutile l'introduzione di filtri necessari ad aumentare la difficoltà.

Come evidenziato dall'output in Figura 5.6, il terzo test mostra il completo fallimento di Amazon Transcribe. L'impossibilità di applicare i filtri custom o di *PII* risiede nel risultato del testo generato ridotto a pochi frammenti di parole prive di logica.

5.3.2 Analisi delle chiamate Amazon Transcribe

Il secondo microservizio, soggetto all'analisi in questa sezione, è adibito alla gestione delle relazioni con i clienti. All'interno delle aziende moderne, infatti, l'analisi delle telefonate di assistenza o di vendita rappresenta una fonte di dati importante; tuttavia, l'ascolto e la valutazione manuale delle registrazioni da parte degli operatori richiedono un dispendio di tempo e risorse non accettabile, limitando il controllo solo a piccoli campioni casuali.

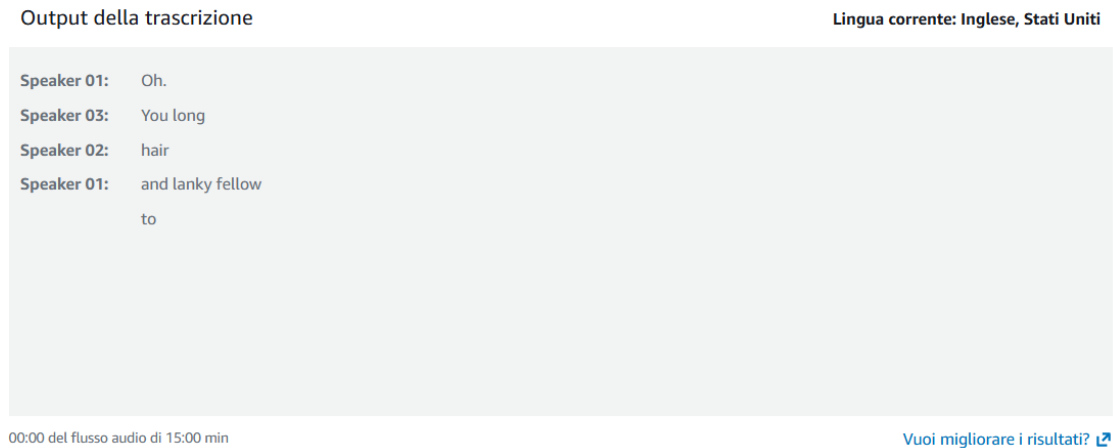


Figura 5.6: Illustrazione dell'elaborazione di una scena del film "Phas Gaye Re Obama"

L'applicazione di tecniche avanzate di *Natural Language Processing*, in questo specifico caso, permette non solo di automatizzare la trascrizione della chiamata, ma di estrarre dati legati alle conversazioni e allo stato emotivo degli interlocutori, trasformando flussi audio non strutturati in strutture dati valutabili.

L'esperimento è stato condotto sfruttando i due campioni audio pre-configurati che la piattaforma AWS mette a disposizione per testare le potenzialità del sistema. Questa scelta consente di valutare le capacità del modello *built-in* su tracce volte a simulare scenari reali di comunicazione telefonica. A differenza della trascrizione standard esaminata nella sezione precedente, l'algoritmo di *analisi in tempo reale* non si limita a convertire il parlato in testo, ma esegue un'analisi parallela volta a misurare alcuni parametri comportamentali, quali:

- *La Sentiment Analysis*: l'identificazione dello stato d'animo (positivo, negativo, neutro o misto) sia del cliente che dell'operatore, mantenendolo monitorato per tutta la durata della chiamata.
- *La rilevazione dei tempi di silenzio e di conversazione*: il calcolo esatto dei periodi di silenzio e delle percentuali di tempo in cui ciascun interlocutore ha occupato il canale (visibile solo nell'utilizzo senza interfaccia grafica).
- *Le casistiche di interruzione e sovrapposizione*: l'individuazione dei momenti esatti in cui un parlante interrompe l'altro, indice importante di problematiche nella gestione del dialogo.

In aggiunta ai parametri sopra menzionati, l'architettura di *analisi delle chiamate* offre la possibilità di salvare l'output di ogni singola chiamata all'interno di un *bucket* Amazon S3. Abilitare questa specifica significa poter attivare la generazione del riepilogo automatico della conversazione. Nel presente *case study*, l'analisi si baserà sull'interfaccia grafica per ragioni di ottimizzazione visiva.

Un ulteriore elemento interessante è rappresentato dalla funzionalità di *"gestione delle categorie"*. Questo strumento permette di istruire l'algoritmo affinché riesca ad identificare determinati termini in tempo reale sulla base di regole logiche predefinite. Al fine di testare tale componente, prima di avviare l'elaborazione, si è proceduto alla creazione di due categorie personalizzate di tipo *"Real-time"*, adottando l'utilizzo dei modelli consigliati da AWS:

- *Categoria "Angry"*: configurata impostando una regola relativa al parlato del cliente. L'algoritmo è stato impostato per intercettare casi in cui il contenuto verbale esprima

uno stato d'animo negativo, unito all'utilizzo di specifiche parole chiave. In particolare, la regola si attiva quando, durante l'intera durata della chiamata, l'interlocutore identificato come "Cliente" pronuncia vocaboli associati alla rabbia o al disappunto (quali "upset" e termini affini). La configurazione di tale categoria è rappresentata nella Figura 5.7.

- **Categoria "Comprehension":** impostata analizzando sia il parlato del cliente che quello dell'agente. In questo caso, la regola mira a monitorare la comprensione dell'agente rispetto allo stato d'animo del cliente e, al contempo, l'empatia del cliente nei confronti dell'azienda, attivandosi in corrispondenza di formule di cortesia o di scusa (quali "sorry") pronunciate da entrambi gli attori del flusso. La configurazione di tale categoria è rappresentata nella Figura 5.8.

Processi di trascrizione
 Modello linguistico personalizzato
 Vocabolario personalizzato
 Filtraggio del vocabolario
 ► **Analisi delle chiamate Amazon Transcribe**
 Analisi in tempo reale
 Analisi post-chiamata
 Gestione delle categorie
 ► **Amazon Transcribe Medical**
 Trascrizione in tempo reale
 Processi di trascrizione
 Vocabolario personalizzato
 ► **AWS HealthScribe** New
 Approfondimenti medici Help
 Processi di trascrizione

Impostazioni di categoria

Nome della categoria
 Angry
Il nome può contenere fino a 200 caratteri. I caratteri validi sono a-z, A-Z, 0-9, ., _ (brattino basso) e - (brattino).

Tipo di categoria Informazioni
 Scegli il tipo di categoria

Metodo di creazione della categoria Informazioni
☐ Usa un modello (consigliato)
Usa un modello per modificare le regole predefinite.
☒ Crea da zero
Se cancelli le regole che desideri definire, scegli questa opzione.

Regole
Affinché un processo di trascrizione possa essere classificato in questa categoria, devono essere soddisfatte tutte le condizioni della regola.

▼ **Regola 1** Elimina regola
Quando una delle parole era menzionata durante l'intera chiamata in cui colui che parla era cliente.

Tipo di regola Informazioni
 Scegli la regola da definire.
 Scegli un tipo di regola

Logica Informazioni
Definisci le condizioni che devono essere soddisfatte.
 Quando una qualsiasi delle parole era
 menzionato
 durante il
 intera chiamata
 quando colui che parla era
 cliente

Parole o frasi Informazioni
Inserisci le parole o le frasi che desideri cercare nella trascrizione.
 Inserisci una parola o una frase
La parola o la frase può contenere fino a 2.000 caratteri.
 Aggiungi una nuova parola o una frase

upset X
 Annoyed X
 Irritated X
 Cross X
 Angry X
 Frustrated X
 Resentful X
 Furious X
 Enraged X
 Livid X

Aggiungi regola
Puoi aggiungere un massimo di 19 altre regole.

Figura 5.7: Illustrazione della configurazione della categoria "Angry" per il microservizio di Transcribe "Analisi delle chiamate"

Una volta completata la configurazione di queste regole e degli spazi di *storage*, è possibile procedere con i test. Per ciascuno dei due campioni analizzati verranno mostrati l'interfaccia grafica dei risultati, la scomposizione temporale dei messaggi e l'eventuale attivazione delle categorie personalizzate sopra descritte. Di seguito vengono riportati i test eseguiti e i contestuali risultati emersi:

- **Analisi del primo campione audio di test (Caso 1):** il primo campione audio pre-configurato analizzato propone la simulazione di una chiamata di assistenza tra un utente e l'opera-

Processi di trascrizione
 Modello linguistico personalizzato
 Vocabolario personalizzato
 Filtraggio del vocabolario

▼ **Analisi delle chiamate Amazon Transcribe**
 Analisi in tempo reale
 Analisi post-chiamata
 Gestione delle categorie

▼ **Amazon Transcribe Medical**
 Trascrizione in tempo reale
 Processi di trascrizione
 Vocabolario personalizzato

▼ **AWS HealthScribe** [New](#)
 Approfondimenti medici [Demo](#)
 Processi di trascrizione

Impostazioni di categoria

Nome della categoria
 Comprehension
 Il nome può contenere fino a 200 caratteri. I caratteri validi sono a-z, A-Z, 0-9, ., _ (trattino basso) e - (trattino).

Tipo di categoria [Informazioni](#)
 Scegli il tipo di categoria

Metodo di creazione della categoria [Informazioni](#)
☐ Usa un modello (consigliato)
 Usa un modello per modificare le regole predefinite.
☒ **Crea da zero**
 Se compili le regole che desideri definire, scegli questa opzione.

Regole

Affinché un processo di trascrizione possa essere classificato in questa categoria, devono essere soddisfatte tutte le condizioni della regola.

▼ **Regola 1** [Elimina regola](#)
 Quando una delle parole era menzionato durante il intero chiamata in cui colui che parla era sia.

Tipo di regola [Informazioni](#)
 Scegli la regola da definire.
 Scegli un tipo di regola

Logica [Informazioni](#)
 Definisci le condizioni che devono essere soddisfatte.
 Quando una qualsiasi delle parole era

menzionato
 durante il
 intero chiamata
 quando colui che parla era
 sia

Parole o frasi [Informazioni](#)
 Inserisci le parole o le frasi che desideri cercare nella trascrizione.
 La parola o la frase può contenere fino a 2.000 caratteri.

Inserisci una parola o una frase
 Aggiungi una nuova parola o una frase

Empathetic
 Compassionate
 Supportive
 Caring
 Considerate
 Kind
 I understand why you are upset
 I completely understand your frustration
 I hear what you are saying
 That sounds incredibly frustrating
 I am truly sorry for the inconvenience
 I apologize for this mistake
 Thank you for bringing this to my attention
 Let's see how we can fix this
 I am going to look into this right away

Figura 5.8: Illustrazione della configurazione della categoria "Comprehend" per il microservizio di Transcribe "Analisi delle chiamate"

tore di una catena alberghiera ("*Suite First Hotel*"). L'audio ha una buona qualità acustica di linea telefonica standard e un parlato chiaro da parte di entrambi gli interlocutori. L'obiettivo del test è verificare la capacità di Amazon Transcribe di gestire correttamente i turni di conversazione (*Agent* e *Customer*), analizzare il *sentiment* di entrambi i soggetti e attivare le categorie personalizzate ("*Angry*" e "*Comprehension*") precedentemente configurate. La Figura 5.9 illustra lo sviluppo temporale della conversazione e l'interfaccia di riepilogo generata dal servizio.

I risultati dell'elaborazione, riportati in Figura 5.9, mostrano l'efficacia del microservizio *analisi delle chiamate*. L'analisi permette di trarre le seguenti considerazioni:

- *Attivazione delle categorie e del sentiment*: il sistema ha dimostrato un'eccellente attivazione dei moduli comportamentali personalizzati. Al blocco temporale 00:09.34 - 00:17.95, la rilevazione della parola "*upset*" unita a un tono negativo ha comportato correttamente l'avviso della categoria "*Angry*". Parallelamente, la risposta dell'operatore al blocco 00:18.34 - 00:24.10, caratterizzata dall'espressione empatica "*I'm so sorry that happened to you*", ha generato l'attivazione della categoria "*Comprehension*", associando il *sentiment* positivo per l'agente.
- *Frammentazione del flusso del cliente*: è emersa una piccola inefficienza nella gestione del parlato del cliente. Nei primi due interventi del blocco sinistro, l'algoritmo ha

frammentato un discorso logicamente continuo in due messaggi distinti: il primo interrotto al termine "but" (00:06.36 - 00:09.14) e il secondo avviato come nuova entità.

- *Sottolineatura del problema*: risulta interessante la capacità di evidenziare i blocchi di testo che determinano la rilevazione del problema tramite il tag "Issue Detected".

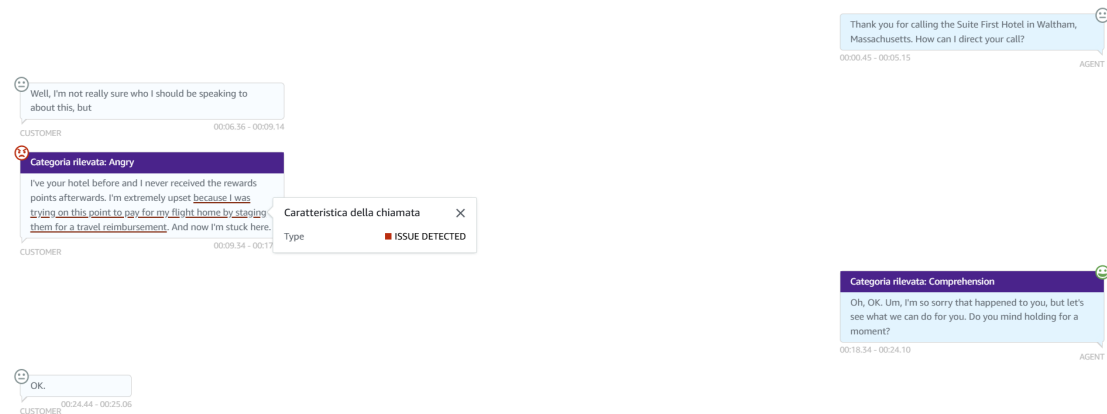


Figura 5.9: Illustrazione dello sviluppo temporale della conversazione e l'interfaccia di riepilogo generata dal microservizio *analisi delle chiamate* per il primo campione audio

- *Analisi del secondo campione di test (Caso 2)*: il secondo audio pre-configurato mostra una conversazione di supporto più lunga, incentrata sulla notifica del cliente per la cancellazione improvvisa della propria polizza auto. Questo scenario introduce un livello di complessità superiore rispetto al precedente, a causa dello spelling di dati anagrafici e della dettatura di numeri di telefono. L'obiettivo del test rimane valutare la robustezza di Amazon Transcribe, tracciare l'evoluzione emotiva del dialogo e gestire il trasferimento di informazioni critiche di contatto. La Figura 5.10 mostra lo sviluppo temporale della conversazione.

I risultati dell'elaborazione, documentati nello sviluppo temporale della Figura 5.10, offrono degli spunti di riflessione sulle capacità analitiche di Transcribe. Di seguito vengono analizzate le principali evidenze emerse:

- *Gestione dello spelling*: il sistema ha mostrato una notevole accuratezza nell'individuare le richieste di chiarimento dell'operatore. Al blocco 00:21.82 - 00:24.75, il sistema ha trascritto correttamente lo spelling in lettere maiuscole del cognome dell'utente ("JAMISON"). Un comportamento interessante si riscontra nella gestione del numero telefonico: nel blocco 00:31.24, la dettatura del cliente viene registrata come stringa numerica frammentata ("61026 51 714"), mentre nel blocco successivo (00:35.37), il sistema ricompone l'entità normalizzandola nel formato ("610-265-1714").
- *Tracciamento del sentiment*: a differenza del primo caso, questo test evidenzia l'efficacia dell'algoritmo nel rilevare le emozioni lungo lo scorrere del tempo. Se nella fase iniziale i blocchi sono dominati da icone neutre a causa della natura prevalentemente informativa dello scambio, nella fase di chiusura si assiste all'attivazione automatica del sentiment positivo. Le espressioni di rassicurazione dell'agente ("send your message to Mr. McAvey right away") e i conseguenti ringraziamenti del cliente diventano dati utili per il monitoraggio della qualità del servizio. Tuttavia,

si riscontra un'anomalia emotiva nel penultimo messaggio dell'agente ("*I would advise not to do anything until...*"), segnalando come *sentiment* negativo un consiglio che sarebbe dovuto passare come neutrale.

- *Sottolineatura del problema*: risulta, ancora una volta, interessante la capacità del sistema di evidenziare i blocchi di testo che determinano la rilevazione del problema tramite il tag "*Issue Detected*".

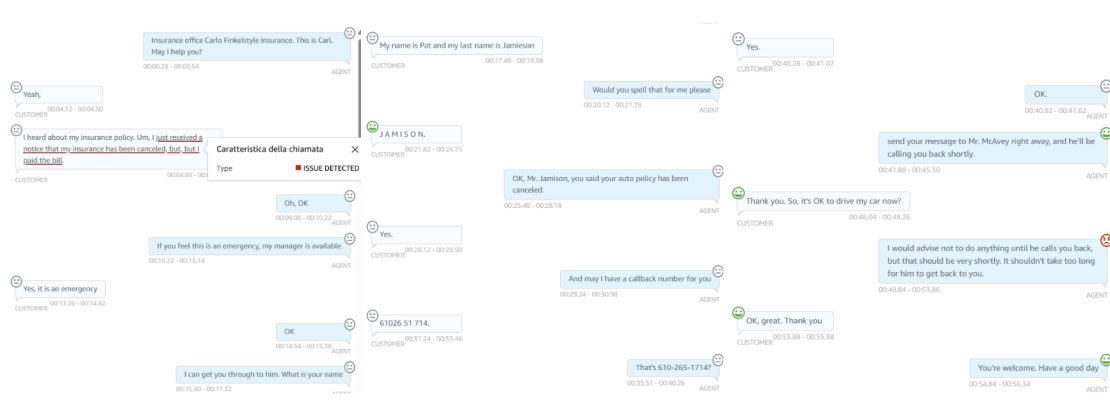


Figura 5.10: Illustrazione dello sviluppo temporale della conversazione e l'interfaccia di riepilogo generata dal microservizio *analisi delle chiamate* per il secondo campione audio

5.3.3 Analisi qualitativa

L'analisi qualitativa dei test eseguiti su Amazon Transcribe ha evidenziato come la risposta dei modelli *built-in* varia al differenziarsi del flusso audio fornito ad esso in ingresso. Il servizio, infatti, è risultato fortemente dipendente dalle varianti dialettali e dal rumore di fondo presente all'interno delle sorgenti audio analizzate.

Per quanto riguarda la *Trascrizione standard* e il *filtraggio del vocabolario*, l'accuratezza del servizio risulta strettamente dipendente dalla chiarezza e dalla pulizia della traccia audio. Nel caso del giuramento di Re Carlo III, il parlato lineare e l'assenza di disturbi di fondo hanno garantito un'eccellente precisione. Queste due caratteristiche, infatti, hanno permesso al sistema di isolare e mascherare correttamente le informazioni sensibili con indici di confidenza pari al 95.70% per il nome proprio e al 99.93% per gli indirizzi; ciò ha anche dato la possibilità di individuare e rimuovere dal testo dei termini corrispondenti a quelli inseriti nel vocabolario personalizzato pre-configurato. Al contrario, è stato mostrato come il peggioramento della sorgente audio condiziona drasticamente le performance. Nel test sull'accento irlandese rurale, sebbene l'algoritmo sia stato aiutato impostando la lingua in (*English, Ireland*) e abbia gestito l'intervista attivando la suddivisione degli interlocutori, sono state evidenziate le prime lacune relative alla confusione di alcuni termini. Il vero e proprio punto di rottura si è registrato, tuttavia, con lo stress test avanzato sul film "*Phas Gaye Re Obama*": la combinazione di un marcato accento anglo-indiano e la presenza della colonna sonora di sottofondo ha completamente inibito i risultati, portando al totale fallimento della trascrizione per oltre il 90% del testo, riuscendo a restituire solo quattro righe frammentate ("*You long hair and lanky fellow*").

Relativamente all'*Analisi delle chiamate*, lo strumento ha dimostrato un'elevata efficienza nel comprendere lo scambio di risposte tra gli interlocutori e nel tradurre le regole logiche in successive categorie. Nel primo test in ambito alberghiero, il sistema ha mostrato con precisione lo sviluppo temporale della conversazione, attivando correttamente la categoria

personalizzata "Angry" (al rilevamento del termine "upset") e la categoria "Comprehension" in corrispondenza delle scuse dell'operatore, registrando, parallelamente, il motivo fondamentale della chiamata. Nel secondo test in ambito assicurativo, caratterizzato da una conversazione più estesa, l'algoritmo ha mostrato un'ottima flessibilità nel gestire lo spelling del cognome "JAMISON" e del numero di telefono. Parallelamente, si è dimostrato ottimale nel monitorare il *sentiment* degli interlocutori, registrando il passaggio da uno stato iniziale interamente neutro a una situazione di *sentiment* positivo. Tuttavia, entrambe le sessioni di analisi telefonica non si sono rivelate perfette; nel primo test il blocco introduttivo di testo si è interrotto dividendo il discorso in due messaggi separati, mentre nel secondo, il penultimo messaggio dell'agente è risultato caratterizzato da *sentiment* negativo senza un apparente motivo.

5.4 Punti di forza e limitazioni riscontrate

Una volta effettuata l'analisi qualitativa dei risultati ottenuti dai test effettuati sui due microservizi, è possibile riassumere i punti di forza e le limitazioni rilevate nell'utilizzo di Amazon Transcribe.

5.4.1 Punti di forza riscontrati

Amazon Transcribe si è dimostrato una risorsa notevolmente utile per l'integrazione dello *Speech-To-Text* nella maggior parte delle situazioni quotidiane. I principali punti di forza rilevati comprendono:

- *Interfaccia user-friendly*: anche in questo caso tutti i microservizi *SaaS* di Amazon Transcribe, in particolare i moduli analizzati, offrono strumenti estremamente intuitivi. Possono essere utilizzati facilmente anche da coloro che non hanno competenze nel settore del *Machine Learning* o di gestione di file audio. Il sistema permette, infatti, di trascrivere i propri contenuti in pochi minuti, persino senza richiedere alcun tipo di configurazione.
- *Automazione aziendale*: durante i test sulla gestione delle relazioni con i clienti, il modulo di *Analisi delle chiamate* si è dimostrato molto efficace nel tracciare in tempo reale l'andamento delle conversazioni. Incrociando regole preimpostate dall'utente (in questo caso l'agente), è risultato, infatti, possibile far sì che il sistema attivasse in modo preciso categorie personalizzate come, ad esempio, "Angry" e "Comprehension", mostrati all'interno del case study. Questo tipo di approccio offre sicuramente un beneficio a livello di sviluppo, automatizzando casistiche di analisi di qualità del servizio in ottica di miglioramento dell'efficienza aziendale.
- *Estrazione corretta di caratteristiche particolari del linguaggio*: il test relativo all'*analisi delle chiamate* si è dimostrato eccellente nell'estrarre caratteristiche particolari del parlato umano quotidiano. Lo strumento, infatti, come dimostrato nel test relativo, ha perfettamente riconosciuto lo *spelling* dell'agente durante la conversazione e, parallelamente, si è dimostrato eccellente nel normalizzare in formati leggibili e impattanti all'occhio umano i numeri di telefono dettati nel corso della conversazione, peculiarità non semplici da rilevare e formattare per questo tipo di modelli.
- *Efficaci nella PII Detection*: in presenza di un flusso audio comprensibile, l'algoritmo di identificazione delle informazioni sensibili ha dimostrato un'eccellente precisione. Il sistema riesce a isolare entità critiche, come nomi propri e riferimenti geografici, associando, anche, indici di confidenza molto elevati. Questa affidabilità rende il microservizio

uno strumento pronto all'uso per l'integrazione di software di mascheramento dei dati sensibili all'interno delle aziende, senza la necessità di addestramenti preventivi.

5.4.2 Limitazioni riscontrate

Nonostante i vantaggi mostrati nella sezione precedente, la fase sperimentale ha mostrato alcune limitazioni da evidenziare. Le principali includono:

- *Vincoli di test nell'analisi delle chiamate*: all'interno dell'interfaccia grafica, nella sezione dedicata all'analisi delle chiamate in tempo reale, è possibile analizzare unicamente i due file audio forniti dalla piattaforma, facendo, così, risultare il microservizio fortemente limitato sul piano sperimentale.
- *Degrado delle prestazioni dipendente dalla qualità dell'audio*: i test intermedio e avanzato del primo microservizio hanno evidenziato le difficoltà dei modelli *built-in*. Infatti, di fronte all'unione di un accento non standard e dei rumori di fondo, la piattaforma ha mostrato un degrado proporzionale alla scarsa qualità del flusso audio. Il fallimento della trascrizione di oltre il 90% del testo originario, nel test avanzato, dimostra che l'accuratezza dei modelli pre-addestrati risulta fortemente dipendente da input qualitativamente elevati, richiedendo, così, metodologie di pulizia preliminari in contesti reali complessi.

Tuttavia, a prescindere dai vantaggi e dagli svantaggi, è necessario evidenziare come il primo microservizio si sia dimostrato impeccabile nel caso semplice del giuramento reale, per poi decadere progressivamente all'aumentare della complessità della traccia audio. L'interpretazione di questo comportamento oscilla tra il vantaggio e il limite tecnologico a seconda dei punti di vista; d'altronde, un flusso audio che risulti di difficile comprensione per un umano a causa di rumori di fondo e del dialetto stretto presenterà la stessa complessità per un modello di questo tipo, se non maggiore. La macchina, infatti, come spiegato nella Sezione 5.1.2, fonda il proprio processo di comprensione sui fonemi rilevati, mancando della capacità cognitiva caratteristica che permette all'uomo di ricostruire il significato basandosi sul contesto di contorno.

Questo capitolo, in primo luogo, si articola presentando una sezione dedicata al confronto dei servizi analizzati tramite i risultati dei case study, all'interno della quale le evidenze quantitative e qualitative di Amazon Comprehend, Amazon Rekognition e Amazon Transcribe, discusse nei precedenti capitoli, vengono messe in relazione per evidenziare analogie e punti di rottura. Successivamente, il capitolo si concentra sulle implicazioni relative al mondo aziendale, analizzando l'impatto, i vantaggi competitivi e i cambiamenti organizzativi determinati dall'adozione di queste architetture SaaS. Un'integrazione ottimale non può non prendere in considerazione i vincoli tecnologici ed etici che le implementazioni di questi sistemi comportano; a tale scopo, il capitolo mostra le strategie per mitigare i rischi e massimizzare i benefici, offrendo un insieme di soluzioni per gestire la sicurezza dei dati, i falsi positivi e, più in generale, i problemi intrinseci dei modelli cloud. Infine, l'elaborato si conclude con i possibili scenari evolutivi della collaborazione tra uomo e macchina. Questa sezione finale esplora l'attuale percezione sociale e professionale nei confronti dei vari sistemi intelligenti oggi esistenti, delineando una sinergia in cui l'algoritmo non sostituisce l'uomo, bensì lo potenzia a livello cognitivo ed operativo.

6.1 Confronto dei servizi analizzati tramite i risultati dei case study

La conclusione della fase sperimentale offre l'opportunità di trarre un riassunto delle tre architetture SaaS oggetto di studio: Amazon Comprehend, Amazon Rekognition e Amazon Transcribe. Sebbene i servizi operino in ambiti totalmente differenti (rispettivamente, il testo non strutturato, i flussi multimediali visivi e i segnali audio), l'analisi simultanea dei test eseguiti permette di evidenziare come essi condividano la stessa logica di funzionamento a livello software, mostrando, tuttavia, nette differenze nel momento in cui vengono sottoposti a condizioni di stress. Nello specifico, per "stessa logica di funzionamento" si intende il fatto che tutti e tre i servizi cloud analizzati si sviluppano attraverso quattro fasi essenziali, ovvero:

1. *Input grezzo*: l'utente fornisce al sistema un file di partenza non strutturato (un testo libero, un'immagine o una traccia vocale) che un'applicazione software tradizionale, priva di logiche predittive, non sarebbe in grado di decodificare e comprendere autonomamente.
2. *Elaborazione in cloud*: qualora si utilizzino i servizi tramite l'interfaccia grafica o a livello base, non è richiesta la scrittura di algoritmi terzi né l'addestramento locale di modelli di *Machine Learning*. Non è necessario allocare o configurare infrastrutture hardware

dedicate sui propri server aziendali, poiché l'intero carico computazionale è gestito dal *provider*.

3. *Output standardizzato*: il sistema, in ognuno dei tre casi, restituisce la risposta elaborata adottando lo stesso formato di testo strutturato JSON. Tale file include metriche, talvolta indici di confidenza statistica (*confidence score*) arricchiti da metadati specifici legati alla natura dell'algoritmo impiegato (come lo spettro emotivo o le marcature temporali).
4. *Integrazione facoltativa*: qualora si avesse la necessità di integrare il servizio cloud all'interno di un software di terze parti, l'infrastruttura mette nativamente a disposizione delle API. Questo consente agli sviluppatori di automatizzare i flussi di dati aziendali.

Questo quadro d'insieme, incentrato sui vantaggi e sui rispettivi limiti, è riassunto sinteticamente nella Tabella 6.1.

Dall'analisi di questa tabella emerge chiaramente come gli ingegneri di AWS abbiano ottimizzato la facilità di utilizzo della piattaforma per gli utenti meno esperti. Al contrario, le funzionalità di addestramento dei modelli avanzati risultano meno intuitive per gli specialisti del settore; tale scelta progettuale è probabilmente guidata dal presupposto che i professionisti possiedano le competenze necessarie per utilizzare metodologie di sviluppo più complesse.

È, inoltre, visibile dalla tabella, ma di più dai capitoli relativi ai case study, che tutti e tre i servizi analizzati offrono prestazioni eccellenti in condizioni ottimali. Tuttavia, in assenza di un addestramento specifico, la maggior parte dei microservizi analizzati nei case study mostra un progressivo decadimento delle performance all'aumentare della complessità del contesto. Questo comportamento è, d'altronde, prevedibile: un singolo algoritmo non può eccellere in ogni scenario e, per specializzarlo in un determinato settore, è indispensabile un addestramento basato su dati mirati.

I modelli pronti all'uso *built-in* di AWS rispondono, quindi, alle esigenze di utenti alle prime armi che cercano algoritmi generalisti. Tuttavia, per scenari applicativi complessi e articolati, l'utilizzo standard non è sufficiente: diventa fondamentale integrare sistemi di controllo, supervisioni umane, software di terze parti e cicli di addestramento supplementari per massimizzare l'accuratezza e raggiungere prestazioni ottimali.

6.2 Implicazioni concrete nel mondo aziendale

L'introduzione delle tecnologie *SaaS*, analizzate nei case study all'interno dei processi aziendali, costituisce una trasformazione profonda nel modo in cui le organizzazioni gestiscono i loro processi e le loro risorse. L'analisi sperimentale e i confronti condotti su Amazon Comprehend, Rekognition e Transcribe dimostrano che, se ottimizzati e utilizzati in modo appropriato, i benefici teorici che essi propongono si trasformano in vantaggi importanti a livello competitivo.

La prima grande implicazione riguarda l'automazione dei servizi di assistenza e supporto dei clienti. Infatti, attraverso l'uso combinato di Amazon Transcribe e di Amazon Comprehend, un'azienda è in grado di digitalizzare la gestione dei propri *call center*. I test eseguiti sul modulo di *Analisi delle chiamate* dimostrano che, venendo queste ultime integrate con dei software terzi (come il centralino cloud di AWS *Amazon Connect*), l'infrastruttura può analizzare automaticamente la totalità delle conversazioni telefoniche aziendali, monitorando l'evoluzione dello stato d'animo dei clienti e attivando alert immediati in tempo reale di fronte a parole chiave con le quali è stato addestrato. Questo esonera i supervisori dall'ascoltare manualmente campioni di registrazioni e, parallelamente, consente all'operatore che sta eseguendo la telefonata di visualizzare in *real-time* l'andamento di quest'ultima, consentendo un intervento solo sulle interazioni realmente critiche e migliorando il supporto del cliente.

Servizio AWS	Vantaggi riscontrati	Svantaggi emersi
Amazon Comprehend	- Interfaccia estremamente intuitiva. - Elevata stabilità e precisione su testi colloquiali e recensioni non strutturate, rispecchiando, nella maggior parte dei casi, il feedback reale dell'utente. - Ottima accuratezza nell'estrazione delle entità e nell'analisi del <i>sentiment</i>. - Continuo e costante apprendimento del modello Cloud.	- I modelli sono capaci di comprendere parzialmente il contesto. - Forte dipendenza dalla lingua per le funzionalità PII (disallineamento con l'italiano). - Forte affidamento a parole chiave superlative, influenzando talvolta eccessivamente l'output. - Rischio di risultati errati in presenza di testi fortemente ironici o ambigui.
Amazon Rekognition	- Interfaccia estremamente intuitiva. - Notevole capacità di operare anche in ambienti a bassa visibilità (come in luoghi notturni e distorti dalla risoluzione). - Riconoscimento automatico degli oggetti e capacità di estrarre molteplici dati del volto (emozioni, età, dettagli biometrici) tutti insieme in un'unica chiamata API. - Possibilità di estendere le rilevazioni dell'algoritmo tramite le <i>Custom Labels</i>.	- Crollo della precisione nei profili e in presenza di occlusioni fisiche del volto. - I dataset dei modelli <i>built-in</i> sono prevalentemente addestrati in ambienti convenzionali, provocando il mancato riconoscimento di diversi soggetti famosi, come evidenziato all'interno della sezione dedicata al case study della <i>Celebrity Recognition</i>. - Alta densità di falsi positivi in contesti ampiamente affollati (scaffali).
Amazon Transcribe	- Interfaccia estremamente intuitiva. - Trascrizione del parlato e <i>PII Redaction</i> ottime in condizioni di flussi audio chiari. - Possibilità di estendere le rilevazioni dell'algoritmo tramite l'inserimento di un <i>vocabolario personalizzato</i>. - Elevata possibilità di automazione tramite l'analisi in tempo reale e la gestione delle categorie.	- Estrema sensibilità ai disturbi esterni, con decadenza progressiva in presenza di accenti non standard e rumori di fondo. - Limitazione relativa al testing limitato del microservizio "analisi delle chiamate" - Limiti relativi alle frammentazioni dei messaggi a causa di brevi esitazioni degli interlocutori. - Limiti relativi al riconoscimento del <i>sentiment</i>.

Tabella 6.1: Analisi comparativa dei vantaggi e degli svantaggi riscontrati nei tre servizi esaminati.

Per quanto riguarda le applicazioni basate su Amazon Comprehend, tramite il supporto di un supervisore umano, diventa vantaggioso integrare questa tecnologia in contesti di moderazione del linguaggio all'interno di piattaforme web e forum aziendali e nell'analisi massiva delle recensioni. Questo approccio permette di monitorare l'andamento e il gradimento di un determinato prodotto sul mercato e, allo stesso tempo, di bloccare potenziali commenti inappropriati, offensivi o non conformi alle linee guida aziendali prima della loro pubblicazione.

Un ulteriore scenario di forte impatto per le aziende, emerso indirettamente durante le fasi preparatorie della *Sentiment Analysis* attraverso l'uso di *Amazon Translate*, risiede nella traduzione e trascrizione automatica dei contenuti multimediali. La combinazione di Transcribe e degli algoritmi di traduzione di Translate abilita, infatti, la generazione automatizzata di sottotitoli accurati partendo da audio in lingue differenti. Questa sinergia

riduce i costi e i tempi per la traduzione e trascrizione dei contenuti, rendendo i materiali, i video e le conferenze accessibili globalmente in modo del tutto automatizzato.

In secondo luogo, l'applicazione della *Computer Vision* rappresenta un'innovazione nell'ambito della logistica e della sicurezza aziendale. La notevole capacità dei modelli *built-in* di operare con stabilità sia in condizioni diurne che notturne consente lo sviluppo di sistemi di monitoraggio attivi H24. Questa flessibilità si estende dalla videosorveglianza perimetrale orientata alla rilevazione automatica di intrusi, fino ai sistemi di sicurezza attiva per il tracciamento dello stato d'attenzione dei conducenti alla guida. Inoltre, qualora l'infrastruttura venga opportunamente addestrata per operare in scenari densi e complessi, la tecnologia si dimostra importante sia nella grande distribuzione, per monitorare in tempo reale l'esaurimento dei prodotti sugli scaffali dei supermercati, sia nel tracciare l'utilizzo non autorizzato di personaggi famosi all'interno di contenuti multimediali. Le aziende possono, così, automatizzare il controllo dei beni in magazzino o la vigilanza delle infrastrutture, riducendo la necessità di una supervisione visiva umana particolarmente fallibile. L'operatore viene, di conseguenza, trasformato da un controllore attivo a un gestore passivo, chiamato a intervenire esclusivamente in corrispondenza delle anomalie segnalate dalla piattaforma cloud.

Un'ultima fondamentale applicazione commerciale risiede nella gestione automatizzata del rispetto delle normative, in linea con gli standard europei del GDPR. La capacità dei servizi di individuare e mascherare istantaneamente le informazioni sensibili (*PII*) all'interno di recensioni scritte, immagini o dialoghi rappresenta uno strumento chiave per la sicurezza aziendale. L'integrazione di queste logiche nei gestionali aziendali assicura che i dati sensibili di utenti e dipendenti vengano intercettati e anonimizzati prima di essere memorizzati nei database interni. In questo modo si azzerà quasi completamente il rischio di divulgazione di dati, proteggendo l'azienda sia da attacchi informatici, sia da pesanti sanzioni legali ed economiche.

6.3 Strategie per mitigare i rischi e massimizzare i benefici

L'integrazione di sistemi di Intelligenza Artificiale basati su architetture cloud *SaaS* offre, come fatto notare, notevoli vantaggi competitivi, ma, allo stesso tempo, crea rischi operativi derivanti dai limiti dei modelli pre-addestrati. Per massimizzare i benefici derivanti dall'uso di strumenti come Amazon Comprehend, Rekognition e Transcribe, è fondamentale che gli ingegneri software aziendali adottino specifiche strategie di mitigazione e di pre-elaborazione del dato precedentemente e successivamente alle chiamate API, trasformando i modelli generali in soluzioni robuste e tolleranti all'errore.

Per quanto riguarda i rischi e le vulnerabilità legate ad Amazon Comprehend, la principale criticità emersa è l'impossibilità dell'estrazione delle informazioni sensibili (*PII*) se il sistema viene applicato ad una lingua differente dall'inglese. Per sopperire a questo vincolo linguistico e riuscire ad utilizzare questa funzionalità in tutte le lingue, una strategia utilizzabile, come fatto nel case study, prevede l'inserimento di un modulo intermedio di traduzione (sfruttando servizi come *Amazon Translate*) direttamente nell'input. Tuttavia, all'interno di questo processo di mascheramento, è assolutamente necessario implementare un controllo su ciò che viene effettivamente censurato e ciò che viene ignorato dal sistema. Il processo di traduzione automatica da una lingua all'inglese introduce, infatti, il rischio che alcune informazioni sensibili vengano interpretate erroneamente e scambiate per parole comuni, oppure codici burocratici appartenenti ad un altro paese potrebbero non essere intercettati perchè non riconosciuti tra quelli standard. Una volta tradotto il testo dalla lingua desiderata all'inglese, si fornisce la possibilità a Comprehend di operare nella lingua standard su cui è stato progettato, procedendo, poi, all'importazione delle etichette di mascheramento nel testo originale. Inoltre,

per contrastare le imprecisioni dovute all'ironia o alle recensioni ambigue, è consigliabile affiancare ai modelli delle logiche di validazione umana a campione, addestrare gli algoritmi con dataset evidenziati come sarcastici, e magari, aggiungere al processo degli algoritmi che evidenzino, per quanto possibile, i testi comprensivi di aggettivi troppo discordanti e accostamenti di parole estremamente positive come ("*fantastico*" e "*geniale*") a punteggi di recensioni molto bassi.

Nell'ambito della *Computer Vision* gestita da Amazon Rekognition, i test condotti sugli scaffali dei supermercati e sui volti di profilo hanno evidenziato la vulnerabilità del sistema davanti ad alta densità visiva e occlusioni. Per superare tali difficoltà, l'azienda non deve limitarsi ai modelli generali, ma deve sfruttare la flessibilità della piattaforma cloud tramite le *Custom Label*. Addestrare il sistema con un dataset relativo al determinato settore permette di migliorare la sensibilità dell'algoritmo al *pattern* specifico del magazzino o del punto vendita. Inoltre, l'adozione di telecamere dedicate a ogni singolo ripiano, l'integrazione di sensori di peso e di prossimità, l'esecuzione di convalide incrociate e la mappatura dello scaffale vuoto basata sull'analisi del "prima e dopo", rappresentano soluzioni ideali per aumentare i benefici della *Computer Vision*. L'insieme di queste tecniche, infatti, permette di ridimensionare il compito richiesto al sistema: l'algoritmo non è più costretto a interpretare e identificare autonomamente uno scaffale affollato con oggetti ambigui o confusi, ma deve esclusivamente verificare se lo spazio predefinito e assegnato a uno specifico codice a barre risulti vuoto o se mostri il colore dello sfondo dello scaffale.

Relativamente al monitoraggio biometrico del volto e ai sistemi di sicurezza stradale, i rischi di crollo della precisione dovuti a cambi di inquadratura o barriere fisiche (come una mano davanti alla bocca) possono essere ridotti a livello hardware e software. Da un lato, sarebbe utile utilizzare sensori ad infrarossi capaci di penetrare occlusioni come occhiali da sole, abbinati a diverse inquadrature dell'autista. Dall'altro lato, a livello software, sarebbe utile utilizzare ulteriori modelli per stimare la posa della testa nei casi di occlusione più ostici e, infine, applicare una sinergia tra i dati visivi e i parametri di guida del veicolo, come gli sbandamenti di corsia.

Per quanto riguarda *Celebrity Recognition*, è sufficiente addestrare e mantenere aggiornato costantemente il dataset, in modo tale che l'algoritmo sia capace di riconoscere qualsiasi volto, e in casi di ambiguità restituire i personaggi più probabili.

Infine, le limitazioni emerse nei test di Amazon Transcribe, come la degradazione dell'output in presenza di forti stress acustici, richiedono contromisure specifiche. Da un lato, è necessario pulire il segnale acustico grezzo attraverso microfoni multipli e l'applicazione di filtri di Wiener¹ (per azzerare i rumori di fondo costanti). Dall'altro lato, è utile far intervenire modelli linguistici avanzati (i cosiddetti LLM) che letteralmente "indovinano" le parole mancanti: quando un suono viene cancellato, il sistema non si blocca ma calcola la massima probabilità statistica della parola corretta basandosi sul contesto logico della frase.

Alcune delle soluzioni appena esaminate sono già implementate nei moderni contesti aziendali; esse, spesso senza che ce ne accorgiamo, ottimizzano le attività di tutti i giorni. Proprio la capacità di questi strumenti di diffondersi con facilità introduce la necessità di guardare oltre la configurazione tecnica, aprendo il dibattito su come cambierà nel lungo termine il rapporto tra l'uomo e l'automazione con i vari sistemi di Intelligenza Artificiale.

¹È un algoritmo statistico avanzato che calcola il livello di pulizia di ogni frequenza, abbassando il volume solo dei canali troppo disturbati e lasciando intatta la voce.

6.4 Prospettive future della cooperazione tra uomo e Intelligenza Artificiale

Il rapido sviluppo dell'Intelligenza Artificiale ha acceso, ormai da qualche tempo, un dibattito sul futuro dell'uomo e dei sistemi intelligenti. Oggi è fondamentale superare l'idea che le macchine sostituiranno in massa l'essere umano, per concentrarsi, invece, su come la tecnologia stia concretamente trasformando il mercato del lavoro.

Sebbene i casi studio analizzati mostrino l'efficacia dell'Intelligenza Artificiale applicata a strumenti separati, le tendenze evolutive più recenti suggeriscono un superamento di questa frammentazione. Il futuro della cooperazione uomo-macchina si direziona, infatti, verso lo sviluppo di un unico assistente linguistico, capace di coordinare e sintetizzare funzionalità differenti attraverso il linguaggio naturale.

Un interessante spunto di riflessione sull'argomento emerge da un recente studio pubblicato da Anthropic "Massenkoff e McCrory [2026]", il quale offre una chiave di lettura innovativa per misurare l'impatto economico dei modelli linguistici. La ricerca mostra una netta distinzione tra l'applicazione teorica dell'Intelligenza Artificiale, ovvero ciò che gli algoritmi sarebbero potenzialmente in grado di fare a livello di automazione dei singoli task, e l'applicazione osservata, che quantifica, invece, l'effettiva adozione di queste tecnologie all'interno dei processi aziendali. Le evidenze emerse mostrano un quadro controintuitivo rispetto alle storiche rivoluzioni industriali, svelando come i profili professionali a maggior rischio di automazione appartengano alla categoria dei lavori d'ufficio e intellettuali, come mostrato in Figura 6.1.

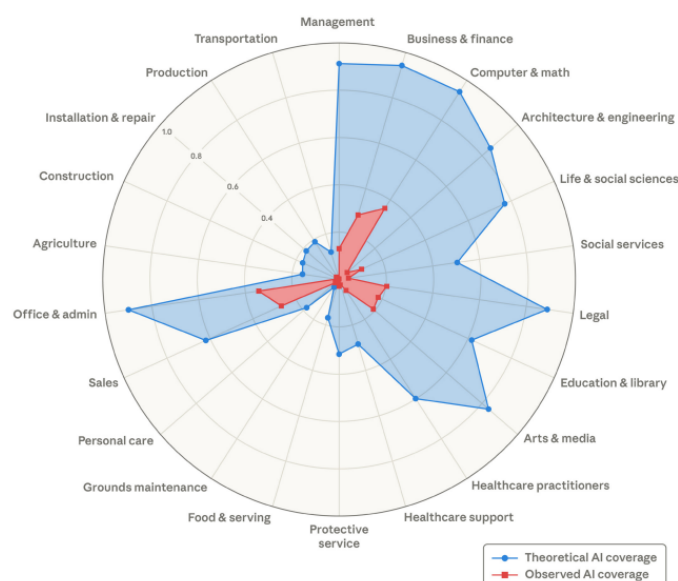


Figura 6.1: Illustrazione delle mansioni lavorative che gli LLM potrebbero teoricamente eseguire (blu), confrontata con la stima di quante professioni siano effettivamente impattate (rosso), basata sui dati reali di utilizzo (fonte: anthropic.com)

I dati in Figura 6.1 evidenziano come settori quali la finanza, il marketing, l'amministrazione, l'ambito legale e la stessa programmazione software registrino indici di esposizione osservata estremamente elevati. È possibile notare e riflettere su un paradosso: esiste una correlazione quasi diretta tra gli anni che un individuo ha dedicato agli studi e il grado di sostituibilità del proprio lavoro da parte dei modelli linguistici. Le professioni che richiedono lauree tradizionalmente ritenute prestigiose risultano le più esposte, poiché gran parte delle loro competenze centrali si basa sulla redazione, sintesi, ricerca di pattern e interpretazione

di testi non strutturati, tutti settori in cui i modelli cloud mostrano capacità straordinarie. Al contrario, lo studio dimostra che i lavori legati al settore manuale e relazionale, come le costruzioni, la manutenzione, l'agricoltura rurale, la ristorazione o i trasporti, registrano un'esposizione pressoché nulla, per via di un'incapacità empatica e fisica delle macchine. Questo scenario mostra una potenziale inversione dell'importanza economica e sociale delle competenze, dove le abilità artigianali e relazionali potrebbero configurarsi come valori fondamentali, simbolo di unicità professionale, per le quali il professionista dovrebbe essere scelto rispetto alla concorrenza.

Il cuore della ricerca di Anthropic risiede, tuttavia, nella comprensione delle differenze esistenti tra la capacità teorica e l'effettivo utilizzo sul mercato dei vari sistemi intelligenti. Molteplici settori, come quello legale, mostrano un potenziale di automazione altissimo a differenza di un utilizzo effettivo ancora parziale. Questa divergenza non deve portare a una falsa percezione di sicurezza; le aziende, per quanto mostrato, si trovano oggi in una fase lenta di transizione e riprogettazione interna, precedente a un'imminente accelerazione. I primi segnali di questa transizione si manifestano non tanto in un picco immediato della disoccupazione, ma in un progressivo rallentamento delle nuove assunzioni dei giovani lavoratori e dei neolaureati nei settori esposti. L'Intelligenza Artificiale, quindi, non sta agendo come un distruttore istantaneo di occupazione, ma come uno scrupoloso selettore di competenze all'interno del quale la variabile dell'ottenere un posto di lavoro non è legata alla professione bensì alla capacità del lavoratore di evolvere le proprie competenze, talvolta servendosi dell'IA.

Per via di tutti i motivi sopra delineati, la sopravvivenza professionale, come fu per l'introduzione delle macchine all'interno delle fabbriche, è legata alla transizione dei professionisti da esecutori di attività a supervisor di processo. Il messaggio che si evince, quindi, dalle prospettive future della cooperazione uomo-IA risiede nella necessità di ripensare e riprogettare costantemente il valore della propria attività lavorativa, in quanto, come ben noto, questi sistemi imparano ed evolvono molto meglio dell'umano. Chi sceglierà di non accettare questo scenario andrà incontro a un'inevitabile decadenza, venendo sostituito non direttamente dall'algoritmo, ma da professionisti capaci di sfruttare i modelli cloud come moltiplicatori delle loro competenze.

Questa transizione impone, tuttavia, una profonda riflessione sulla natura stessa dell'intelligenza e sull'incredibile accelerazione evolutiva a cui stiamo assistendo. Se la natura biologica ha impiegato circa 4,5 miliardi di anni di evoluzione, imparando dagli errori, per modellare il cervello dell'Homo Sapiens, l'elettronica e l'informatica sono riuscite a compiere un salto paragonabile a quello dell'evoluzione umana in appena 100 anni di storia. Funzionando allo stesso modo del cervello, come mostrato nella Sezione 1.3.3, le reti neurali nel tempo sono state addestrate con un'enorme mole di dati provenienti dall'era dell'inizio di internet ad oggi. Rispetto alla mente umana, capace di processare una frazione limitata di informazioni al secondo, i moderni sistemi sono capaci di analizzare milioni di dati nello stesso intervallo temporale, migliorando costantemente i propri modelli predittivi grazie al feedback continuo restituito dagli utenti di tutto il mondo. Ci troviamo, quindi, senza dubbio, di fronte a un traguardo tecnologico straordinario.

Questo scenario porta ad abbandonare la vecchia metrica secondo la quale l'intelligenza era misurabile quantitativamente tramite un singolo valore. Come teorizzato dallo psicologo *Howard Gardner*, nel libro "*Gardner [2011]*", attraverso il modello delle intelligenze multiple, l'intelletto non può essere ridotto a un singolo punteggio numerico, ma deve essere mappato come una ragnatela multidimensionale che include vari tipi di intelligenza, come quella creativa e quella empatica, come mostrato in Figura 6.2. Competere direttamente contro macchine composte da miliardi di circuiti sull'intelligenza puramente logico-matematica è una battaglia persa a prescindere, poiché, in quel settore, l'automazione ha già superato le

capacità umane. La vera sfida, e allo stesso tempo innovazione evolutiva, risiede nell'atto della delega: incaricando l'Intelligenza Artificiale dei compiti ripetitivi e logici, l'uomo ha l'opportunità unica di riappropriarsi delle restanti dimensioni cognitive e relazionali.

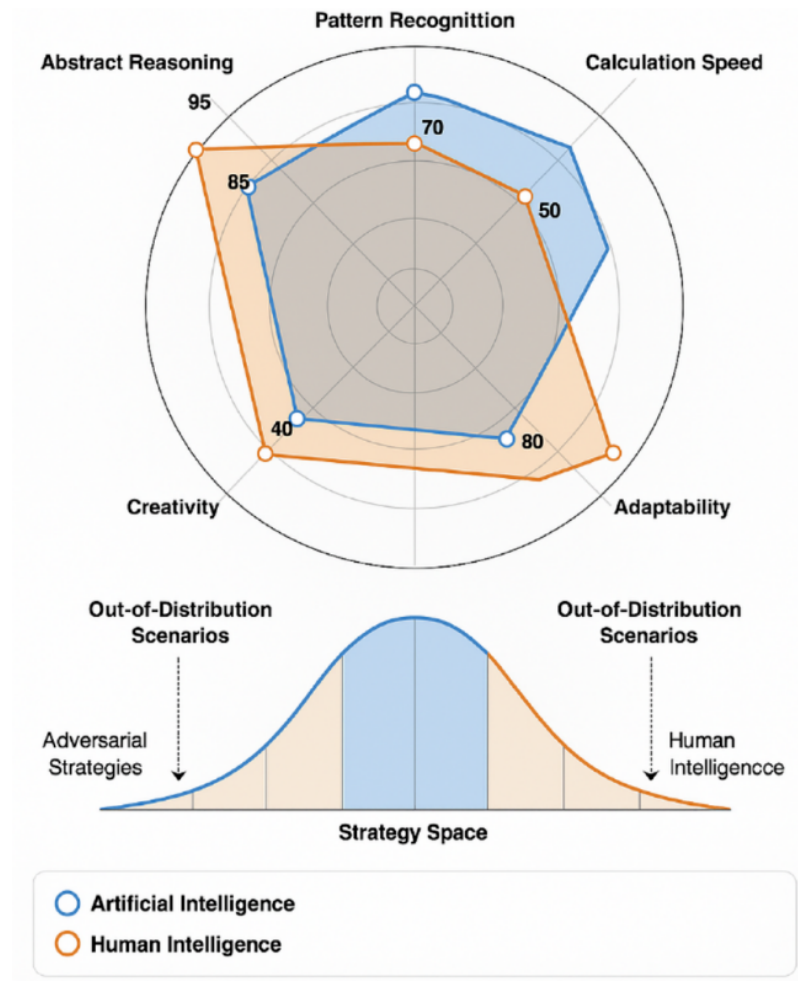


Figura 6.2: Illustrazione dell'intelligenza a ragnatela (fonte: researchgate.net)

Giunti al termine della tesi, possiamo affermare di aver pienamente conseguito gli obiettivi che ci eravamo inizialmente proposti. Infatti, per quanto riguarda la comprensione del funzionamento teorico e pratico dell'Intelligenza Artificiale, abbiamo inizialmente delineato un'ampia panoramica teorica incentrata sui fondamenti tecnici, sugli approcci algoritmici e sulle logiche di funzionamento di tale disciplina, ponendo le basi concettuali necessarie per comprendere i meccanismi del *Deep Learning* e delle reti neurali profonde. Una volta compresi questi meccanismi, con l'obiettivo di rimanere costantemente aggiornati sulle tipologie di soluzioni che l'Intelligenza Artificiale riesce a creare, ci si è concentrati sulla genesi, sulla struttura architeturale e sui principali servizi offerti dalla piattaforma Amazon Web Services (AWS), per poi, mediante la progettazione e l'esecuzione di specifici *case study*, analizzare i reali punti di forza e le limitazioni strutturali dei tre servizi selezionati, ossia Amazon Comprehend, Amazon Rekognition e Amazon Transcribe.

In questo contesto, i tre servizi sopra riportati hanno costituito il centro del lavoro svolto. Infatti, tramite il case study relativo al servizio Amazon Comprehend, abbiamo delineato la *Sentiment Analysis*, abbiamo definito i suoi aspetti tecnici e abbiamo dimostrato come tale piattaforma cloud implementi l'elaborazione automatica del linguaggio naturale per l'interpretazione di testi e per il monitoraggio dei contenuti testuali che ogni giorno vengono pubblicati online. Per quanto riguarda Amazon Rekognition, tramite il suo approfondimento, ci è stato possibile studiare e presentare l'*Image Recognition*, accostandola, anche, ai suoi aspetti più tecnici e alle sue applicazioni più attuali. Questo servizio si è dimostrato in grado di individuare oggetti, scenari, espressioni facciali e personaggi celebri nonché di monitorare, tramite videocamere di sorveglianza, perimetri critici. Infine, l'analisi dei processi di trascrizione e delle chiamate dei *call center*, effettuate tramite Amazon Transcribe, ci ha permesso di definire lo *Speech-To-Text*, i suoi fondamenti tecnici e le sue applicazioni moderne. Parallelamente, tale servizio ci ha permesso di mostrare come l'Intelligenza Artificiale possa essere inserita all'interno dei contesti aziendali e quotidiani per automatizzarne le attività più ripetitive e standardizzate.

Relativamente ai risultati ottenuti dai *case study*, abbiamo evidenziato come gli strumenti analizzati siano eccellenti, sebbene presentino alcune limitazioni che risultano, comunque, superabili attraverso configurazioni più profonde e un addestramento mirato dei servizi. Abbiamo, infatti, dimostrato come l'accuratezza di alcuni microservizi decadesse a livello prestazionale nel momento in cui i servizi stessi venivano messi alla prova con situazioni complesse e ambigue o con file di *input* disturbati e di scarsa qualità.

Allargando lo sguardo all'evoluzione complessiva del settore, le prospettive future ci suggeriscono traguardi che ridefiniranno il rapporto tra la tecnologia e la società. Da un lato,

la ricerca si sta muovendo verso il superamento dei modelli pre-addestrati a favore di sistemi di Intelligenza Artificiale Generale (AGI). Dall'altro lato, la vera sfida dei prossimi decenni risiederà nell'unione tra i modelli di *Machine Learning* e l'hardware dei computer quantistici, una sinergia strategica capace di annullare i colli di bottiglia legati alla latenza e alla velocità di elaborazione delle informazioni.

Essendo ormai l'Intelligenza Artificiale la nuova infrastruttura del futuro, come avvenne storicamente per l'avvento dell'elettricità e di Internet, l'auspicio finale di questo percorso di ricerca risiede nella speranza che tale tecnologia non venga più percepita come un limite per l'essere umano. Configurandosi come il primo strumento mai concepito dalla scienza capace di prendere decisioni e sviluppare idee in autonomia, essa deve essere interpretata come il più potente acceleratore di competenze oggi a disposizione dell'uomo. Un'ulteriore sfida del futuro non riguarderà la capacità degli algoritmi di replicare le funzioni cognitive, ma bensì la volontà della società e dei futuri professionisti di sviluppare quelle dimensioni che nessuna macchina potrà mai sostituire, ovvero l'empatia profonda, l'intuizione creativa, il giudizio etico e la ricerca spirituale. Delegare la componente puramente ripetitiva e logico-matematica all'automazione, pur mantenendo rigorosamente un grado di supervisione umana, deve diventare l'occasione per riflettere e riappropriarsi di quei valori profondi che oggi l'uomo, sopraffatto dai ritmi frenetici della vita, non ha il tempo di coltivare. In questo nuovo scenario tecnologico, la macchina rimarrà un dispositivo volto a ottimizzare la produttività, mentre l'uomo sarà di nuovo l'unico a far progredire la società e comprendere ciò che lo circonda.

- BROSSA, C. (2026), *ANTROPOTECH: 30 leggi universali per l'umanesimo tecnologico*, Libri d'Impresa, disponibile su Amazon.it.
- BRYNJOLFSSON, E. e MCAFEE, A. (2017), «The business of artificial intelligence», *Harvard business review*, vol. 7 (1), p. 1–2.
- FORTUNA, P. e NUNES, S. (2018), «A survey on automatic detection of hate speech in text», *Acm Computing Surveys (Csur)*, vol. 51 (4), p. 1–30.
- GARDNER, H. (2011), *Frames of mind: The theory of multiple intelligences*, Basic books. (Cited at page 106)
- LECUN, Y., BENGIO, Y. e HINTON, G. (2015), «Deep learning», *nature*, vol. 521 (7553), p. 436–444.
- LIU, B. (2020), *Sentiment analysis: Mining opinions, sentiments, and emotions*, Cambridge university press.
- MASSENKOFF, M. e MCCRORY, P. (2026), «Labor market impacts of AI: A new measure and early evidence», URL <https://www.anthropic.com/research/labor-market-impacts>. (Cited at page 105)
- MURPHY, K. P. (2012), *Machine learning: a probabilistic perspective*, MIT press.
- ORTEGA, J. D., KOSE, N., CAÑAS, P., CHAO, M.-A., UNNERVIK, A., NIETO, M., OTAEGUI, O. e SALGADO, L. (2020), «DMD: A Large-Scale Multi-modal Driver Monitoring Dataset for Attention and Alertness Analysis», in BARTOLI, A. e FUSIELLO, A., curatori, «Computer Vision – ECCV 2020 Workshops», p. 387–405, Springer International Publishing. (Cited at page 66)
- RUSSELL, S., NORVIG, P. e INTELLIGENCE, A. (1995), «A modern approach», *Artificial Intelligence. Prentice-Hall, Egnlewood Cliffs*, vol. 25 (27), p. 79–80.
- SULEYMAN, M. (2023), *The coming wave: technology, power, and the twenty-first century's greatest dilemma*, Crown.
- SZELISKI, R. (2022), *Computer vision: algorithms and applications*, Springer Nature.
- TEGMARK, M. (2018), *Life 3.0: Being human in the age of artificial intelligence*, Vintage.

- TURING, A. M. (2007), «Computing machinery and intelligence», in «Parsing the Turing test: Philosophical and methodological issues in the quest for the thinking computer», p. 23–65, Springer.
- VASWANI, A., SHAZEER, N., PARMAR, N., USZKOREIT, J., JONES, L., GOMEZ, A. N., KAISER, Ł. e POLOSUKHIN, I. (2017), «Attention is all you need», *Advances in neural information processing systems*, vol. 30.
- WANKHADE, M., RAO, A. C. S. e KULKARNI, C. (2022), «A survey on sentiment analysis methods, applications, and challenges», *Artificial intelligence review*, vol. 55 (7), p. 5731–5780.

- AI4business – www.ai4business.it
- apprendoo – www.apprendoo.com
- AWS Documentation – docs.aws.amazon.com
- AWS, Amazon Web Services – aws.amazon.com
- botpress – botpress.com
- Cloudflare – www.cloudflare.com
- Commissione europea – health.ec.europa.eu
- Covisian – covisian.com
- DataMasters – datamasters.it
- DaviAI – www.davi-ai-web.com
- ENCORD – encord.com
- Epiclink, your service provider – epiclink.it
- IBM, International Business Machines – www.ibm.com
- Innovaformazione, Informatica Specialistica – innovaformazione.net
- ISI – www.isipc.it
- NetApp – www.netapp.com
- NetCom GROUP – www.netcomgroup.eu
- Nutanix next – www.nutanix.com
- OKpedia – www.okpedia.it
- Parlamento europeo – www.europarl.europa.eu
- Performance Trading – www.performancetrading.it
- PMF Research – pmf-research.eu

- Ranstad – digitalcontent.randstad.it
- Rete del dono – www.retedeldono.it
- Rivista di Neurologia – www.rivistadineurologia.it
- scientific reports – www.nature.com
- skilla – www.skilla.com
- Skillbook – www.skillbook.it
- topnetwork – www.topnetwork.it
- UC Berkeley Electrical Engineering & Computer Sciences – people.eecs.berkeley.edu
- UNIBO, Laboratory of Advanced Research on Computer Science – lia.disi.unibo.it
- UniverseIT Gruppo OTS – universeit.blog
- WEB LEADERS, winning people – webleaders.it
- Wikipedia – it.wikipedia.org
- YouTube – www.youtube.com

Ringraziamenti

Desidero, prima di tutto, esprimere la mia più profonda gratitudine al Professor Domenico Ursino, per la sua costante guida, la disponibilità, le correzioni e i preziosi consigli che hanno reso possibile la realizzazione di questa tesi.

Se ripenso a quando ho iniziato questo percorso, ricordo che vedevo questo traguardo lontanissimo, quasi uno scherzo a pensare che ci sarei riuscito davvero. Come alcuni sanno, sopraffatto dalle difficoltà e dalle forti emozioni che ho vissuto, mi è capitato più volte di pensare di fermarmi o di prendermi una pausa; tuttavia, quello che credo mi contraddistingua maggiormente è la resilienza che riesco a dimostrare a fronte di queste situazioni. Piano piano, passo dopo passo, esame dopo esame, l'obiettivo si faceva sempre più chiaro e, non mentirò, da solo, senza il supporto di tutte le persone a me care, probabilmente, anche la mia resilienza avrebbe ceduto. Desidero, per questo motivo, ringraziare tutte le persone che mi hanno affiancato in questo percorso: coloro che sono rimasti al mio fianco fin dall'inizio e chi, invece, per fortuna, è arrivato lungo la strada.

*Ai miei genitori,
Vi ringrazio per tutti i sacrifici che avete fatto per la mia felicità,
Vi prometto che non appena ne avrò la possibilità, ricambierò ogni vostro singolo sacrificio.*

*A Gerry, mio pezzo di cuore più grande,
Per tutte le volte che hai dormito serenamente a pancia in sù al mio fianco mentre studiavo,
Per tutte le volte che con il tuo faccino mi facevi capire di aver bisogno di una pausa,
Questo traguardo lo dedico soprattutto a te, che con amore puro mi sei sempre stato vicino,
Oggi sono felice di dirti: "ce l'abbiamo fatta Gerry!"*

*A tutti i miei amici più grandi che oggi sono qui.
Grazie per aver alleggerito i momenti più intensi, trasformando ogni piccola distrazione in un ricordo prezioso.*

A tutti coloro che hanno creduto in me dal primo giorno.

*A me stesso,
Alla determinazione e alla costanza assidua nel raggiungere e superare i miei obiettivi,
Per l'impegno, la resilienza e l'ambizione che mi appartengono,
Alle sfide che ho combattuto e contro cui combatto ogni giorno in silenzio,
Alla continua scoperta di me e delle mie parti più profonde,*

*Alla mia continua ricerca di consapevolezza,
Alla mia empatia,
Dico con orgoglio e umiltà:
Sono fiero di me, non perché sono arrivato,
Ma perché non mi sono mai arreso.
Che io possa raggiungere tutti i miei obiettivi, superare tutte le mie paure e costruire il futuro che merito.*