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 case study relativi all'Intelligenza Artificiale: Riconoscimento di Immagini, Speech-To-Text e Chatbot

Design and implementation of case studies related to Artificial Intelligence: Image Recognition, Speech-To-Text and Chatbots

Relatore

Prof. Domenico Ursino

Candidata

Letizia Lucrezia Mulieri

ANNO ACCADEMICO 2025-2026

E però non c'è dubbio che i progressi della ragione
e lo spegnimento delle illusioni producono la barbarie,
e un popolo oltremodo illuminato non diventa mica civilissimo,
come sognano i filosofi del nostro tempo, la Staël ec., ma barbaro;
al che noi c'incamminiamo a gran passi e quasi siamo arrivati. [.....]
La ragione è un lume: la natura vuol essere illuminata dalla ragione, non incendiata.

Giacomo Leopardi, Zibaldone

Sommario

Negli ultimi anni, l'Intelligenza Artificiale sta acquisendo un ruolo sempre più rilevante all'interno della nostra quotidianità. I principali provider di servizi di cloud computing stanno dedicando un'attenzione sempre maggiore alla distribuzione di prodotti che includono funzionalità di Intelligenza Artificiale. Nella presente trattazione vengono illustrati alcuni case study realizzati mediante i servizi offerti dal provider Amazon Web Services. Per ciascuno dei servizi analizzati viene proposta un'analisi, finalizzata a evidenziarne gli aspetti positivi, le potenziali problematiche e i limiti operativi. In particolare, è stato innanzitutto approfondito Amazon Rekognition, un servizio dedicato al riconoscimento delle immagini; nello specifico, è stata analizzata la funzionalità di riconoscimento del testo nelle immagini. Successivamente, è stato esaminato il servizio Amazon Transcribe, dedicato allo Speech-to-Text. In conclusione, è stato approfondito Amazon Lex, un servizio per la realizzazione di chatbot.

Keyword: Intelligenza Artificiale (IA), Machine Learning (ML), Cloud computing, Amazon Web Services (AWS), Riconoscimento ottico dei caratteri (OCR), Speech-to-Text, Chatbot

Indice

Introduzione **1**

1 Introduzione all'Intelligenza Artificiale **3**
1.1 Definizione di Intelligenza Artificiale . . . 3
1.1.1 L'approccio del test di Turing . . . 4
1.1.2 L'approccio del modello cognitivo . . . 4
1.1.3 L'approccio delle "Leggi del pensiero" . . . 5
1.1.4 L'approccio dell'agente razionale . . . 5
1.1.5 Macchine vantaggiose . . . 6
1.2 Classificazione . . . 6
1.2.1 Tipologie di Intelligenza Artificiale basata sulle capacità . . . 7
1.2.2 Tipologie di Intelligenza Artificiale basata sulle funzionalità . . . 10
1.3 Machine Learning e modelli di apprendimento . . . 13
1.3.1 Algoritmi di apprendimento . . . 14
1.4 Deep Learning . . . 16
1.5 Stato dell'arte . . . 17
1.6 Rischi e benefici dell'Intelligenza Artificiale . . . 19

2 Amazon Web Services **21**
2.1 Introduzione ad Amazon Web Services . . . 21
2.2 Introduzione al cloud computing . . . 22
2.2.1 Livelli e tipologie di cloud . . . 22
2.2.2 Istanza nel cloud computing . . . 24
2.3 Infrastruttura globale di AWS . . . 25
2.3.1 Architettura della rete globale di AWS . . . 25
2.3.2 Sicurezza e affidabilità della rete AWS . . . 25
2.3.3 Prestazioni della rete globale ed efficienza energetica . . . 26
2.3.4 Soluzioni per l'infrastruttura . . . 26
2.3.5 Posizionamento di AWS nel mercato del cloud . . . 27
2.4 Servizi di AWS . . . 29
2.4.1 Accesso ai servizi AWS . . . 32
2.5 Servizi AWS per l'Intelligenza Artificiale . . . 33
2.5.1 Esempi di utilizzi di AWS per l'Intelligenza Artificiale . . . 35

3 Un case study relativo al riconoscimento di immagini **38**
3.1 Riconoscimento di immagini . . . 38
3.1.1 Difficoltà nel riconoscimento di immagini . . . 39
3.1.2 Casi d'uso di riconoscimento di immagini . . . 40
3.2 Riconoscimento ottico dei caratteri . . . 41
3.2.1 Funzionamento dell'OCR . . . 41
3.2.2 Tipologie di OCR . . . 42
3.2.3 Casi d'uso dell'OCR . . . 43
3.3 Amazon Rekognition . . . 44
3.3.1 Funzionamento di Amazon Rekognition . . . 45
3.4 Case study relativi all'individuazione di testo nelle immagini . . . 46
3.4.1 Immagini con testo chiaramente leggibile . . . 47
3.4.2 Immagini con testo leggibile . . . 50
3.4.3 Immagini con testo presente su cartelli stradali . . . 53
3.4.4 Immagini con testo presente su insegne . . . 56
3.4.5 Immagini con testo presente su locandine di film . . . 58
3.4.6 Immagini con testo presente su graffiti . . . 58
3.4.7 Immagini con testo caratterizzato da aspetti particolari . . . 60

4 Un case study relativo allo Speech-to-Text **64**
4.1 Introduzione allo Speech-to-Text . . . 64
4.1.1 Funzionamento dello Speech-to-Text . . . 64
4.1.2 Difficoltà nel riconoscimento del linguaggio parlato . . . 66
4.1.3 Casi d'uso dello Speech-to-Text . . . 67
4.2 Amazon Transcribe . . . 67
4.2.1 Funzionamento di Amazon Transcribe . . . 67
4.2.2 Funzionamento di Amazon S3 . . . 72
4.3 Case study relativi alla trascrizione del linguaggio parlato . . . 72
4.3.1 Criteri di analisi e organizzazione dei risultati . . . 74
4.3.2 Trascrizione in condizioni standard . . . 74
4.3.3 Gestione di contenuti multilingua . . . 77
4.3.4 Analisi di condizioni acustiche particolari . . . 80
4.3.5 Influenza degli accenti sulla trascrizione . . . 86
4.3.6 Gestione del linguaggio specialistico . . . 88
4.3.7 Trascrizione di contenuti musicali . . . 92
4.3.8 Analisi di un audio ad elevata complessità . . . 94

5 Un case study relativo ai chatbot **99**
5.1 Introduzione ai chatbot . . . 99
5.1.1 Tipologie di chatbot . . . 99
5.1.2 Funzionamento dei chatbot . . . 101
5.1.3 Sfide e rischi dei chatbot . . . 102
5.1.4 Casi d'uso dei chatbot . . . 103
5.2 Amazon Lex . . . 103
5.2.1 Integrazione con altri servizi AWS . . . 103
5.2.2 Funzionamento di Amazon Lex . . . 104
5.3 Case study relativo alla creazione di un chatbot . . . 106
5.3.1 Criteri di analisi . . . 106
5.3.2 Architettura adottata . . . 107
5.3.3 Configurazione del chatbot . . . 109
5.3.4 Flusso di conversazione . . . 109

5.3.5 Gestione degli intent 111
5.3.6 Funzioni Lambda 120
5.3.7 Dynamo DB 122
5.3.8 Integrazione del chatbot in un'applicazione web 124
5.3.9 Verifica e valutazione del chatbot 126

6 Discussione **130**
6.1 Discussione in merito ai case study 130
6.1.1 Discussione in merito al riconoscimento di immagini 130
6.1.2 Discussione in merito al riconoscimento vocale automatico 133
6.1.3 Discussione in merito alla realizzazione di un chatbot 136

Conclusioni **140**

Bibliografia **142**

Sitografia **144**

Ringraziamenti **145**

Elenco delle figure

1.1 Diagramma di Venn che mostra la relazione fra le diverse discipline dell'IA (fonte: https://www.machinedesign.com/) 14
1.2 Immagine che mostra la risposta di Midjourney al prompt: "un'immagine iperrealistica di Harry Potter" (fonte: https://hai.stanford.edu/) 18

2.1 Servizi di Amazon Web Services (fonte: https://aws.plainenglish.io/) 21
2.2 Principali differenze tra i servizi cloud (fonte: https://www.capterra.it) 24
2.3 Componenti principali di un'istanza Amazon EC2 e struttura della denominazione di un tipo di istanza (fonte: https://docs.aws.amazon.com) 25
2.4 Mappa della rete globale AWS (fonte: https://aws.amazon.com/) 26
2.5 Magic Quadrant per i servizi di piattaforme cloud strategiche (fonte: https://www.gartner.com/) 28
2.6 Servizi di AWS suddivisi per categoria (fonte: https://www.acs.it/) . . . 29
2.7 Pagina iniziale della AWS Console Mobile Application (fonte: https://aws.amazon.com/it/console/mobile/) 33
2.8 Relazione gerarchica tra Intelligenza Artificiale (AI), Machine Learning (ML), Intelligenza Artificiale Generativa, Classification AI e Predictive AI (fonte: https://docs.aws.amazon.com) 35
2.9 Architettura In-Console Cloud Assistant (fonte: https://aws.amazon.com/) 36

3.1 Diagramma di flusso di una CNN (fonte: https://www.unidformazione.com/) 39
3.2 Esempio di rilevamento delle etichette con relativi risultati 45
3.3 Esempio di riconoscimento di volti celebri con relativi risultati 45
3.4 Rilevamento del testo nell'immagine contenente una citazione di *George Orwell* e relativi risultati 48
3.5 Rilevamento del testo nell'immagine contenente una citazione di *Luigi Pirandello* e relativi risultati 49
3.6 Rilevamento del testo nell'immagine contenente una scritta in lingua inglese ben visibile e relativi risultati 49
3.7 Rilevamento del testo nell'immagine contenente una pagina di libro con il componimento *L'infinito* e relativi risultati 50
3.8 Rilevamento del testo nell'immagine contenente una scritta in lingua inglese con caratteri disposti verticalmente e relativi risultati 51

3.9 Rilevamento del testo nell'immagine contenente la targa commemorativa dedicata all'autore *William Shakespeare* e relativi risultati 52
3.10 Rilevamento del testo contenuto in un'immagine raffigurante una poesia di *Giovanni Pascoli*, con ciascun verso disposto su un gradino di una scala, e relativi risultati 53
3.11 Individuazione del testo nell'immagine contenente segnaletica stradale newyorkese e relativi risultati 54
3.12 Individuazione del testo nell'immagine contenente segnaletica stradale in alfabeto arabo e relativi risultati 55
3.13 Rilevamento del testo nell'immagine contenente cartello stradale in alfabeto cirillico e relativi risultati 55
3.14 Individuazione del testo nell'immagine contenente indicazioni stradali in lingua giapponese e relativi risultati 56
3.15 Rilevamento del testo nell'immagine contenente l'insegna dell'Hard Rock Cafe di Copenhagen e relativi risultati 57
3.16 Individuazione del testo nell'immagine contenente un'insegna luminosa e relativi risultati 57
3.17 Rilevamento del testo nell'immagine contenente la locandina del film *Inception* e relativi risultati 59
3.18 Individuazione del testo nell'immagine contenente una pittura murale e relativi risultati 60
3.19 Individuazione del testo nell'immagine contenente una pittura murale con caratteri disposti in ordine inverso rispetto al normale senso di lettura e relativi risultati 60
3.20 Rilevamento del testo nell'immagine contenente una scritta composta da palloncini e relativi risultati 61
3.21 Confronto delle prestazioni di Amazon Rekognition nel riconoscimento del testo al variare del suo orientamento 61
3.22 Confronto dei risultati relativi alla Figura 3.21 62
3.23 Rilevamento del testo nell'immagine contenente una scritta distorta con effetto di dissolvenza e relativi risultati 62
3.24 Rilevamento degli elementi testuali che compongono la parola LIFE e relativi risultati 62

4.1 Schema illustrativo del processo di riconoscimento vocale automatico (`https://it.shaip.com/`) 65
4.2 Sezione *Processi di trascrizione* della console di Amazon Transcribe 68
4.3 Sezione *Dettagli processo* 68
4.4 Sezione *Anteprima della trascrizione* 69
4.5 Schermata per la selezione delle impostazioni del processo 71
4.6 Schermata relativa alla configurazione di processo e alle funzionalità opzionabili 73
4.7 Schermata di gestione degli oggetti contenuti in un bucket Amazon S3 73
4.8 QR code per l'accesso alla traccia audio in lingua inglese (fonte: `https://docs.aws.amazon.com/`) 75
4.9 QR code per l'accesso alla traccia audio in lingua inglese caratterizzata dalla presenza di due parlanti (fonte: `https://www.bbc.co.uk/`) 76
4.10 QR code per l'accesso alla traccia audio in lingua italiana (fonte: `https://www.focus.it/`) 76
4.11 QR code per l'accesso alle tracce audio caratterizzate da alternanza di lingua inglese e italiana 77

4.12 QR code per l'accesso alle tracce audio caratterizzate da multilinguismo (fonte: `https://it.babbel.com/`) . . . 79
4.13 QR code per l'accesso alla traccia audio contenente più parlanti (fonte: `https://www.youtube.com/watch?v=r8_eI-stGQc`) . . . 81
4.14 Output restituito dal servizio per l'individuazione di linguaggio potenzialmente tossico nella traccia audio . . . 82
4.15 QR code per l'accesso alla traccia audio contenente rumore di sottofondo (fonte: `https://www.youtube.com/watch?v=jCjuXt7CiIY`) . . . 84
4.16 QR code per l'accesso alla traccia audio in lingua inglese (fonte: `https://education.nationalgeographic.org`) . . . 85
4.17 QR code per l'accesso alle tracce audio caratterizzate da accento particolare . 86
4.18 QR code per l'accesso alla traccia audio in lingua inglese con accento russo . . 87
4.19 QR code per l'accesso alla traccia audio caratterizzata da linguaggio tecnico (fonte: `https://youtu.be/u7hwHh_-7_g?si=vlmbixb9FhIerf5K`) . . 89
4.20 QR code per l'accesso alla traccia audio caratterizzata da linguaggio finanziario (fonte: `https://youtu.be/DApfoOmMBFw?si=IOISIZouXdmdg7Fo`) . . . 91
4.21 QR code per l'accesso alle tracce audio caratterizzate da contenuti musicali . 93
4.22 QR code per l'accesso alla traccia audio caratterizzata dalla combinazione di diversi fattori (fonte: `https://www.motogp.com/`) . . . 95
4.23 QR code per l'accesso alla traccia audio contenente diverse scene tratte dal film *The Dark Knight* . . . 96
4.24 QR code per l'accesso ai sottotitoli generati dall'analisi della traccia audio contenente scene riprese dal film *The Dark Knight* . . . 97

5.1 Schema illustrativo del processo di funzionamento di un chatbot multi-turn task-oriented (fonte: https://www.researchgate.net/) . . . 101
5.2 Architettura impiegata per la realizzazione del bot . . . 107
5.3 Flusso logico principale del chatbot . . . 110
5.4 Sezione *Acquisizione slot: risposta al successo* con abilitazione della *Ramificazione condizionale* . . . 113
5.5 Schermata per la configurazione dei *Session attributes* . . . 115
5.6 Ramificazione condizionale *ChosenCourses* . . . 115
5.7 Ramificazione condizionale relativa all'intent *Third Year Bachelor Degree* . . . 116
5.8 Ramificazione condizionale relativa all'intent *Second Year Master Degree* . . . 118
5.9 Abilitazione dei passaggi di conversazione *Adempimento* e *Riposta di chiusura* . 119
5.10 Ramificazione condizionale relativa a *FirstYearMaster* . . . 120
5.11 Database NoSQL realizzato mediante Amazon DynamoDB . . . 124
5.12 Schema dell'architettura adottata per l'interfaccia utente (fonte: `https://github.com/aws-samples/aws-lex-web-ui`) . . . 124
5.13 Interfaccia web del chatbot . . . 126
5.14 Grafico delle prestazioni dell'intento . . . 127

Elenco delle tabelle

2.1 Principali servizi AWS organizzati per categoria funzionale 29

3.1 Principali differenze fra OCR tradizionale e ICR 43

4.1 Vocabolario personalizzato per il riconoscimento dei termini tecnici 90
4.2 Vocabolario personalizzato con termini finanziari 92
4.3 Vocabolario filtro 98

5.1 Principali intent del bot organizzati per funzione nel flusso conversazionale . 111
5.2 Principali slot del bot organizzati per funzione 112
5.3 Prestazioni delle conversazioni nelle diverse configurazioni di test 128
5.4 Prestazioni degli intent nelle diverse configurazioni di test 128
5.5 Riconoscimento degli enunciati nelle diverse configurazioni di test 128

6.1 Caratteristiche delle immagini e livelli di complessità dei casi di test (prima parte) 132
6.3 Risultati di rilevamento e trascrizione del testo per i casi di test 134
6.4 Caratteristiche del parlato e livelli di complessità dei casi di test (prima parte) 136
6.6 Risultati di rilevamento e trascrizione del parlato per i casi di test 138

Elenco delle trascrizioni

4.1 Trascrizione della traccia audio in lingua inglese . . . 75
4.2 Trascrizione della traccia audio in lingua inglese caratterizzata dalla presenza di due parlanti . . . 76
4.3 Trascrizione della traccia audio in lingua italiana . . . 77
4.4 Trascrizione di una traccia audio caratterizzata dall'alternanza fra italiano e inglese . . . 78
4.5 Trascrizione di una traccia audio caratterizzata dall'alternanza fra inglese e italiano con rilevamento automatico della lingua . . . 79
4.6 Trascrizione di una traccia audio contenente più lingue specificate esplicitamente 80
4.7 Trascrizione di una traccia audio contente più lingue rilevate automaticamente 80
4.8 Trascrizione di una traccia audio contenente numerosi parlanti . . . 81
4.9 Trascrizione con diarizzazione di una traccia audio contenente numerosi parlanti 83
4.10 Trascrizione di una traccia audio caratterizzata da rumori di sottofondo . . . 84
4.11 Trascrizione a velocità di riproduzione normale . . . 85
4.12 Trascrizione di una traccia audio in lingua inglese con accento indiano . . . 86
4.13 Trascrizione di una traccia audio in lingua inglese con accento scozzese . . . 87
4.14 Trascrizione di una traccia audio in lingua inglese con accento russo mediante rilevazione automatica della lingua . . . 88
4.15 Trascrizione di una traccia audio in lingua inglese con accento russo mediante indicazione manuale della lingua . . . 88
4.16 Trascrizione di una traccia audio in lingua inglese contenente linguaggio tecnico in assenza di vocabolario personalizzato . . . 90
4.17 Trascrizione di una traccia audio in lingua inglese contenente linguaggio tecnico in presenza di vocabolario personalizzato . . . 91
4.18 Trascrizione di una traccia audio in lingua inglese caratterizzata da linguaggio finanziario in assenza di vocabolario personalizzato . . . 92
4.19 Trascrizione di una traccia audio contenente la canzone *Imagine* . . . 93
4.20 Trascrizione di una traccia audio contenente la canzone *Heal The World* . . . 94
4.21 Trascrizione di una traccia audio caratterizzata dalla combinazione di diversi fattori . . . 95
4.22 Trascrizione di una traccia audio contenente scene tratte dal film *The Dark Knight* 97
4.23 Trascrizione di una traccia audio contenente scene tratte dal film *The Dark Knight* abilitando la funzionalità di *Filtraggio del vocabolario* . . . 98

Introduzione

I sistemi di Intelligenza Artificiale (IA) stanno assumendo un ruolo sempre più importante all'interno della società odierna. Essi sono sempre più integrati nei processi aziendali e organizzativi, rendendo evidente il loro continuo inserimento nello svolgimento delle attività quotidiane. Ad oggi, uno degli impieghi maggiormente discussi è rappresentato dai chatbot. L'Intelligenza Artificiale rappresenta una disciplina in rapida evoluzione, che interessa non solo aspetti puramente tecnici relativi a discipline quali la matematica, l'ingegneria e l'informatica, ma ha anche visto il coinvolgimento di altri settori, quali le neuroscienze, la psicologia e la sociologia. Inoltre, l'Intelligenza Artificiale viene integrata con una frequenza sempre maggiore nell'ambito della ricerca, contribuendo al raggiungimento di importanti risultati in campo scientifico e tecnologico. Le tecniche di IA sono già parte integrante di numerosi flussi di lavoro, con gli hyperscaler che forniscono servizi di cloud computing caratterizzati da un'integrazione sempre maggiore di funzionalità di IA. I prodotti offerti sono numerosi e si adattano a diversi casi d'uso, quali la robotica, la cybersecurity, l'assistenza ai clienti e molti altri ancora.

Questa trattazione ha, dunque, l'obiettivo di analizzare alcuni di questi servizi, in particolare quelli offerti da Amazon Web Services (AWS). AWS svolge, ad oggi, un ruolo centrale nel mercato dei provider di servizi di cloud computing. La scelta di utilizzare i servizi forniti da AWS è dovuta alla sua ampia offerta di soluzioni basate sull'Intelligenza Artificiale. Grazie alle API messe a disposizione, le aziende e le organizzazioni possono integrare tali funzionalità all'interno dei propri sistemi. Per poter sfruttare questi servizi non è necessaria una conoscenza approfondita di discipline quali il Machine Learning e il Deep Learning. Questo aspetto rappresenta un importante vantaggio per le aziende che intendono modernizzare i propri flussi di lavoro. Oltretutto, AWS è concepito come un ecosistema all'interno del quale i diversi servizi possono essere combinati per realizzare soluzioni adattate a specifiche esigenze.

Il presente lavoro ha lo scopo di offrire un contributo allo studio del comportamento di soluzioni dedicate a funzionalità di IA. In particolare, attraverso le osservazioni ottenute mediante l'analisi, si intende evidenziare gli aspetti maggiormente positivi e le problematiche ancora presenti in questo tipo di sistemi. Infatti, per ciascuno dei casi di studio presentati vengono precedentemente esposti i limiti dei sistemi appartenenti alla stessa categoria, con l'obiettivo di confrontarli con i risultati ottenuti.

A tale scopo, viene, innanzitutto, introdotta la disciplina dell'Intelligenza Artificiale, cercando di fornirne una definizione e inquadrandola all'interno del contesto attuale attraverso una panoramica sullo stato dell'arte. In aggiunta, vengono descritti i principali rischi e benefici associati a tale tecnologia. In conclusione a questa introduzione, viene presentato il

concetto di cloud computing, in quanto strettamente correlato alle modalità di erogazione di tali servizi, e viene introdotto AWS, per comprenderne il ruolo all'interno del mercato attuale. In particolare, vengono illustrate le principali soluzioni offerte e analizzate nell'elaborato, con particolare attenzione all'adozione di tali servizi da parte di diverse aziende.

Nello specifico, l'attenzione viene rivolta a tre servizi, attraverso i quali vengono analizzati tre differenti campi di applicazione delle tecnologie di IA.

Il primo riguarda possibili applicazioni nel campo della computer vision mediante il servizio Amazon Rekognition, che permette di analizzare immagini e video al fine di recuperare caratteristiche rilevanti sfruttando tecniche di Machine Learning. Nella presente trattazione, l'attenzione è rivolta al riconoscimento del testo all'interno delle immagini.

Il secondo riguarda l'applicazione delle tecnologie di IA nel campo dei sistemi di riconoscimento vocale automatico, mediante l'utilizzo del servizio Amazon Transcribe. Per questi due strumenti, sia le immagini sia le tracce audio verranno selezionate in modo da offrire un livello di complessità progressivamente maggiore, al fine di osservare la risposta del sistema in presenza di condizioni che si avvicinano maggiormente a contesti applicativi reali.

L'ultimo settore analizzato riguarda l'applicazione di tecniche di IA per la realizzazione di bot conversazionali intelligenti. Questa tipologia di tecnologia viene, ad oggi, impiegata principalmente dalle aziende per l'assistenza ai clienti. Quest'ultima analisi ha lo scopo di far emergere, nello specifico, l'architettura alla base delle interfacce conversazionali con le quali gli utenti si interfacciano sempre più frequentemente all'interno di siti web e software, evidenziandone le principali caratteristiche.

Questa tesi è strutturata come di seguito specificato:

- Nel Capitolo 1 viene introdotto il concetto di Intelligenza Artificiale, fornendo una definizione di tale tecnologia, per proseguire poi con una classificazione delle diverse tipologie e una panoramica sul Machine Learning e sul Deep Learning; infine, viene fornito un quadro sullo stato dell'arte e sui principali rischi e benefici associati a tale tecnologia.

- Nel Capitolo 2 viene presentato l'ecosistema AWS in relazione al cloud computing, illustrando i principali servizi suddivisi per categoria funzionale e approfondendo le soluzioni dedicate all'IA, con esempi di utilizzo da parte di diverse aziende.

- Nel Capitolo 3 viene, innanzitutto, fornita una panoramica sul funzionamento del riconoscimento delle immagini, concentrandosi, successivamente, sul riconoscimento ottico dei caratteri, al fine di migliorare la comprensione dei case study relativi all'individuazione del testo nelle immagini mediante Amazon Rekognition.

- Nel Capitolo 4 viene fornita un'introduzione alla tecnologia Speech-to-Text, al suo funzionamento e ai suoi principali casi d'uso, al fine di approfondire i casi di studio relativi al riconoscimento vocale automatico mediante Amazon Transcribe.

- Nel Capitolo 5 viene presentata la tecnologia dei chatbot, illustrandone le principali tipologie, il funzionamento e i principali rischi e le sfide a essa associate, in modo da introdurre il caso di studio dedicato alla realizzazione di chatbot mediante Amazon Lex.

- Infine, nel Capitolo 6 vengono esposte le osservazioni emerse dall'analisi dei tre strumenti AWS.

CAPITOLO 1

Introduzione all'Intelligenza Artificiale

Questo capitolo si presenta come un'introduzione al concetto di Intelligenza Artificiale. Nella prima parte viene fornita una definizione di IA, partendo dall'approccio del test di Turing fino a giungere al concetto di macchine vantaggiose. Successivamente viene riportata una classificazione basata: sulle capacità, che include tre categorie: AI ristretta, AI generale (di cui fa parte anche AI generativa) e AI superintelligente; sulle funzionalità, di cui fanno parte quattro categorie: le macchine reattive, AI a memoria limitata (che include la RAG), la teoria della mente, AI autocosciente. Per completare il quadro di presentazione dell'Intelligenza Artificiale si propone una definizione di machine learning, con un approfondimento sugli algoritmi di apprendimento e le varie tipologie di machine learning. In seguito viene mostrato il concetto di deep learning, per fornire una visione sui principali approcci e una distinzione dal concetto di machine learning. A conclusione del capitolo viene riportata una panoramica sullo stato attuale dell'utilizzo dell'IA, presentando anche i suoi benefici, ma anche i potenziali rischi a cui aziende e governi devono prestare attenzione.

1.1 Definizione di Intelligenza Artificiale

Nel 2014 Stephen Hawking in una intervista rilasciata alla BCC (fonte: BBC News, 2014), ha espresso una sua precisa opionione riguardo l'Intelligenza Artificiale, nel suo senso più ampio, soffermandosi sul suo impatto sulle generazioni future: "The development of full Artificial Intelligence could spell the end of the human race.", "Lo sviluppo di un'Intelligenza Artificiale completa potrebbe segnare la fine della razza umana.". Apperentemente le sue parole possono sembrare molto forti e lasciano l'evidenza ad un approccio alquanto pessimistico nei riguardi di questa tecnologia.

Invece esperti del settore guardano all'Intelligenza Artificiale in modo meno critico, ad esempio Rollo Carpenter, creatore di *Jabberwacky* e *Cleverbot*, che si è espresso sul tema così: "I believe we will remain in charge of the technology for a decently long time and the potential of it to solve many of the world problems will be realised", "Credo che manterremo il controllo della tecnologia per un periodo di tempo ragionevolmente lungo e che il suo potenziale per risolvere molti dei problemi del mondo si realizzerà.".

Partendo da queste due visioni contrastanti, offriamo una definizione di Intelligenza Artificiale, di cui storicamente esistono molte versioni. Alcuni hanno definito intelligenza in termini di fedeltà alle capacità umane, altri hanno preferito una definizione astratta, formale di intelligenza, nel senso di razionalità. La materia stessa varia: alcuni considerano intelligenza l'abilità di saper processare pensieri interni e di ragionamento, mentre altri si focalizzano sul comportamento intelligente, una caratterizzazione esterna.

Da queste dimensioni, l'umano a confronto con il razionale e il pensiero messo di fronte al comportamento, esistono quattro possibili combinazioni e sono presenti opportuni programmi di ricerca per tutte.

I metodi utilizzati sono necessariamente differenti: la ricerca di un'intelligenza simile a quella umana dovrebbe essere parte di una scienza empirica che coinvolge anche la psicologia, includendo osservazioni ed ipotesi sul reale comportamento umano e sui processi di pensiero; si tratta di un approccio razionalista che coinvolge maggiormente la sfera matematica e ingegneristica, connessa alla statistica, alla teoria del controllo e all'economia. I gruppi si sono vicendevolmente aiutati.

1.1.1 L'approccio del test di Turing

In molti credono che l'Intelligenza Artificiale sia una tecnologia arrivata soltanto agli albori del ventunesimo secolo; in realtà questa tecnologia trova le sue radici ben prima, all'incirca negli anni '50 del Novecento.

Infatti, nel 1950, Alan Turning propone il *Turing Test*, pensato come esperimento mentale per aggirare la vaghezza filosofica della domanda: "Una macchina può pensare?".

Pertanto, un computer passa il test se un interrogatore umano, dopo aver fatto delle domande scritte, non può comprendere se le risposte siano date da una macchina oppure da una persona.

Per poter passare il test, il computer doverebbe essere in grado di:

- *elaborare il linguaggio naturale*, per comunicare in modo chiaro in linguaggio umano;
- *rappresentare la conoscenza*, per poter immagazzinare ciò che conosce o ascolta;
- *ragionare autonomamente*, per rispondere a domande e giungere a nuove conclusioni;
- *apprendere*, per adattarsi a nuove circostanze e per determinare ed estrapolare nuovi schemi.

Per esporre le sue conclusioni, Turing vide la simulazione *fisica* di una persona come non necessaria al fine di dimostrare l'intelligenza. Comunque, altri ricercatori, sulla base del lavoro di Turing, hanno proposto un *total Turnig* test, che richiede interazioni con oggetti e persone nel mondo reale. Per poterlo passare, è necessario che il robot abbia:

- *computer vision*[1] e *riconoscimento vocale*, per percepire l'ambiente circostante;
- *robotica*, per manipolare oggetti e muoversi nello spazio.

Da questi punti appena descritti si sono sviluppate le discipline che caratterizzano la maggior parte dell'AI. I ricercatori hanno devoluto poche risorse per passare il test di Turing, ritenendo più importante studiare i principi fondamentali dell'intelligenza.

1.1.2 L'approccio del modello cognitivo

Per poter capire il modo in cui un programma può ragionare come un umano, è necessario comprendere come ragionano gli umani stessi. Per tale scopo, possiamo approcciare al pensiero in questi tre modi:

[1] Con il termine *computer vision* si indica un sottocampo dell'Intelligenza Artificiale che fornisce alle macchine la capacità di analizzare, elaborare ed interpretare immagini e video. In particolare, la computer vision può essere immaginata come l'interazione fra tre processi, ovvero riconoscimento, ricostruzione e riorganizzazione, che operano insieme.

- *introspezione*, cercando di catturare i nostri pensieri così come fluiscono;
- *esperimenti psicologici*, osservando le persone mentre eseguono un compito;
- *brain imaging*, osservando il cervello nella sua attività.

Una volta apprese sufficienti conoscenze sulla teoria della mente, è possibile esprimere la teoria come un programma del computer. Se il modo di elaborare la risposta da parte del programma incontra quello corrispondente dell'umano, vuol dire che è presente un'evidenza del fatto che il meccanismo del programma possa operare come un umano. Ad esempio, Allen Newell e Herbert Simon, che hanno sviluppato il "General Problem Solver" (GPS), non erano soddisfatti del fatto che il loro programma fosse in grado di risolvere il problema correttamente, ma di come, comparando la sequenza di ragionamenti e le tempistiche dei passaggi svolti dal programma, incontrassero quelle dei soggetti umani che risolvevano gli stessi quesiti.

Da queste considerazioni prende fondamentamento il campo interdisciplinare della *scienza cognitiva*, che mette insieme la modellazione dei computer dell'Intelligenza Artificiale e le tecniche sperimentali della psicologia per costruire, teorie precise e verificabili sulla mente umana.

Agli albori delle tecniche di Intelligenza Artificiale era presente una confusione fra i due approcci, ma la loro distinzione ha permesso sia all'AI che alla scienza cognitiva di crescere più rapidamente. Chiaramente i due campi si sono influenzati fortemente a vicenda, in particolare nella computer vision, che incorpora evidenze neuropsicologiche nei modelli computazionali.

1.1.3 L'approccio delle "Leggi del pensiero"

I primi tentativi di codificare "il giusto modo di pensare", ovvero un processo cognitivo irrifiutabile, fu attuato da parte del filosofo greco Aristotele. I suoi *sillogismi* forniscono schemi per strutturare argomentazioni per raggiungere conclusioni sempre corrette partendo da premesse corrette.

Un esempio canonico che può essere proposto è il seguente: *Socrate è un uomo e tutti gli uomini sono mortali, quindi Socrate è mortale.* Queste leggi sono quelle supposte a governare le operazioni della mente, che hanno dato inizio al cossidetto campo della *logica*.

Nel 1965, i programmi, in principio, potevano risolvere qualsiasi problema descritto per mezzo di una notazione logica, con la speranza, servendosi di questi programmi, di creare sistemi intelligenti.

La logica richiede, convenzionalmente, di carpire le nozioni dal mondo in modo certo. Questo requisito viene soddisfatto dalla teoria della *probabilità*, garantendo ragionamenti rigorosi con informazioni incerte. In principio, la logica permette di costruire modelli completi di pensiero, partendo da informazioni grezze per giungere ad una comprensione di come il mondo funziona ed elaborare previsioni sul futuro. Quello che non garantisce è la generazione di un comportamento intelligente; per quello è necessaria una teoria dell'azione razionale, poichè il pensiero razionale, in sè, non risulta sufficiente.

1.1.4 L'approccio dell'agente razionale

Un *agente* è semplicemente qualcosa che agisce. Chiaramente, tutti i programmi dei computer sono in grado di fare qualcosa, ma dai computer agent ci si aspetta qualcosa di più, ovvero la capacità di operare autonomamente, percepire l'ambiente, persistere per un periodo prolungato di tempo, adattarsi ai cambiamenti, creare e raggiungere obiettivi. Un

agente razionale è un qualcosa che agisce in modo tale da raggiungere il miglior risultato o, quando questo è incerto, quello migliore atteso.

Nell'approccio delle "leggi del pensiero" all'IA, l'enfasi è sulla correttezza delle deduzioni, per cui dedurre che una certa azione sia la migliore e le conclusioni che possono derivare da questo, è parte dell'essere un agente razionale.

In opposizione, ci sono modi del pensare razionalmente che non coinvolgono lo svolgere deduzioni, ad esempio tutti i riflessi spontanei, come togliere una mano da una fonte di calore.

La rappresentazione della conoscenza e il ragionamento, così come tutte le capacità illustrate nel test di Turing, permettono ad un agente di agire in modo razionale.

L'approccio ad agente razionale risulta più vantaggioso degli altri per due motivi principali. La prima è che è più generale a dispetto delle "leggi del pensiero", che forniscono solo uno dei possibili meccanismi per raggiungere la razionalità. In seconda istanza, questo approccio presta maggiore attenzione allo sviluppo scientifico, con dimostrazioni a supporto della progettazione di agenti, non ottenibili se l'obiettivo è soltanto quello di imitare il comportamento umano o proccessi di pensiero.

Questi due motivi hanno portato l'Intelligenza Artificiale a concentrarsi sullo studio e la costruzioni di agenti in grado di fare la cosa giusta, valutata sulla base dell'oggetto fornito all'agente. Questo paradigma è stato così pervasivo da poter essere definito *modello standard*, perchè non ha solo influenzato l'IA, ma anche la teoria del controllo, dove il controllore minimizza una funzione di costo. In conclusione, è importante fare una precisazione su questo modello, ovvero l'impossibilità di raggiungere la razionalità perfetta in ambienti complessi, motivo per cui si parla di *limited rationality*.

1.1.5 Macchine vantaggiose

Il modello standard è stato utile come guida per la ricerca sull'Intelligenza Artificiale sin dalla sua introduzione, ma probabilmente non è la scelta migliore sul lungo periodo. La motivazione principale è che il modello standard assume che venga fornito alla macchina un oggetto specifico.

Infatti diventa sempre più difficile specificare un oggetto completo e corretto nel mondo reale, in particolare, quando si approccia all'interazione umano-robot. Il problema di raggiungere un punto di congiunzione fra le vere preferenze e l'obiettivo posto alla macchina è denominato *value alignment problem*. Quest'ultimo specifica che i valori o gli obiettivi forniti alla macchina devono essere allineati con quelli umani.

Nella pratica, è impossibile prevedere tutti i possibili modi nell'agire della macchina, dato un certo obiettivo, per cui si può sviluppare un comportamento ritenuto errato. Per tale motivo, il modello standard appare più inadatto. Infatti, la volontà non risiede in macchine che sono intelligenti nel senso che vogliono raggiungere i loro obiettivi, ma che devono rispondere agli obiettivi umani. Se non si è in grado di trasferire questi oggetti alla macchina perfettamente, allora è necessaria una nuova formulazione, in cui la macchina persegue gli obiettivi forniti, ma è necessariamente incerta su cosa siano. Nel momento in cui la macchina sa di non conoscere l'obiettivo completo, ha un incentivo ad agire con maggior cautela, a chiedere il consenso e imparare di più sulla base delle preferenze umane, attraverso l'osservazione e la delega del controllo all'uomo.

1.2 Classificazione

L'Intelligenza Artificiale sfrutta algoritmi e metodi complessi per costruire macchine in grado di prendere decisioni in modo autonomo e adattarsi ai cambiamenti. Ma la caratteristica

che rende maggiormente interessante l'IA è la sua notevole capacità di apprendimento autonomo e il miglioramento delle proprie prestazioni nel tempo. Questi ultimi aspetti sono ciò che la distingue fortemente dalle altre tecnologie e la rende uno strumento degno di nota.

Non è possibile riferirsi ad una sola tipologia di IA, bensì esistono categorie diverse che si differenziano per complessità, capacità e, soprattutto, ambito di applicazione.

Le due macroaree principali nelle quali è possibile suddividere l'IA sono le seguenti:

- *Classificazione basata sulle capacità*: classifica i sistemi di IA sulla base delle loro capacità generali o livelli di intelligenza. Questo spettro si estende, da strumenti specifici fino alla superintelligenza artificiale.
- *Classificazione basata sulle funzionalità*: classifica i sistemi di IA sulla base di come operano o di quali caratteristiche funzionali possiedono.

1.2.1 Tipologie di Intelligenza Artificiale basata sulle capacità

La classificazione si concentra nello specifico sul livello di avanzamento di un sistema di IA in termini di portata dell'intelligenza e autonomia operativa.

In particolare è possibile identificare tre categorie: *AI ristretta*, *AI generale*, *AI superintelligente*.

- L'*Intelligenza Artificiale ristetta* (ANI), o anche conosciuta come *Narrow AI*, o *Weak AI*, è un tipo di IA specializzata in una specifica attività o in un compito limitato. Infatti, è progettata con l'intento di eseguire compiti specifici in modo efficente, ma non è in grado di comprendere o affrontare attività al di fuori di quelle per cui è stata concepita.

  Attualmente rappresenta il paronama commerciale dell'IA. Secondo Gartner [2], quasi tutte le implementazioni dell' IA fino al 2026 rimarrano nell'ambito dell'Intelligenza Artificiale ristretta.

  L'ANI è utilizzata in una varietà di applicazioni. Fra gli usi comuni ricorrono: il riconosimento vocale e di immagini, il trading algoritmico, i motori di ricerca e molte altre applicazioni. Pertanto, questi sistemi sono progettati specificamente per l'analisi predittiva, l'elaborazione del linguaggio naturale (NLP) [3], la visione artificiale o automazione dei flussi di lavoro.

  Nella maggior parte dei casi, le ANI sono ciò che viene richiesto da parte delle aziende per soddisfare le proprie richieste.

  Le prestazioni delle ANI sono spesso dipese non dall'IA in sè, ma dalla specificità del dominio di addestramento. Il cambiamento che sta permettendo di far diffondere tale tecnologia è il passaggio da modelli lineari d'apprendimento (LLM) all'IA verticalizzata, ovvero sistemi addestrati su dati specifici del settore, con l'obiettivo di eliminare le ambiguità generiche che emergono quando viene richiesto al modello generico di ragionare su domini specializzati. Un esempio è *Bloomberg GPT*, addestrato su documenti finanziari, in grado di superare i modelli generici con un margine significativo. Altro esempio di modello verticalizzato è *Harvey*, progettato per flussi di lavoro legali,

[2] Gartner Inc. è una società per azioni multinazionale statunitense che si occupa di consulenza strategica, ricerca di mercato e analisi nel campo della tecnologia dell'informazione e della Supply Chain. Gartner si è specializzata negli ultimi anni sulle tendenze dell'AI e della sicurezza cloud-native.

[3] L'elaborazione del linguaggio naturale (NLP, Natural Language Processing) è una sottobranca della linguistica computazionale e dell'Intelligenza Artificiale che si concentra principalmente sull'interazione fra computer e linguaggio umano, nello specifico su come programmare gli elaboratori allo scopo di analizzare grandi moli di dati in linguaggio naturale.

gestione di ricerca dei casi e analisi dei contratti con un grado di precisione che i modelli generali non riescono ad eguagliare senza una costante correzione umana.

Il vantaggio operativo di questa tipologia di IA è stato dimostrato per mezzo di ricerche. Ad esempio la ricerca di McKinsey [4] ha dimostrato che l'integrazione di IA ristretta nelle operazioni di vendita, in particolar modo in un CRM [5], può generare un aumento del 15-20% della produttività delle vendite.

I suoi limiti sono altrettanto chiari: l'AI ristretta non è trasferibile; difatti tra i principali vincoli abbiamo i seguenti:

- Un modello addestrato per identificare le frodi nel settore bancario non può essere riutilizzato per il rilevamento di anomalie all'interno di un inventario del settore manifatturiero senza ricostruirlo partendo da un base di addestramento differente.
- Ogni implementazione di ANI in un campo specifico equivale all'assunzione di uno specialista del settore, eccezionale nel suo campo, ma incapace di addattarsi facilmente ad un campo diverso dal suo.
- La specificità del dominio deve essere perfettamente allineata alla richiesta dell'azienda, altrimenti le proiezioni sul ritorno sull'investimento (ROI) non saranno soddisfatte.

- L'*Intelligenza Artificiale Generale*, conosciuta anche come *Strong AI* o anche *AGI*, è la fase evolutiva dell'IA in cui le macchine avranno la capacità di pensare e prendere le decisioni come un essere umano. L'obiettivo è quello di creare delle macchine in grado di svolgere compiti multifunzionali e di agire come assistenti reali per operare in attività quotidiane esattamente come un umano.

  Attualmente non sono presenti esempi di IA forte; tuttavia le basi per tale tecnologia sono riposte nei supercomputer, nell'hardware quantistico e nei modelli di *Intelligenza Artificiale Generativa*, di cui l'esempio più eclatante di tutti è *ChatGPT*.

  In particolare, l'aspetto principale di questa tipologia di IA è la capacità di operare in diversi ambiti, senza il bisogno di dover riaddestrare il modello nel momento in cui cambia il compito.

  Il divario attualmente presente fra i modelli di apprendimento lineari (LLM) e la AGI non riguarda la scalabilità, bensì l'architettura su cui si fonda il ragionamento. I modelli più avanzati attuali si basano su quello che viene definito da parte degli scienziati cognitivi, pensiero di *Sistema 1*: veloce, intuitivo ed emotivo, basato su schemi associativi. Esso è straordinariamente efficiente nel riconoscere e completare schemi presenti nei dati di addestramento, ma manchevole di ciò che caratterizza il pensiero di *Sistema 2*: lento, ponderato e causale[6]. Infatti, ad oggi, quando viene richiesto ad un modello di risolvere un problema che necessita l'applicazione di una catena logica di pensieri senza dei precedenti chiari nei dati di addestramento, il sistema elabora risposte plausibili piuttosto che ragionare sul problema partendo da principi fondamentali.

  Una vera AGI richiederebbe un apprendimento per domini trasversali, caratterizzati da livelli qualitativamente differenti. Di seguito è riportato un esempio, che viene definito da parte dei ricercatori, lo scenario di riferimento:

[4] McKinsey & Company è una società di consulenza strategica.

[5] Con il termine automazione CRM si intende l'uso di tecnologie e strumenti di automazione per semplificare e ottimizzare le attività ripetitive all'interno di un sistema di gestione delle relazioni con i clienti (CRM, Customer Relationship Management).

[6] La presentazione del funzionamento della mente umana come interazione fittizia fra due personaggi irreali (Sistema 1 e Sistema 2) è contenuta nella prima parte del libro *Thinking, Fast and Slow* di Daniel Kahmeman, Premio Nobel per l'economia 2002.

  - Un sistema AGI che ha appreso la fluidodinamica doverebbe essere in grado di applicarla anche al contesto della liquidità finanziaria e sarebbe in grado di farlo non perchè è stato addestrato su dati finanziari, ma perchè ha compreso la logica sottostante di flusso, resistenza e pressione. Inoltre, l'apprendimento avviene senza riaddestramento, istruzioni esplicite o ottimizzazione relativa allo specifico dominio.

  Un sondaggio condotto da Deloitte[7] del 2026, attribuisce una probabilità del 50% che si possa raggiungere un capacità di risoluzione autonoma dei problemi simili all'AGI entro il 2030. Tale tempistica, anche se soltanto approssimativa, cambierebbe la prospettiva di come viene vista l'Intelligenza Artificiale da parte delle organizzazioni: da strumento a partecipante.

  L'implicazione strategica centrale non riguarda il rischio di sostituzione, bensì la predisposizione architetturale dell'azienda:

  - Le organizzazione dotate di infrastrutture dati più pulite assorbiranno più rapidamente le funzionalità dell'AGI.
  - I flussi di lavoro modulari che presentano una stretta collaborazione uomo-IA richiederanno meno ristrutturazioni.
  - Le parti che già utilizzano l'IA nei diversi cicli decisionali saranno caratterizzate da tempi di addattamento più brevi.

- La *Superintelligenza Artificiale*, anche detta *Artificial Super Intelligence* (ASI), è lo stadio dell'IA in cui le capacità della tecnologia supereranno quelle degli umani. Si tratta di un concetto teorico che si riferisce ad una forma di Intelligenza Artificiale altamente avanzata, che rappresenta il livello più elevato di Intelligenza Artificiale possibile.

  La sua caratteristica distintiva non è il superamento dell'intelligenza umana in un dominio specifico, dato che l'IA ristretta lo fa già in molti campi, ad esempio negli scacchi o nella radiologia. Un tratto caratteristico è l'automiglioramento ricorsivo, un sistema in grado di potenziare le proprie capacità più rapidamente di quanto un qualsiasi team umano possa impegnarsi per limitarlo o reindirizzarlo.

  L'ipotesi espressa da parte di Nick Bostrom [8], già messa in causa da parte di I. J. Good e altri pensatori, di cui il filosofo svedese ha dato l'analisi più rigorosa, descrive un punto fondamentale dello scenario, che coincide con il momento in cui l'IA avrà raggiunto una certa soglia di capacità; essa sarà in grado di riprogettarsi in modo autonomo per diventare più evoluta, in un processo esponenziale che supera velocemente la soglia delle prestazioni cognitive umane. Non è una preoccupazione di breve termine, ma un quadro concettuale sulla quale si basano la maggior parte dei lavori sulla sicurezza dell'IA che, ad oggi, si stanno svolgendo.

  I vincoli pratici che limitano l'ASI non sono solo teorici, ma anche reali.

  - *Limiti hardware*: l'infrastruttura di calcolo exascale necessaria per contenere la complessità a livello ASI non esiste ancora su scala implementabile.
  - *Vincoli energetici*: la quantità di energia elettrica per supportare il funzionamento exascale in modo prolungato supera la capacità attualmente disponibile dalla rete elettrica a livello globale.

[7] Deloitte & Touche, nota anche solo come Deloitte, è un'azienda multinazionale britannica nota per servizi di consulenza e revisioni.

[8] Nick Bostrom è un filosofo svedese, noto per le sue riflessioni relative al rischio esistenziale dell'umanità e sul principio antropico.

  - *Lacune di allineamento*: garantire che un sistema superintelligente persegua obiettivi compatibili con quelli del benessere umano rimane comunque un problema etico e tecnico non risolto.

Organizzazioni come *Future of Life Insitute* stanno lavorando per salvaguardie strutturali, tra cui la presenza di meccanismi di disattivazione automatica, framework di interpretabilità e protocolli di coordinamento internazionale per sistemi che, in teoria, potrebbero superare le capacità umane in termini di allocazioni delle risorse o operazioni di sicurezza.

L'ASI merita attezione a livello di pianificazione strategica, non come rischio operativo immediato, ma come fattore che plasma la pianificazione sul lungo termine. Per i leader aziandali saranno fondamentali decisioni di governance, traiettorie normative e tesi di investimento.

1.2.2 Tipologie di Intelligenza Artificiale basata sulle funzionalità

La classificazione basata sulle funzionalità permette di descrivere l'architettura operativa dei sistemi di IA. In particolare, si concentra sulla gestione dell'input, su cosa viene trattenuto fra un'interazione e un'altra e le modalità in cui elabora rappresentazioni del mondo al di là dei dati immediati. Ogni categoria serve a rappresentare un livello di complessità differente nel modo in cui il sistema si relaziona al tempo, ai diversi contesti e agli umani con cui esso interagisce.

Le categorie che è possibile identificare sono le seguenti.

- Le *Macchine reattive* rappresentano uno strato fondamentale nell'architettura dell'IA. Queste ultime lavorano senza adoperare memoria o apprendimento, o un modello che sia in grado di andare oltre l'input immediato. Infatti, ogni evento si registra come una nuova transazione, il che non rappresenta una limitazione nei sistemi in cui si tratta di una scelta progettuale coerente.

  Il vantaggio principale di servirsi di sistemi reattivi, la cui esecuzione è, per l'appunto, priva di stato, è la prevedibilità. Poichè essi non sono dotati della capacità di conservare alcuno stato storico, ne consegue l'impossibilità di perdite di memoria, errori cumulativi dovuti a interazioni passate o a distorsioni storiche. Questo aspetto si rivela particolarmente utile in contesti critici per la sicurezza, in cui il comportamento deterministico di queste macchine è proprio ciò che le rende affidabili. Tra i principali casi d'uso in cui le macchine reattive rimangono la scelta architetturale più adeguata vi sono i seguenti:

  - Sistemi in cui i tempi di risposta adeguati rappresentano un requisito fondamentale di sicurezza, ad esempio produzione di sensori per freni e sistemi di rilevamento dei guasti.
  - Controllo di conformità basato su regole ben determinate, in cui la verificabilità richiede che ogni decisione possa essere riconducibile ad un albero logico fisso e privo di stato.
  - Sistemi di controllo industriale, per cui una breve latenza, nel recupero delle informazioni per mezzo di una memoria crea un rischio operativo troppo elavato, pertanto inaccettabile.

  Due esempi di IA reattiva sono:

  - Deep Blue, il supercomputer IBM che sconfisse il grande campione di scacchi Garry Kasparov in una partita, disputata il 10 febbraio del 1996, la prima mai vinta

da parte di un computer. Esso era in grado di valutare le posizioni sulla scacchiera utilizzando delle euristiche predefinite e calcoli a forza bruta, senza servirsi di alcuna memoria di partite precedenti e apprendimento fra una mossa e l'altra.

- Il motore di raccomandazioni di Netlflix, prima di adottare l'apprendimento per rinforzo, associava l'input dell'utente a categorie predefinite, restituendo il risultato e cancellando lo stato. Esso non utilizzava un modello continuamente aggiornato con le preferenze dell'utente che si evolvesse nel tempo.

- L' *Intelligenza Artificiale a memoria limitata* è l'architettura che, allo stato attuale, si trova alla base di quasi tutti i sistemi che vengono adoperati su scala aziendale.

  Tra questi sono inclusi: i modelli linguistici di grande portata, sistemi di percezione per veicoli autonomi, motori di rilevamento di frodi fiscali e Agenti AI per flussi di lavoro complessi costituiti da più fasi. La caratteristica distintiva di questa categoria è la capacità di utilizzare i dati storici entro un intervallo predefinito di tempo, al fine di orientale decisioni attuali.

  Alla base della sua architettura è presente il modello di apprendimento profondo che sfrutta il meccanismo della auto-attenzione, pesando in modo differente la significatività di ogni parte dei dati di ingresso. Questa innovazione è stata introdotta da parte di un gruppo di ricercatori di Google Brain nel 2017 per mezzo del concetto di *Transformer*, conferendo la capacità ai modelli linguistici di mantenere un contesto coerente durante interazioni a più turni. Analogamente alle neurali ricorrenti (RNN), esse sono in grado di elaborare dati sequenziali, come linguaggio naturale, applicazioni alla traduzione e sintetizzazione di testi. Tuttavia, a differenza delle RNN, essi permettono un maggior grado di parallellismo, elaborando l'intero insieme di dati in ingresso contemporaneamente. Per tale motivo, esse rappresentano il modello preferenziale nei problemi di elaborazione del linguaggio naturale, andando a sostituire le RNN, e garantendo una riduzione dei tempi di addestramento.

  Il funzionamento nei diversi contesti di implemtazione vede un uso limitato della memoria. In particolare:

  - Nel contesto dei veicoli a guida autonoma, essa permette di mantenere una finestra mobile degli ultimi 30 secondi di dati del sensore per continuare ad aggiornare le previsioni sulla traiettoria, senza dover archiviare mesi di filmati irrilevanti.
  - Nei motori di rilevamento frodi, essa viene usata per valutare modelli di transazione in un intervallo di tempo definito, segnalando eventuali anomalie basandosi sulla deviazione dei parametri di riferimento previsti a regime stabile.
  - Per gli agenti di Intelligenza Artificiale aziendali, essa viene usata per la cronologia di conversazioni e del contesto di sessione per mantenere una coerenza nell'esecuzione delle attività su più fasi, durante lo svolgimento di un unico flusso di lavoro.

  Lo sviluppo più significativo è rappresentato dalla *Retrieval-Augmented Generation* (RAG), che permette di estendere l'IA a memoria limitata ai casi d'uso aziendali. Questa tecnica permette ai modelli linguistici di grandi dimensioni di recuperare e incorporare nuove informazioni estrapolate da fonti esterne. Per di più, i modelli non rispondono a richieste da parte dell'utente fino a che non si faccia riferimento ad uno specifico dominio, ovvero ad un insieme di documenti, che aggiungono conoscenza ai dati già forniti per mezzo dell'addestramento. In tal modo, si permette di usare informazioni specifiche relative al dominio o aggiornare informazioni che non erano presenti in sede

di addestramento. Ciò permette di riallenare il modello con nuovi dati, risparmiando a livello computazionale e finanziario.

In generale, esempi di sistemi RAG sono chatbot IA con accesso ad Internet, GitHub Copilot e altri strumenti di assistenza alla programmazione, oltre a strumenti specialistici che attingono a domini settoriali, per esempio in campo medico o legale.

- La *teoria della mente* nell'IA non ha ancora una forma effettivamente implementabile. Il concetto sulla quale si fonda trae origine dalla teoria della simulazione, in particolar modo dall'osservazione del comportamento dei bambini che usano le proprie emozioni per prevedere ciò che faranno gli altri, proiettando i propri stati mentali su di essi. In tal modo, descriverebbe la capacità di attribuire credenze, intezioni, emozioni, desideri e conoscenze a se stessi e agli altri, oltre alla possibilità di comprendere stati mentali diversi dai propri.

  L'applicazione al campo dell'Intelligenza Artificiale permette di passare da "cosa ha detto l'utente" a "perchè l'utente ha detto questo, cosa sta cercando di ottenere e come potrebbe reagire alle diverse risposte". Il divario fra il funzionamento attuale e quest'ultimo è significativo, ma si sta progressivamente riducendo, articolando l'avanzamento in tre fasi distinte:

  – il sistema identifica l'intento basandosi su modelli linguistici superficiali;
  – il sistema modella ciò che sta cercando di fare l'utente, andando oltre ciò che è esplicitamente dichiarato;
  – il sistema non solo adatta la sua risposta all'obiettivo del compito, ma anche allo stato emotivo dedotto.

  Le tendenze nello sviluppo dei nuovi sistemi si orientano sempre più verso soluzioni in cui la risposta della macchina non si limita ad adattarsi alle richieste dell'utente, ma è anche in grado di analizzarne lo stato emotivo durante l'interazione. Infatti, la ricerca si sta concentrando su sistemi in grado di rilevare microespressioni, schemi di stress vocale e segnali fisiologici per dedurre lo stato emotivo e questi stanno già influenzando le soluzioni di nuova generazione per le interfacce di assistenza clienti.

  La ricerca si è concentrata, in particolar modo, sulla modellazione degli stati mentali, esplorata per applicazioni di comunicazione organizzativa. Questa tipologia di soluzione sarebbe in grado di simulare come diversi gruppi di stekeholder reagiranno ad annunci finanziari, in materia di politica aziendale, ad esempio lanci di nuovi prodotti o possibili cambiamenti, giocando di anticipo, modellando possibili reazioni di tali gruppi.

- L'*Intelligenza Artificiale autocosciente* è puramente teorica. Ad oggi non esiste alcun sistema in grado di soddisfare una definzione seria di coscienza o autocoscienza artificiale. Attualmente, ciò di cui si dispone è un dibattito scientifico rigoroso sulla coerenza del concetto e su quali potrebbero essere le prove a sua dimostrazione.

  La *teoria dell'Informazione Integrata* (IIT), proposta originariamente dal neuroscienziato Giulio Tononi nel 2004, è il quadro scientifico più sviluppato nell'affrontare la questione. Infatti quest'ultima cerca di fornire una spiegazione al perchè alcuni sistemi fisici (il cervello umano) sono coscienti, e quali caratteristiche dovrebbero possederne degli altri affinchè anche essi siano considerabili tali.

  In linea di principio, se la teoria venisse completata e verificata in condizioni controllate, potrebbe fornire una misura su quanto un sistema fisico sia cosciente e, perfino, una descrizione di quale esperienza stia vivendo.

L'IIT sostiene che la coscienza non emerge solo dalla capacità di un sistema di elaborare informazioni, ma anche di integrarle in maniera unitaria. Difatti un sistema altamente integrato non può essere ridotto a parti indipendenti, altrimenti si perderebbe un qualcosa di essenziale riguardo al suo stato o funzione.

1.3 Machine Learning e modelli di apprendimento

I compiti che possono sembrare più difficili per l'uomo, ovvero quelli astratti e formali, in realtà sono i più semplici per un computer. Gli aspetti della vita quotidiana dell'uomo hanno bisogno di un immenso quantitativo di informazioni sul mondo. La maggior parte di queste vengono reperite in modo soggettivo e intuitivo, difficile da articolare in modo formale. Infatti, uno degli obiettivi dell'IA è proprio quello di determinare come far ottenere ad un computer conoscenza informale.

Nella realtà, diversi progetti hanno cercato di codificare in modo esplicito, per mezzo di linguaggi formali, la conoscenza del mondo arrivando ad un approccio con l'uso di regole di inferenza logica, conosciuto come *knowledge base*.

Inoltre, le difficoltà riscontrate da parte dei sistemi di IA basati su conoscenze predefinite, suggeriscono che questi debbano acquisire regole, estraendole a partire dai dati grezzi, capacità che va sotto il nome di *machine learning*.

In particolare, definiamo con questo termine un insieme di metodi in grado di individuare automaticamente schemi nei dati e di utilizzarli per prevedere dati futuri o per prendere decisioni in condizioni di incertezza.

Quindi, l'introduzione del machine learning ha permesso ai computer di tracciare problemi che riguardano il mondo reale e prendere decisioni che appaiono soggettive. Alcuni esempi di algoritmi semplici sono: il *logistic regression*, che, ad esempio, può essere utilizzato in campo medico per determinare se raccomandare un parto cesareo; un altro è il cosiddetto *naive Bayes*, che, ad esempio, può separare le email legittime da quelle di spam.

L'efficienza di questi algoritmi di machine learning è fortemente influenzata dalla rappresentazione dei dati forniti. Infatti, operare una ricerca all'interno di una raccolta di dati per una macchina può essere esponenzialmente più rapido se la raccolta è strutturata e indicizzata in modo intelligente.

Molti compiti possono essere risolti progettando l'insieme di caratteristiche adatto da estrarre e fornendole ad un semplice algoritmo di apprendimento automatico. Tuttavia è spesso difficile sapere quali caratteristiche siano rilevanti.

Una soluzione a questo problema è usare il machine learning non solo per scoprire il mapping dalla rappresentazione all'output, ma anche la rappresentazione stessa; ciò è conosciuto come *rapresentation learning*.

Naturalmente, può apparire molto difficile estrarre caratteristiche così astratte e di alto livello a partire da dati grezzi, visto che ottenere una rappresentazione appare complesso quanto risolvere il problema originale. Il *deep learning* si pone proprio questo obiettivo, introducendo delle rappresentazioni di alto livello espresse in termini di altre rappresentazioni più semplici. Anche la profondità gioca un ruolo importante; infatti, questa consente al computer di apprendere un programma articolato in più fasi. Ogni livello della rappresentazione può essere concepito come lo stato della meomoria del computer in seguito all'esecuzione di un ulteriore insieme di istruzioni in parallelo. Le istruzioni sequenziali offrono un grande potere, poichè quelle successive fanno riferimento ai risultati ottenuti da quelle precedenti.

Nella Figura 1.1 è possibile osservare un diagramma che mostra come il deep learning sia un tipo di representation learning, che, a sua volta, è una tipologia di machine learning, che viene utilizzato per molti ma non per tutti gli approcci all'Intelligenza Artificiale. Per ogni sezione del diagramma viene riportato un esempio di tecnologia IA.

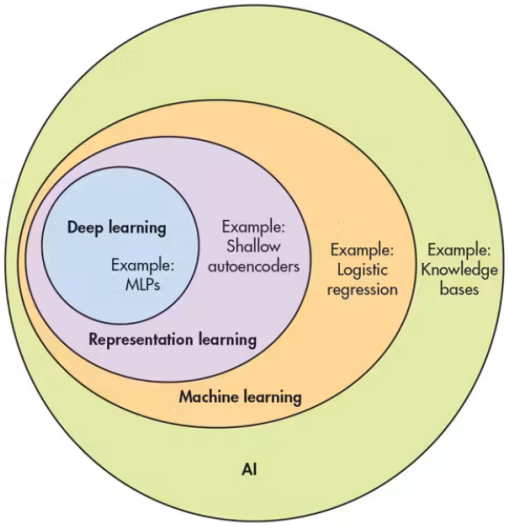


Figura 1.1: Diagramma di Venn che mostra la relazione fra le diverse discipline dell'IA (fonte: `https://www.machinedesign.com/`)

1.3.1 Algoritmi di apprendimento

Il machine learning è un processo in grado di apprendere dai dati. Con apprendimento, in questo caso, intendiamo la definizione fornita da Tom M. Mitchell: "Un programma si dice sia in grado di apprendere dall'esperienza E rispetto a certa classe di task T e misura di prestazione P, se la sua prestazione nei task appartenenti a T, misurata tramite P, migliora con l'esperienza E." In generale per definire correttamente un problema di apprendimento, è necessario identificare queste tre caratteristiche: la classe dei task, la misura di performance da migliorare e la sorgente dell'esperienza.

In questa definizione formale, il processo di apprendimento in sè non costituisce il compito, ma è il mezzo per cui si acquisisce l'abilità di eseguire il task. I task vengono geralmente descritti in base a come il sistema deve elaborare un *esempio*, ovvero un insieme di caratteristiche misurate quantitativamente a partire da un oggetto o da un evento che vogliamo far elaborare. Tipicamente esso è rappresentato per mezzo di un vettore $x \in \mathbb{R}^n$ dove ogni componente x_i rappresenta un'altra caratteristica.

Alcuni dei compiti più comuni includono:

- *Classificazione*: in questo compito ciò che è richiesto al programma informatico è di specificare a quale fra le k categorie appartenga un determinato input. Al fine di risolvere tale problema l'algoritmo produce una funzione $f : \mathbb{R}^n \to \{1, \ldots, k\}$. Assumendo $y = f(x)$ per una funzione sconosciuta f, il compito è di stimare la funzione f dato un set di addestramento etichettato e quindi di fare previsioni usando $\hat{y} = \hat{f}(x)$. Chiaramente l'obiettivo principale è di fare previsioni su nuovi input, ovvero quelli che non sono stati mai visti prima, poichè farne su set di addestramento è semplice.

- *Regressione*: questa è analoga alla classificazione con la differenza che la variabile di risposta è continua.

- *Trascrizione*: in questo tipo di task, viene chiesto al sistema di osservare una rappresentazione non strutturata di qualche tipo di dato e trascriverla in una forma discreta, testuale.

- *Identificazione di anomalie*: il programma esamina una serie di eventi o oggetti e segnala l'eventuale presenza di alcuni come atipici o insoliti.

Chiaramente molti altri tipi di task sono possibili; quelli selezionati sono solo un esempio e non constiutiscono una rigida tassonomia.

Di norma la misura quantitativa delle prestazioni è specifica per il compito svolto dal sistema. Ad esempio, per quelli di classificazione e trascrizione viene utilizzata l'*accuratezza* del modello, con cui si indica la proporzione di esempi per cui quel modello fornisce l'output corretto. In modo equivalente può essere utilizzato il *tasso di errore*, ovvero la proporzione di esempi per i quali il modello produce un output errato. Generalmente, ciò di cui si vuole essere a conoscenza, è quanto bene l'algoritmo lavora sui dati che non ha mai visto prima e, per tal motivo, spesso, la scelta della misura da adoperare è piuttosto complessa.

Tipi di machine learning

Il machine learning viene suddiviso in due categorie, *supervisionato* o *non supervisionato*, sulla base dell'esperienza che ad esso è permesso avere durante il processo di apprendimento.

- In quello *supervisionato* l'obiettivo è di apprendere un mapping da un input x ad un output y, dato un insieme etichettato di coppie input-output $D = \{(x_i, y_i)\}_{i=1}^N$, dove con D si indica il *training set* ed N è il numero di esempi di addestramento. Ogni training input x_i è un vettore D-dimensionale; le dimensioni vengono definite come caratteristiche. In modo similare l'output, in linea di principio, può essere qualsiasi cosa, ma la maggior parte dei metodi assume che y_i è una variabile categorica appartenente a un insieme finito, $y \in 1, \ldots, C$, oppure uno scalare a valori reali. Inoltre, quando y_i è categorica, il problema ricade nella classificazione o nel riconoscimento di pattern, mentre se è a valori reali ricade nella regressione. Ognuno degli esempi è associato ad un *target* o *label*, ad esempio il dataset Iris è annotato con la specie di ciascuna pianta di iris. Generalmente, l'apprendimento del predire y a partire da x avviene stimando $p(y \mid x)$. Inoltre, il termine *apprendimento supervisionato* deriva dall'idea che il target y sia fornito da un istruttore che mostra al sistema cosa deve svolgere.

- Nel secondo tipo, ovvero quello *non supervisionato*, sono forniti soltanto gli input $D = \{x_i\}_{i=1}^N$, ovvero un dataset contenente diverse caratteristiche; quindi, l'obiettivo è apprendere le proprietà della struttura, proprio a partire dal dataset. Nel contesto del deep learning, generalmente, si vuole apprendere l'intera distribuzione di probabilità che genera il dataset. Pertanto, questa tipologia di apprendimento coinvolge l'osservazione di diversi esempi del vettore casuale x e cerca di apprendere, implicitamente o esplicitamente, la distribuzione di probabilità $p(x)$, oppure alcune proprietà di quella determinata distribuzione. Inoltre, a differenza della metodologia precedente, non è presente un istruttore; quindi l'algoritmo deve imparare a dare un senso ai dati senza l'aiuto da parte di una guida.

In realtà supervised e unsupervised learning sono termini non formalmente definiti; infatti, spesso il loro confine non è ben delineato.

Altre varianti di paradigmi di apprendimento sono possibili.

- Il *reinforcement learning* coinvolge un apprendimento basato su come associare situazioni ad azioni, in modo da massimizzare un segnale di ricompensa numerico. Si tratta di problemi *closed-loop*, poichè le azioni del sistema influenzano i suoi successivi input. Per di più, non viene detto al sistema quale azioni intraprendere, bensì esso scopre in autonomia quali azioni generano la ricompensa migliore. Le azioni non influenzano solo la ricompensa immediata, ma anche la situazione successiva, e quindi tutte le ricompense seguenti. Pertanto, la natura intrinseca del sistema a ciclo chiuso, l'assenza di istruzioni dirette sulle azioni da compiere e il fatto che le conseguenze delle azioni si

manifestino su archi di tempo prolungati, costituiscono i tre tratti distintivi dei problemi di apprendimento per rinforzo.

L'introduzione dell'apprendimento per rinforzo nasce dall'esigenza di lavorare con problemi di tipo interattivo, dove è spesso impraticabile l'ottenimento di esempi del comportamento desiderato che sono sia corretti che rappresentativi di tutte le situazioni in cui l'agente attua; ciò differisce dall'apprendimento supervisionato. Tuttavia, il reinforcement learning è distante anche da quello non supervisionato, dato che, esso cerca di massimizzare un segnale di ricompensa, piuttosto che tantare di individuare una struttura nascosta.

Un elemento importante, in questa tipologia di apprendimento, è il compromesso tra esplorazione e sfruttamento. In generale, l'agente, per ottenere un buon compenso, dovrebbe preferire azioni che ha già provato nel passato e scoprire che effettivamente queste portano ad una ricompensa. Nonostante ciò, per trovare queste azioni, deve provare quelle che non ha selezionato in precedenza. Quindi l'agente deve sfruttare ciò che già conosce, ma anche esplorare per migliorare la selezione nel futuro.

Altro elemento fondamentale è considerare esplicitamente l'intero problema di un agente orientato ad un obiettivo che entra in interazione con un ambiente incerto, in contrasto ad altri approcci che considerano sottoproblemi, senza affrontare come questi si inseriscano in un quadro più ampio. Perciò, si parte da un agente completo, con degli obiettivi espliciti, in grado di percepire aspetti del proprio ambiente (anche con condizioni di significativa incertezza) e di scegliere azioni per influenzarlo.

1.4 Deep Learning

Inizialmente i primi algoritmi di apprendimento che conosciamo erano concepiti come modelli computazionali dell'apprendimento biologico. Di conseguenza, uno dei nomi con cui era noto il "deep learning" è quello di *Artificial Neural Networks* (ANN). La prospettiva su questi modelli è che essi siano ispirati al funzionamento del cervello biologico (sia esso umano o di un altro animale).

Sebbene le tipologie di reti neurali impiegate nel machine learning siano state talvolta fonte di comprensione al funzionamento del cervello, esse non sono concipite in generale come modelli realistici di funzioni biologiche. Infatti questo approccio si basa su due idee fondamentali: la prima è che il cervello offra la dimostrazione concreta della possibilità di comportamento intelligente; pertanto, un modo per realizzare l'intelligenza consiste nel ricostruire, all'inverso, i principi computazionali alla base del cervello al fine di replicarne le funzionalità. La seconda è il grande interesse scientifico nel comprendere il cervello e i principi che governano l'intelligenza umana, quindi i modelli di machine learning sono in grado di fare luce anche al di là della loro utilità prettamente ingegneristica.

I primi predecessori del moderno deep learning erano semplici modelli lineari motivati da prospettive neuroscientifiche; pertanto, questi erano disegnati per prendere un insieme di n valori in ingresso $x_1, x_2, \ldots, x_n$ e associare essi all'output y. Questi modelli apprendono un insieme di pesi $w_1, \ldots, w_n$ e calcolano il loro output $f(x, w) = x_1 w_1 + \cdots + x_n w_n$.

Ad oggi, la neuroscienza è un importante sorgente di ispirazione per il deep learning, ma non più la guida predominante nel settore; questo è dovuto all'assenza di informazioni relative al cervello umano da poter sfruttare come linea guida.

Negli anni '80, una seconda ondata di ricerche suelle reti neurali emerse in gran parte in un movimento chiamato *connectionism* o anche *parallel distributed processing*. L'idea alla base di quest'ultimo è che un gran numero di unità computazionali semplici possono raggiungere un comportamento intelligente quando sono connesse insieme. Questa intuizione si applica

allo stesso modo ai neuroni del sistema nervoso biologico e alle unità nascoste nei modelli computazionali.

Molte dei concetti sviluppati con il connectionism sono rimasti centrali negli sviluppi attuali. Tra questi quello di *rappresentazione distribuita*, ovvero ogni input di un sistema dovrebbe essere rappresentato da molte caratteristiche ed ognuna di esse dovrebbe contribuire alla rappresentazione di molti possibili input.

La terza ondata che si è venuta a sviluppare negli anni successivi fino ad oggi; essa concentra la sua attenzione sulle nuove tecniche di *apprendimento non supervisionato* e sull'abilità dei modelli di generalizzare efficacemente a partire da dataset ridotti, anche se l'interesse rimane comunque alto per algoritmi di apprendimento supervisionato più datati e per la capacità dei modelli di sfruttare ampi dataset etichettati.

Infine un ultimo obiettivo del deep learning è l'estensione al dominio del *reinforcement learning*. In questo contesto, un agente autonomo impara ad eseguire un compito con un approccio del tipo trial-and-error, senza la guida di un operatore umano.

1.5 Stato dell'arte

Una risposta coincisa a che cosa sia in grado di fare l'Intelligenza Artificiale ad oggi è abbastanza difficile, dato che esistono così tante attività in molti sottocampi differenti. Per dare un' idea dello stato di avanzamento di questa tecnologia attualmente, è stato utilizzato l'*AI Index* report 2026, una ricerca portata avanti dallo Stanford HAI, riconosciuta come risorsa affidabile per governi e aziende leader. Essa costituisce uno strumento in grado di fornire informazioni rigorose e oggettive sui progressi tecnici, sull'influenza economica e nell'impatto sociale dell'IA.

Le risorse alla base dello sviluppo dell'IA sono continuate a crescere nel 2025, ma con un numero di modelli di rilievo minore a dispetto dell'anno precedente. Inoltre i sistemi all'avanguardia sono sempre più nelle mani di gruppi ristretti di organizzazioni, oltre che essere meno trasparenti; infatti, codice di addestramento, numero di parametri e dimensioni dei dataset sono spesso tenuti riservati. Il numero di parametri è cresciuto notevolmnte dal 2010 fino al 2022, a causa della crescente complessità nell'architettura dei modelli, stabilizzandosi da allora, ma è probabile che si sottostimi la reale crescita per la mancanza di informazioni. La potenza di calcolo avanza di 3,3 volte all'anno dal 2022. Attualmente i chip di Nvidia per l'IA rappresentano oltre il 60% della capacità complessiva, con Google e Amazon che ne forniscono la maggior parte della quota restante.

Tuttavia quasi tutta quest'ultima transita attraverso un'unica fonderia di chip a Taiwan, il che rende fragile la catena di approvvigionamento globale dell'hardware.

Lo sviluppo dei modelli di rilievo continua a concentrarsi nel settore industriale, accompagnato solo in parte da quelli provenienti dal mondo accademico. Tra i principali contribuenti sono presenti: OpenAI, Google, che possiede il primato seguita da Meta e Alibaba, mentre in ambito accademico vi sono Stanford University e Tsinghua University.

Un altro elemento importante sono i dati reali utilizzati durante la fase di addestramento dei modelli, di cui si è registrato un esaurimento, spesso definito "peak data" per mancanza di testi umani e dati web di alta qualità. Infatti, le ultime ricerche si sono concentrate sulla possibilità di utilizzare dati sintetici (prodotti dall'AI) per migliorare le prestazioni dei modelli successivi, insieme ad una maggiore attenzione alla qualità dei dataset esistenti. Per di più, le ultime ricerche hanno osservato che Internet sarà ben presto invasa da contenuti prodotti con IA; infatti, a inizio gennaio 2025 più del 50% delle nuove pubblicazioni è stata generato per mezzo dell'IA.

Chiaramente l'espansione delle capacità di calcolo comporta un costo energetico diretto. La potenza complessiva dei data center dedicata all'IA ha raggiunto entro il quarto trimestre

del 2025 un quantitativo sufficiente a soddisfare il fabbisogno energetico dello stato di New York durante i picchi di domanda. Di questa solo una parte è effettivamente dedicata alla potenza assorbita dai chip, tenendo conto del restante utilizzo da parte delle infrastrutture (sistemi di raffredamento, infrastruttura di rete e altre componenti del data center).

I modelli di generazione di testo e di immagini rimangono quelli maggiormente utilizzati ad oggi.

Le pipeline RAG standard risultano inadeguate quando è necessario rispondere a delle domande che richiedono di sintetizzare informazioni provenienti da più documenti, problema che è stato affrontato da parte di Microsoft Reaserch introducendo una soluzione che permette di strutturare, per mezzo di un grafo delle conoscenze, il materiale di partenza. Altre varianti si concentrano sul recupero in più fasi o sul riordinamento dei passaggi prima della generazione.

In aggiunta, i modelli possono processare una quantità di testo maggiore in un solo input, quasi equiparabile a più libri o un intero codebase in un singolo passaggio. Anche se si è mostrato ancora un importante scostamento nella comprensione di lingue regionali, mantenendo il primato della lingua inglese.

Nell'ambito della grafica, i modelli di generazione di video non sono più solamente in grado di produrre contenuti realistici, ma iniziano a comprendere come il mondo fisico effettivamente funziona, dove i sistemi puntano a predire come una scena si evolva nel tempo. Il progresso dei modelli generativi risulta sempre più evidente, come è possibile osservare in Figura 1.2



Figura 1.2: Immagine che mostra la risposta di Midjourney al prompt: "un'immagine iperrealistica di Harry Potter" (fonte: https://hai.stanford.edu/)

Le capacità dei modelli di risolvere problemi che richiedono astrazione e generalizzazione fra formati differenti sono notevolemente migliorate, anche se ancora si riscontrano errori all'interno delle risposte. Un esempio è quello della difficoltà dei modelli di leggere gli orologi analogici, o in modo correlato, task che combinano percezione visiva con aritmetica semplice.

A livello di ragionamento per risoluzione di problemi matematici, la ricerca si è concentrata non solo sul raggiungimento di un risultato corretto, ma soprattutto sul processo di costruzione di un'argomentazione rigorosa e svolta passaggio per passaggio. Tra i casi più eclatanti, il modello GPT-5.4 Pro di ChatGPT è stato in grado, da un singolo prompt [9] fornito da un appassionato di matematica Liam Price, di restituire una dimostrazione convincente del problema numero 1196 di Erdös in 80 minuti e 17 secondi di elaborazione, rimasto irrisolto per 60 anni. L'AI non ha, chiaramente, inventato una nuova dimostrazione, ma sfruttato

[9] Con il termine prompt si definisce un testo in linguaggio naturale che richiede ad un modello di IA generativa di eseguire una specifica attività.

tecniche già esistenti in modo originale, che hanno poi richiesto una revisione da parte di matematici professionisti per la pubblicazione [10]. Questo non è l'unico esempio per cui l'AI, è stata sfruttata nella ricerca scientifica, orientandosi verso la sostituzione di interi flussi di lavoro; basti pensare al Premio Nobel per la Chimica 2024 o ancona all'adozione operativa di modelli meteorologici basati su AI presso il Centro europeo per le previsioni metereologiche a medio termine.

Inoltre, gli assistenti AI si dimostrano sempre più capaci di gestire compiti che necessitano di più passaggi e che interagiscono con il mondo reale, anche servendosi di web browsing, gestioni di file e abilità di ragionamento da più fonti. In aggiunta, molte aziende stanno concentrando i loro sforzi sulla progettazione di modelli che possono essere sfruttati nella robotica per eseguire compiti interagendo con il mondo reale, anche se le ricerche mostrano che i livelli di sicurezza necessari per completare i task non sono ancora soddisfacenti. Tuttavia, i veicoli autonomi rappresentano un'eccezione significativa, avendo raggiunto una diffusione su larga scala con risultati iniziali promettenti in termini di sicurezza.

1.6 Rischi e benefici dell'Intelligenza Artificiale

L'uso sempre più crescente dell'IA in diversi campi, fra cui economia, scienza, medicina, finanza, pone un'analisi su quelli che sono i suoi rischi e benefici.

Il potenziale dell'IA, insieme alla robotica, di liberare l'uomo dalle mansioni ripetitive, oltre all'incremento drastico della produzione di beni e servizi, potrebbe preannunciare un'era di prosperità. La capacità di accelerare la ricerca scientifica potrebbe aiutare a portare cure per malattie e soluzioni al cambiamento climatico, oltre alla gestione della mancanza di risorse.

Tra i potenziali rischi, uno dei aspetti maggiormente discussi è la sostituzione delle mansioni da parte dell'IA, causando gravi sconvolgimenti nell'occupazione.

Altro argomento fortemente discusso, negli ultimi tempi, è il consumo di energia, con conseguente impatto sull'ambiente sempre più visibile. Le tendenze in ambito di calcolo e infrastrutture necessitano di un elevato fabbisogno in termini di energia e acqua, nonchè causando ingenti emissioni di carbonio.

I chip più avanzati offrono una capacità di calcolo per watt circa dieci volte superiore rispetto di quelli disponibili un decennio fa. Fra quelli più efficenti figurano Nvidia B200 e Google TPU v5e. Tuttavia la crescita delle dimensioni dei modelli ha portanto ad un continuo aumento della potenza totale richiesta per l'addestramento dei sistemi, che arrivano a necessitare di oltre 100 milioni di watt. Una certa attenzione non è richiesta soltanto dai costi per l'addestramento; anche l'inferenza rappresenta una quota crescente nell'impronta energetica complessiva dell'IA. Infatti una volta distribuito un modello su larga scala, i costi per le richieste possono, nel giro di pochi mesi, superare quelli necessari per l'addestramento di una tantum.

I risultati di recenti ricerche mostrano che i modelli a maggior intensità energetica superano i 29 Wh per prompt lungo, oltre 65 volte di più rispetto a sistemi più efficenti. Anche una query breve (di 0,42 Wh) quando proiettata su un volume di 700 milioni di query al giorno, raggiunge un consumo annuo di elettricità paragonabile a quello di 35.000 famiglie statunitensi, un consumo di acqua dolce per evaporazione pari al fabbisogno idrico annuo di 1,2 milioni di persone ed, infine, un quantativo di emissioni di carbonio che richiederebbero una foresta delle dimensioi di Chicago per poterlo compensare.

Per metterlo in prospettiva, la domanda cumulativa di potenza richiesta da tutti i sistemi di AI è comparabile al consumo elettrico nazionale di Svizzera o Austria.

[10] La dimostrazione completa è stata pubblicata il primo maggio 2026: Primitive sets and Von Mongloldt chains: Erdös Problem 1196 and beyond.

Inoltre, l'interesse per sviluppare l'infrasfruttura per un'IA responsabile sta crescendo, ma i progressi stanno facendo difficoltà a mantenere il passo con la velocità di diffusione dell'IA. Sono stati introdotti nuovi parametri di sicurezza e un numero sempre crescente di organizzazioni sta adottando politiche per utilizzare in modo responsabile i dati, in particolare, per questioni legate alla privacy e altri aspetti legali. Infatti, IA può essere utilizzata su larga scala per eseguire sorveglianza di massa degli individui, per estrarre le attività di interesse, che, per di più, si unisce alla definizione poco chiara della titolarità dei sistemi e su chi debba essere ritenuto responsabile, se l'azienda che sviluppa il sistema o il consumatore che li acquista.

In termini di sicurezza, le tecniche di AI si sono rivelate utili nella difesa contro cyber-attacchi, ma, al contempo, sono stati anche sfruttati modelli di apprendimento per rinforzo per creare strumenti per attacchi di phishing [11].

Nella Sezione 1.5 è stata illustrata l'importanza delle informazioni per l'addestramento dei modelli di IA, a cui si aggiunge il problema dalla generazione, da parte di questa, di informazioni plausibili, ma errate (che spesso vanno sotto il nome di allucinazioni), in particolare in ambiti come legge o medicina.

[11] Con il termine phishing si indica una particolare tipologia di truffa realizzata sulla rete Internet, facendo uso principalmente di e-mail, apparentemente provenienti da istituti finanziari o da siti web che richiedono l'accesso previa registrazione. Solitamente, per rassicurare falsamente l'utente, viene indicato un collegamento che rimanda solo in apparenza al sito originale, emulandolo perfettamente.

CAPITOLO 2

Amazon Web Services

In questo capitolo viene presentata una panoramica dei servizi offerti da Amazon Web Services (AWS). Inizialmente vengono illustrati gli obiettivi di questa divisione di Amazon, introducendo il concetto di cloud computing, le motivazioni alla base della sua adozione e le diverse tipologie, suddivise in base ai modelli di deployment e di servizio. Successivamente, viene descritto il concetto di istanza nel cloud computing, con particolare attenzione alle soluzioni adottate da AWS, di cui vengono illustrate l'infrastruttura e le relative componenti. Nella seconda parte del capitolo viene proposta una panoramica dei principali servizi forniti da AWS, illustrandone le modalità di accesso. Infine, vengono descritti i servizi dedicati all'IA e alcuni dei principali casi d'uso in ambito aziendale.

2.1 Introduzione ad Amazon Web Services

Amazon Web Services (AWS) è una sussidiaria di Amazon[1], che fornisce servizi di cloud computing attraverso un'omonima piattaforma on-demand. Amazon Web Services offre un'ampia gamma di prodotti (Figura 2.1), tra cui risorse di calcolo (compute), storage, database, analytics, networking, strumenti per lo sviluppo (developer tools), IoT, sicurezza e applicazioni enterprise.



Figura 2.1: Servizi di Amazon Web Services (fonte: `https://aws.plainenglish.io/`)

[1]Amazon è un'azienda multinazionale di commercio elettronico statunitense, con sede principale a Seattle. Fondata il 5 luglio 1994 con il nome di *Cadabra.com* da Jeff Bezos, è stata lanciata nel 1995 ed è attualmente la più grande Internet company al mondo.

Gli oltre 200 servizi disponibili coprono un'ampia varietà di casi d'uso e settori applicativi, spaziando dal data warehousing[2] agli strumenti di deployment, dalle directory alla distribuzione dei contenuti.

Per contestualizzare i servizi offerti da AWS, è opportuno introdurre il concetto di cloud computing, paradigma sul quale essi si basano.

2.2 Introduzione al cloud computing

Il *cloud computing* rappresenta un modello di erogazione di risorse IT basato sull'utilizzo di infrastrutture virtualizzate accessibili tramite la rete. Esso consente di fornire capacità di calcolo, archiviazione e servizi software secondo il paradigma dell'utility computing, in cui le risorse vengono allocate dinamicamente e gli utenti pagano esclusivamente per il loro effettivo utilizzo, secondo il modello *pay-as-you-go* (in modo analogo a quanto avviene per altre tipologie di servizi, quali acqua, gas, elettricità o telefonia).

Il cloud computing nasce, anche, come risposta ai limiti delle infrastrutture di grid computing[3]. Infatti, esso introduce maggiore flessibilità, scalabilità e disponibilità delle risorse, permettendo a utenti e aziende di adattare rapidamente la capacità computazionale alle proprie esigenze senza dover effettuare importanti investimenti iniziali in infrastrutture hardware.

I principali provider cloud offrono servizi di calcolo, storage e applicazioni software attraverso piattaforme *as a service*. Questo è il paradigma alla base di AWS, che fornisce un insieme di servizi e strumenti che consentono agli sviluppatori di realizzare applicazioni scalabili e di gestire la propria infrastruttura IT tramite il cloud.

Dal punto di vista infrastrutturale, il cloud si basa su grandi data center e tecnologie di virtualizzazione che permettono di eseguire più *macchine virtuali* sulla stessa piattaforma fisica, garantendo isolamento, consolidamento delle risorse e possibilità di migrazione dei carichi di lavoro. Questa flessibilità risulta particolarmente importante nell'ambito dell'Intelligenza Artificiale, dove i moderni modelli richiedono elevate capacità computazionali difficili da gestire tramite infrastrutture dedicate.

2.2.1 Livelli e tipologie di cloud

Quando si adotta un'architettura cloud, è possibile scegliere diversi modelli e strategie di implementazione in funzione delle specifiche esigenze degli utenti. Le infrastrutture cloud realizzate dai principali provider commerciali, tra cui Amazon, Google e Microsoft, sono distribuite geograficamente e interconnesse attraverso Internet, garantendo scalabilità, efficienza dei servizi e tolleranza ai guasti.

Modelli di deployment

Nell'ambito dei modelli di deployment, le infrastrutture cloud possono essere classificate in base al livello di accessibilità e al controllo esercitato sulle risorse, come riportato di seguito.

[2] Un Data Warehouse aggrega i dati provenienti da diverse fonti, quali i sistemi transazionali, i database operativi e le piattaforme CRM, e li immagazzina in un data store ottimizzato per l'analisi quasi in tempo reale. In genere, sfrutta processi di estrazione, trasformazione e caricamento (ETL) o di estrazione, caricamento e trasformazione (ELT) per pulire, preparare e organizzare i dati destinati alla Business Intelligence (BI) e per altri casi d'uso di analisi dei dati.

[3] Il termine *grid computing* indica un'infrastruttura di calcolo distribuito utilizzata per l'elaborazione di grandi quantità di dati mediante l'uso di una vasta gamma di risorse.

- Un *cloud pubblico* è un'infrastruttura di proprietà di un provider commerciale accessibile agli utenti tramite un'interfaccia remota. I servizi vengono erogati secondo un modello tariffario flessibile basato sull'utilizzo effettivo delle risorse, consentendo alle organizzazioni di sviluppare nuove applicazioni (direttamente nel cloud) o di migrare un'infrastruttura esistente, sfruttando la scalabilità e la flessibilità del cloud computing.
- Un *cloud privato* è costruito su una rete aziendale di proprietà di una singola organizzazione. Il suo sviluppo non è finalizzato alla vendita di capacità attraverso Internet né all'accesso al pubblico tramite interfacce. Infatti, tale modello viene adottato principalmente quando sono richiesti elevati livelli di sicurezza e specifici requisiti organizzativi, rinunciando ad alcuni vantaggi tipici del cloud computing, ad esempio l'elevata scalabilità.
- Un *cloud ibrido* combina infrastrutture cloud pubbliche e private, consentendo di integrare le risorse locali con le capacità di calcolo offerte da un provider esterno. Questo approccio permette di sfruttare la flessibilità del cloud pubblico mantenendo il controllo su alcune componenti dell'infrastruttura.
- Un ambiente *multi-cloud* prevede l'uso combinato di più servizi cloud forniti da almeno due diversi provider, privati o pubblici. Tale approccio aumenta la flessibilità, permettendo di scegliere la soluzione più adatta alle esigenze specifiche dell'azienda e riducendo, al contempo, il rischio di *vendor lock-in*. Il cloud ibrido può essere considerato un caso particolare di multi-cloud quando integra servizi forniti da diversi provider pubblici.

Nel contesto dei modelli di deployment descritti, i servizi di gestione di AWS consentono di ridurre la complessità, offrendo un piano di controllo unico per la gestione di ambienti ibridi e multi-cloud. In particolare, soluzione come *AWS Systems Manager* permettono di centralizzare il controllo e l'automazione delle attività su infrastrutture distribuite. Inoltre, i servizi di AWS dedicati agli ambienti ibridi consentono di integrare risorse cloud con infrastrutture on-premises[4] o edge,[5] permettendo alle organizzazioni di sfruttare le capacità del cloud mantenendo, quando necessario, risorse locali o distrubuite vicino agli utenti per esigenze di controllo o di bassa latenza.

Modelli di servizio

Accanto ai modelli di deployment, il cloud computing prevede una classificazione dei servizi basata sul livello di astrazione delle funzionalità offerte. I diversi modelli possono essere rappresentati come un'architettura a strati, in cui i servizi di un livello superiore si basano su quelli sottostanti; inoltre, questi possono essere utilizzati in combinazione.

Di seguito sono illustrati i principali modelli di servizio del cloud computing, le cui differenze sono riportate nella Figura 2.2.

- L'*Infrastructure as a Service* (IaaS) rappresenta il livello più basso dell'architettura dei sistemi di cloud computing. Essa fornisce risorse di calcolo, rete e archiviazione dei dati virtualizzate on-demand, garantendo elevata flessibilità e un pieno controllo nella gestione dell'infrastruttura IT.

[4] In un modello*on-premises* è l'organizzazione stessa ad acquisire e installare il software nei data center. Quindi, l'implementazione IT interna dell'azienda gestisce in modo sicuro e su larga scala tutta l'infrastruttura di cui il software ha bisogno.

[5] Con il termine *edge computing* si descrive il processo di avvicinamento delle capacità di archiviazione ed elaborazione delle informazioni ai dispositivi che le producono e agli utenti che ne usufruiscono. Spostando le capacità di elaborazione più vicino agli utenti, questi sistemi migliorano notevolmente le prestazioni delle applicazioni, riducono i requisiti di banda e forniscono approfondimenti più rapidi in tempo reale.

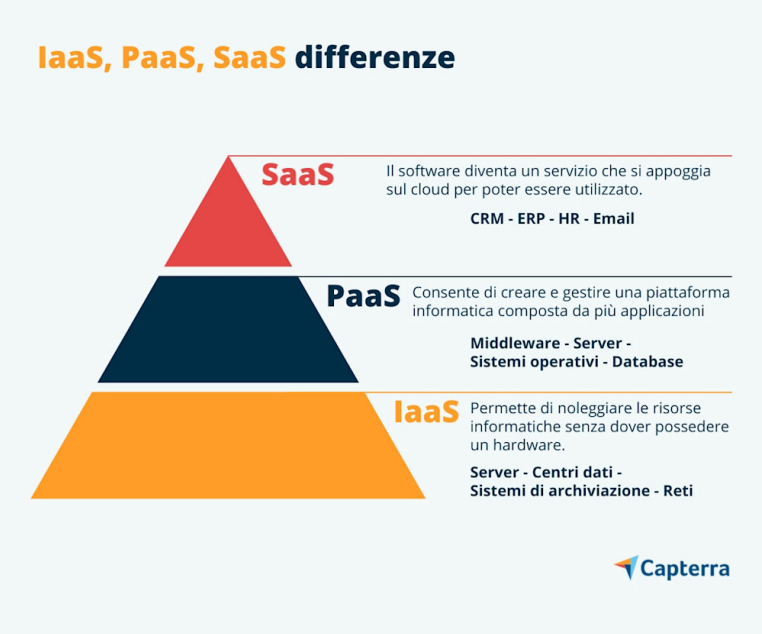


Figura 2.2: Principali differenze tra i servizi cloud (fonte: `https://www.capterra.it`)

- A un livello di astrazione superiore si colloca il *Platform as a Service* (PaaS). Esso offre un ambiente per lo sviluppo e la distribuzione delle applicazioni, sollevando gli sviluppatori dalla gestione dell'infrastruttura sottostante.
- Il *Software as a Service* (SaaS) rappresenta il livello più alto di astrazione e consente agli utenti di accedere direttamente alle applicazioni tramite portali web, riducendo gli oneri di installazione, manutenzione e aggiornamento del software.

Anche AWS organizza la propria offerta secondo questi modelli di servizio, mettendo a disposizione soluzioni IaaS, PaaS e SaaS, che rappresentano il riferimento per la classificazione dei principali servizi offerti dalla piattaforma.

2.2.2 Istanza nel cloud computing

Un elemento fondamentale nella gestione delle risorse di cloud computing è rappresentato dall'istanza cloud, che consente di eseguire carichi di lavoro senza che gli sviluppatori debbano gestire direttamente l'hardware sottostante. Le istanze cloud presentano due vantaggi principali: la scalabilità e la tolleranza ai guasti, permettendo di adattare dinamicamente le risorse disponibili in funzione delle esigenze applicative. Sebbene sia possibile gestire e mantenere le risorse fisiche on-premises, questa soluzione può risultare più costosa e meno flessibile; per questo motivo, le istanze cloud vengono, spesso, adottate per consentire l'esecuzione di un'ampia gamma di carichi di lavoro. Il concetto di istanza è strettamente legato a quello di *macchina virtuale* (Virtual Machine, VM), introdotto nella Sezione 2.2. In ambito AWS, il servizio *Amazon Elastic Compute Cloud* (Amazon EC2) consente di creare e gestire istanze virtuali utilizzando *Amazon Machine Images* (AMI). Le istanze sono organizzate in diverse famiglie, caratterizzate da differenti capacità di calcolo, memoria e archiviazione, così da adattarsi ai requisiti del software che si intende eseguire, come mostrato in Figura 2.3.

Le istanze EC2 possono essere create e amministrate tramite diverse interfacce, descritte nel dettaglio nella Sezione 2.4.1.

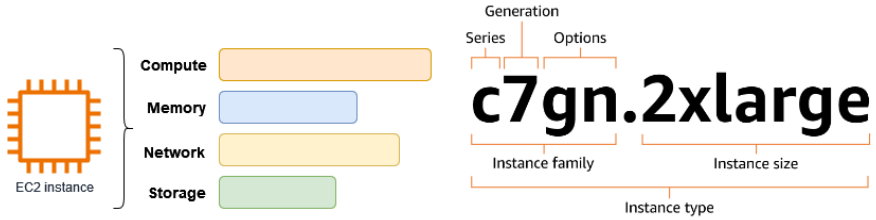


Figura 2.3: Componenti principali di un'istanza Amazon EC2 e struttura della denominazione di un tipo di istanza (fonte: `https://docs.aws.amazon.com`)

2.3 Infrastruttura globale di AWS

Dopo aver introdotto i principi alla base del cloud computing e gli elementi che ne caratterizzano il funzionamento, è possibile analizzare l'infrastruttura globale di AWS, che costituisce il fondamento per l'erogazione dei relativi servizi. L'infrastruttura fornisce una serie di soluzioni che consentono l'esecuzione di applicazioni in diverse aree geografiche. Essa è progettata per massimizzare la resilienza, le prestazioni e l'innovazione, attraverso una rete globale composta da regioni, zone di disponibilità (Availability Zone, AZ) e data center ottimizzati. I principali vantaggi dell'infrastruttura globale di AWS sono la sicurezza e l'affidabilità, la scalabilità e l'ampia disponibilità di soluzioni infrastrutturali. Nelle prossime sezioni vengono analizzati i principali elementi che caratterizzano tale infrastruttura, con particolare attenzione alle soluzioni che integrano funzionalità di IA.

2.3.1 Architettura della rete globale di AWS

La backbone di rete globale di AWS consente una trasmissione dei dati più rapida, una riduzione della latenza e un miglioramento delle prestazioni, grazie a una rete privata in fibra ottica distribuita a livello globale. Secondo i dati pubblicati da AWS al momento della consultazione, l'infrastruttura di AWS copre 123 zone distribuite in 39 regioni geografiche, con piani annunciati per l'aggiunta di altre 7 zone e altre 2 regioni, rispettivamente in Arabia Saudita e in Cile. La distribuzione geografica delle regioni AWS evidenzia una maggiore concentrazione in Nord America, Europa e Asia orientale, come illustrato nella Figura 2.4.

2.3.2 Sicurezza e affidabilità della rete AWS

La rete globale di AWS è stata progettata per garantire resilienza e alta disponibilità nel cloud, anche grazie al monitoraggio e all'instradamento centralizzati del traffico e agli strumenti automatizzati per la gestione della rete. Inoltre, essa è concepita per garantire la massima sicurezza nel cloud, offrendo diversi livelli di crittografia per il trasferimento dei dati, informazioni di routing e funzionalità esclusive di intelligence sulle minacce e di sicurezza automatizzata. Grazie all'estensione della rete, i dati rimangono più a lungo sulla rete di AWS, evitando così potenziali rischi per la sicurezza legati al passaggio attraverso più reti esterne. L'Intelligenza Artificiale e l'apprendimento automatico possono essere di supporto agli analisti nell'elaborazione di enormi quantità di dati. Un esempio dell'impiego dell'IA nell'ambito della sicurezza di rete è rappresentato da *Mithra*, un modello interno basato su reti neurali a grafo, sviluppato da AWS e impiegato per attività di threat intelligence. Attraverso modelli di apprendimento automatico, Mithra identifica e classifica i domini potenzialmente dannosi con cui i clienti entrano in contatto.

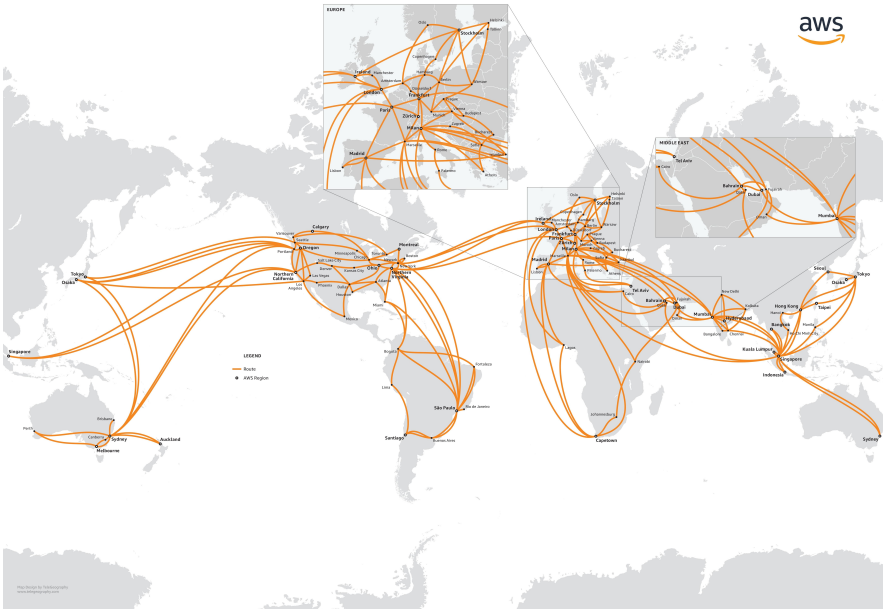


Figura 2.4: Mappa della rete globale AWS (fonte: `https://aws.amazon.com/`)

2.3.3 Prestazioni della rete globale ed efficienza energetica

La dorsale della rete globale di AWS utilizza collegamenti ad alta capacità (fino a 400 GbE), una caratteristica particolarmente rilevante nella gestione di carichi di lavoro ad alta intensità di dati. Inoltre, AWS sfrutta modelli avanzati di machine learning per analizzare grandi quantità di dati sull'uso della rete e implementare in modo proattivo capacità di rete aggiuntive con largo anticipo rispetto alla domanda dei clienti. A tale scopo, AWS dispone di una suite completa di strumenti per la gestione della rete, in grado di monitorarne lo stato, individuare potenziali problemi e risolvere in modo autonomo le anomalie senza alcun intervento umano.

Inoltre, l'infrastruttura è progettata per massimizzare l'utilizzo dell'energia nei data center, ottimizzando il posizionamento dei rack e utilizzando un software basato sull'Intelligenza Artificiale che prevede il modo più efficiente di ridurre la quantità di energia inutilizzata o sottoutilizzata[6]. Tale attenzione all'efficienza energetica riflette una tendenza sempre più diffusa nel settore tecnologico, in cui le aziende attribuiscono crescente importanza alla sostenibilità e alla riduzione dei consumi energetici delle infrastrutture IT.

2.3.4 Soluzioni per l'infrastruttura

AWS propone un'ampia gamma di soluzioni infrastrutturali, che garantiscono un'esperienza cloud coerente per le applicazioni in esecuzione nel cloud, on-premises o edge. Grazie a *Regions*, *Local Zones*, *Dedicated Local Zones*, *CloudFront*, *Outposts* e *Wavelength*, è possibile eseguire i carichi di lavoro in qualsiasi posizione desiderata. Di seguito vengono approfondite le soluzioni per l'infrastruttura maggiormente rilevanti ai fini dell'analisi condotta.

- Ogni *Region* è un'area geografica distinta, progettata per essere isolata dalle altre. La maggior parte dei servizi AWS utilizza risorse regionali, che vengono create e gestite all'interno della Region selezionata. La scelta della Region ricade sulle esigenze dell'applicativo, ad esempio sulla necessità di ridurre la latenza scegliendo un'area

[6]Con il *Climate Pledge*, AWS si impegna a raggiungere l'azzeramento delle emissioni nette di carbonio entro il 2040.

geografica vicina agli utenti. Per questo motivo, durante la creazione delle risorse tramite il proprio account AWS, è possibile selezionare la Region più adatta ai propri requisiti. Inoltre, ogni Region comprende più *Availability Zones* (AZ), collegate tramite una rete a bassa latenza e a elevata larghezza di banda. Ciascuna Availability Zone è costituita da uno o più data center ospitati in strutture separate e dotati di infrastrutture ridondanti, così da garantire continuità operativa anche in caso di guasti localizzati.

- Le *Local Zone* di AWS estendono i principali servizi della piattaforma, quali l'elaborazione, l'archiviazione, la rete, l'IA e il ML, a diverse aree metropolitane a livello globale, consentendo di ridurre la latenza per applicazioni che richiedono tempi di risposta contenuti. Inoltre, permettono di soddisfare i requisiti di residenza dei dati, qualora questi debbano rimanere in uno specifico stato o città per ragioni normative, contrattuali o di sicurezza.

- Le *Dedicated Local Zone* offrono funzionalità aggiuntive, volte a supportare le organizzazioni, in particolare quelle appartenenti al settore pubblico (come le agenzie dei governi federali, statali o locali) nell'adempimento dei requisiti di sovranità digitale e conformità per i propri carichi di lavoro sensibili. Tali zone sono particolarmente indicate per le organizzazioni che intendono migrare al cloud mantenendo l'integrazione con infrastrutture on-premises e una bassa latenza.

- *Amazon CloudFront* è un servizio web progettato per accelerare la distribuzione di contenuti web statici e dinamici agli utenti. A tal fine, esso utilizza una rete mondiale di data center, denominati *edge location*, instradando le richieste verso il nodo con la minore latenza. Inoltre, CloudFront migliora l'affidabilità del servizio grazie alla replica dei file (noti anche come *oggetti*) su più edge location dislocate su scala globale.

- *Amazon Outposts* è un insieme di soluzioni che estendono l'infrastruttura e i servizi di AWS ad ambienti on-premises o edge, al fine di garantire un'esperienza di cloud ibrido uniforme.

- *Amazon Wavelength* è un'infrastruttura AWS che consente la distribuzione di carichi di lavoro in prossimità di utenti e dispositivi connessi alla rete 5G. Nello specifico, le Wavelength Zone sono data center integrati nelle reti dei provider di telecomunicazioni e, come parte integrante di una VPC[7], consentono di accedere ai servizi della relativa Region AWS, mantenendo bassa latenza e alto throughput.

2.3.5 Posizionamento di AWS nel mercato del cloud

Oltre agli aspetti tecnologici, è opportuno considerare anche il posizionamento di AWS nel mercato dei principali hyperscaler. A tale scopo, viene riportata l'analisi proposta da Gartner nel *Magic Quadrant 2025* dedicato agli *Strategic Cloud Platform Services* (SCPS), nel quale AWS si posiziona al primo posto sia per *ability to execute* sia per *completeness of vision*, come mostrato in Figura 2.5.

Secondo la definizione fornita da Gartner, gli SCPS rappresentano offerte di cloud pubblico caratterizzate da elevati livelli di standardizzazione e automazione, che integrano servizi di infrastruttura, piattaforma e trasformazione. La valutazione di Gartner considera differenti aspetti, tra cui la completezza dell'offerta di servizi, le capacità di infrastruttura e piattaforma (IaaS e PaaS), la presenza globale dell'infrastruttura, la sicurezza, la resilienza, nonchè la

[7]Una *VPC* (Virtual Private Cloud) è una rete virtuale simile a una rete tradizionale utilizzata all'interno di un data center. *Amazon Virtual Private Cloud* (Amazon VPC) consente di avviare risorse AWS in una rete virtuale logicamente isolata con i vantaggi dell'infrastruttura scalabile di AWS.

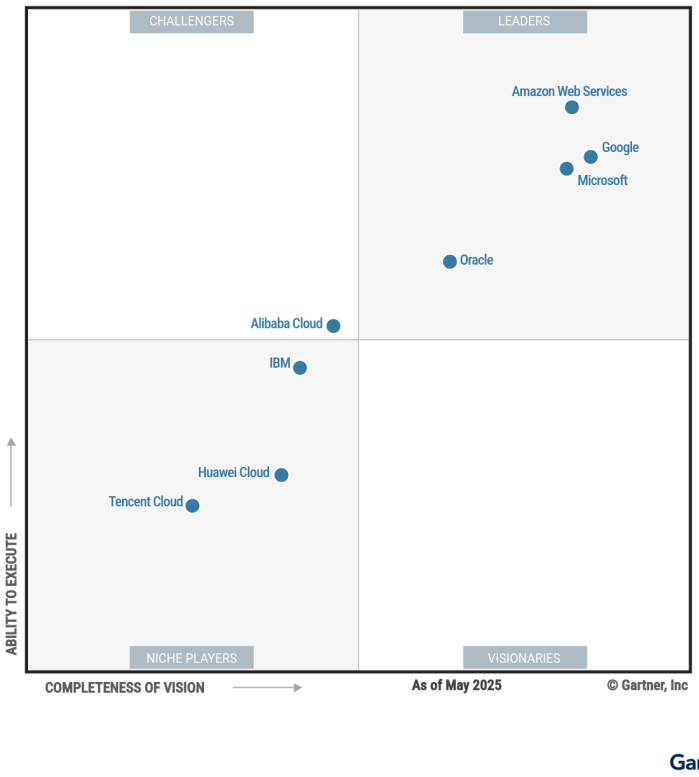


Figura 2.5: Magic Quadrant per i servizi di piattaforme cloud strategiche (fonte: `https://www.gartner.com/`)

disponibilità di strumenti dedicati all'Intelligenza Artificiale e alla gestione automatizzata delle risorse. Sulla base di tali criteri, il report evidenzia per AWS i seguenti punti di forza:

- *La più ampia comunità* dedicata al cloud pubblico, che favorisce la formazione, lo sviluppo delle competenze e l'adozione dei servizi AWS.
- *La competenza maturata nel cloud computing*, che ha consentito ad AWS di sviluppare chip personalizzati per applicazioni generiche e casi d'uso specifici dell'Intelligenza Artificiale, migliorando l'integrazione tra hardware e software e garantendo una maggiore efficienza energetica.
- *L'elevata quota di mercato* detenuta da AWS, che ha consentito all'azienda di sviluppare una solida rete di partner che supporta le organizzazioni nei processi di adozione del cloud.

Tuttavia, Gartner evidenzia per AWS i seguenti aspetti critici:

- AWS dispone di un'*offerta di applicazioni SaaS meno sviluppata* rispetto ad altri leader del settore.
- Solo *pochi componenti* della piattaforma AWS *possono operare al di fuori del cloud AWS*; infatti, la maggior parte di questi è limitata al cloud AWS o a quello distribuito con il marchio AWS.
- AWS offre una varietà di modelli di IA attraverso partnership e tramite la piattaforma Amazon Bedrock; tuttavia, *alcuni modelli proprietari*, come Amazon Nova, presentano ancora *prestazioni inferiori* rispetto ai principali benchmark del settore.

2.4 Servizi di AWS

Dopo aver descritto l'infrastruttura alla base della piattaforma AWS, vengono analizzate le categorie di servizi maggiormente rilevanti ai fini del presente lavoro, alcune delle quali sono illustrate nella Figura 2.6. Inoltre, la Tabella 2.1 riassume i principali servizi richiamati nella presente sezione, organizzati in base alla relativa categoria funzionale di appartenenza, mentre i servizi dedicati all'Intelligenza Artificiale saranno trattati separatamente nella Sezione 2.5.



Figura 2.6: Servizi di AWS suddivisi per categoria (fonte: `https://www.acs.it/`)

Categoria	Servizio	Descrizione
Compute	Amazon Lambda	Servizio di elaborazione serverless che consente di eseguire codice senza gestire direttamente l'infrastruttura sottostante.
Databases	Amazon DynamoDB	Database (NoSQL) chiave-valore e a documenti, completamente gestito, multi-regione, con funzionalità integrate di sicurezza, backup e ripristino.
Management and Governance	Amazon CloudFormation	Servizio per la definizione, la creazione e la gestione di risorse AWS tramite modelli (*template*), che consente un provisioning automatizzato e ripetibile dell'infrastruttura.
	Amazon CloudWatch	Servizio per il monitoraggio di applicazioni e risorse AWS, che consente di raccogliere e analizzare metriche operative, rilevare cambiamenti nelle prestazioni e ottimizzare l'utilizzo dell'infrastruttura.
Security, identity, and compliance	AWS Identity and Access Management (IAM)	Servizio per la gestione di identità e autorizzazioni, che consente di controllare l'accesso a risorse e servizi AWS mediante policy.
Storage	Amazon Simple Storage Service (S3)	Servizio di archiviazione a oggetti, che consente di memorizzare e gestire dati su larga scala, configurando organizzazione, accesso e protezione in base ai diversi requisiti applicativi.

Tabella 2.1: Principali servizi AWS organizzati per categoria funzionale

- *Analytics*. La piattaforma AWS comprende servizi dedicati all'analisi dei dati, progettati per supportare le diverse fasi del ciclo di gestione del dato, dall'archiviazione alla governance dei dati. Tali servizi si integrano con strumenti di Intelligenza Artificiale, come *Amazon SageMaker*, per semplificare la gestione dei dati e dei flussi di lavoro di IA. I servizi possono essere ulteriormente suddivisi sulla base della categoria di analisi, come di seguito specificato:

  - *Streaming*: servizi che consentono di creare e gestire pipeline di dati e applicazioni in tempo reale senza dover amministrare l'infrastruttura sottostante (ad esempio, *Amazon Kinesis*).
  - *Data warehouse, data lake*: servizi per l'archiviazione, l'interrogazione e l'analisi di dati presenti all'interno di data warehouse, data lake e data lakehouse (ad esempio, *Amazon Redshift*).
  - *Elaborazione dei dati*: servizi che si occupano di preparare, trasformare e integrare i dati mediante framework open source, al fine di supportare le attività di analisi e Intelligenza Artificiale (ad esempio, *AWS Glue* e *Amazon Athena*).
  - *Business Intelligence*: strumenti per la creazione di dashboard interattive, report, visualizzazioni e interrogazioni in linguaggio naturale, al fine di supportare le decisioni aziendali (ad esempio, *Amazon QuickSight*).
  - *Ricerca e analisi*: servizi che consentono di effettuare ricerche, monitorare dati e analizzare informazioni operative e aziendali in tempo reale (ad esempio, *Amazon OpenSearch Service*).
  - *Governance dei dati e dell'IA*: servizi che si concentrano sulla catalogazione, classificazione e condivisione di dati provenienti da ambienti AWS, infrastrutture on-premises e sorgenti di terze parti.

- *Cloud Financial Management*. I servizi appartenenti a questa categoria consentono di monitorare, analizzare e ottimizzare i costi associati all'utilizzo delle risorse cloud, supportando la pianificazione del budget e delle spese e una gestione più efficiente delle risorse. A tal fine, AWS offre strumenti come *AWS Billing Console, AWS Saving Plans e AWS Compute Optimizer*.

- *Compute*. I servizi di elaborazione di AWS consentono alle organizzazioni di gestire una varietà di carichi di lavoro e possono essere suddivisi nelle seguenti sottocategorie:

  - *Instances (virtual machine)*: servizi per la gestione delle capacità di calcolo nel cloud, ad esempio *Amazon Elastic Compute Cloud* (Amazon EC2).
  - *Container*: servizi per l'esecuzione di container in modo sicuro, affidabile e scalabile, su infrastrutture gestite dal cliente, ad esempio *Amazon Elastic Container Service*. Inoltre, il servizio *Amazon Elastic Kubernetes Service* (Amazon EKS) consente di eseguire Kubernetes[8] sia nel cloud AWS che nei data center on-premises.
  - *Serverless*: servizi che riducono le attività di gestione dell'infrastruttura, consentendo agli sviluppatori di concentrarsi sull'implementazione della logica di business.
  - *Edge and Hybrid*: soluzioni per la gestione di infrastrutture in ambienti ibridi ed edge, come, ad esempio, *Amazon Snow Family*.

[8] *Kubernetes* (K8s) è un software open-source che consente di automatizzare il deployment, il ridimensionamento e la gestione di applicazioni all'interno di container.

  - *Cost and capacity management*: strumenti per monitorare e gestire i costi in base alla capacità di utilizzo, ad esempio *AWS Savings Plan*.

- *Database*. Tra i servizi offerti da AWS rientrano numerose soluzioni di database, diversificate in funzione del modello di gestione dei dati e dei relativi scenari di utilizzo. Nello specifico, AWS offre due categorie di database OLTP[9]: relazionali e non relazionali. La famiglia di database relazionali comprende otto motori di database disponibili attraverso due servizi principali: *Amazon Aurora*, che include *PostgreSQL-Compatible Edition*, *MySQL-Compatible Edition* e *DSQL*, e *Amazon Relational Database Service* (Amazon RDS), che supporta i motori quali *PostgreSQL, MySQL, MariaDB, SQL Server, Oracle e Db2*. I database non relazionali utilizzano una varietà di modelli per l'accesso e la gestione dei dati, inclusi quelli chiave-valore, in-memory, document, wide-column, graph e time-series. Tali modelli sono ottimizzati per grandi volumi di dati, bassa latenza e maggiore flessibilità.

- *Management and Governance*. La categoria Management and Governance comprende strumenti dedicati al provisioning, al monitoraggio, alla gestione degli accessi, delle operazioni e della configurazione, nonché all'automazione tramite codice dell'intero ambiente cloud. Tali strumenti interagiscono e si integrano con tutte le componenti della piattaforma AWS, consentendo una gestione centralizzata ed efficiente dell'infrastruttura cloud.

- *Security, Identity and Compliance*. La sicurezza, l'identità e la governance sono elementi cruciali per preservare l'integrità e la sicurezza dei dati e dei servizi. A tale scopo, AWS offre servizi dedicati per l'archiviazione crittografata dei dati, la sicurezza della rete e il monitoraggio continuo di potenziali minacce; alcuni esempi sono *AWS Shield*, *Amazon Detective* e *Amazon GuardDuty*. In particolare, i servizi di gestione dell'identità consentono di amministrare in modo sicuro e scalabile le risorse e le relative autorizzazioni, supportando sia le applicazioni rivolte al personale interno sia quelle destinate ai clienti. Inoltre, è possibile implementare la governance del cloud su tutte le risorse AWS, avvalersi degli standard integrati e automatizzare i processi di conformità attraverso servizi dedicati, come *AWS Certificate Manager*.

- *Storage*. I servizi di storage disponibili sulla piattaforma AWS supportano diverse esigenze di archiviazione e sulla base della tipologia di storage utilizzata possono essere classificati in:

  - *Block*: l'archiviazione a blocchi è una tecnologia che gestisce l'archiviazione di qualsiasi tipo di dato e dei dispositivi di archiviazione. I dati vengono suddivisi in blocchi indipendenti e memorizzati sul supporto fisico sottostante in modo ottimizzato, garantendo un accesso e un recupero rapidi. Un esempio è *Amazon Elastic Block Store* (Amazon EBS).
  - *File system*: i file system memorizzano i dati in una struttura gerarchica composta da file e cartelle. Negli ambienti di rete viene spesso utilizzata la tecnologia *Network Attached Storage* (NAS), che consente agli utenti di accedere ai dati archiviati in rete in modo simile a quanto avviene con un disco rigido locale. Tra i principali servizi offerti rientrano *Amazon EFS, Amazon FSx ed Amazon S3 File Gateway*.

[9] Il termine *Online Transaction Processing* (OLTP) indica una tipologia di sistema di gestione di database progettata per supportare applicazioni di elaborazione delle transazioni online. Tale modello si contrappone all'Online Analytical Processing (OLAP), orientato all'archiviazione e all'analisi di grandi quantità di dati in modo rapido ed efficiente. Per questo tipo di carichi di lavoro, AWS offre *Amazon Redshift*.

  - *Object*: l'archiviazione a oggetti gestisce e memorizza i dati in un formato non strutturato denominato *oggetto* (object). Ogni oggetto è contrassegnato da un identificatore univoco e contiene metadati che ne descrivono il contenuto.
  - *Cache*: una cache è un livello di archiviazione dei dati ad alte prestazioni utilizzato per memorizzare temporaneamente i dati ai quali si accede con maggiore frequenza, con l'obiettivo di migliorare le prestazioni del sistema e ridurre la latenza. AWS offre il servizio *Amazon File Cache*.
  - *Hybrid ed Edge*: questa tipologia di storage combina l'infrastruttura di storage on-premises con i servizi di storage cloud, consentendo la mobilità dei dati tra i due ambienti in funzione dei requisiti relativi alle prestazioni, ai costi e alle conformità. Un esempio è *AWS Storage Gateway Tape Gateway*.

  A completamento dei servizi di storage, AWS offre diverse opzioni per la migrazione dei dati, adattabili alle esigenze di trasferimento. Tali soluzioni includono sia modalità di trasferimento dei dati e delle applicazioni attraverso la rete, come *AWS Transfer Family*, sia strumenti per il trasporto fisico di grandi volumi di dati mediante dischi rigidi esterni o altri supporti di archiviazione, come *AWS Snowball*.

I servizi presentati costituiscono solo una parte dell'ampia offerta AWS, che comprende soluzioni dedicate a molteplici ambiti applicativi. La loro descrizione evidenzia come AWS non si limiti a fornire servizi isolati, ma metta a disposizione un ecosistema di componenti progettati per essere integrati tra loro nella realizzazione di soluzioni complete. Questo aspetto verrà approfondito nei capitoli successivi mediante l'analisi di specifici casi d'uso.

2.4.1 Accesso ai servizi AWS

Per completare la panoramica sulla piattaforma AWS, vengono, infine, illustrate le principali modalità di accesso e gestione dei servizi.

- *AWS Management Console* è un'applicazione web che fornisce un accesso centralizzato a tutte le console dei singoli servizi AWS. Attraverso la funzionalità *Unified Navigation* è possibile ricercare i servizi, visualizzare le notifiche, consultare le informazioni relative all'account (ad esempio la fatturazione) e personalizzare le impostazioni generali della console. Inoltre, AWS fornisce *AWS Console Mobile Application*, che consente di monitorare e gestire una selezione di risorse direttamente da dispositivi mobili e di ricevere notifiche push sul loro stato, come illustrato nella Figura 2.7.
- *AWS Command Line Interface* (AWS CLI) è uno strumento open source che consente di interagire con i servizi di AWS mediante comandi eseguiti dalla command-line shell, fornendo funzionalità equivalenti a quelle disponibili tramite AWS Management Console direttamente dal terminale. AWS CLI può essere utilizzato nelle seguenti modalità:

  - *shell Linux*: consente l'esecuzione di comandi su sistemi Linux e macOS tramite shell comunemente adottate, come `bash`, `zsh` e `tcsh`;
  - *riga di comando di Windows*: permette l'esecuzione di comandi su Windows da *Prompt dei comandi* o *PowerShell*;
  - *da remoto*: consente l'esecuzione di comandi su istanze Amazon EC2 tramite client *SSH*, *PuTTY* oppure AWS Systems Manager.

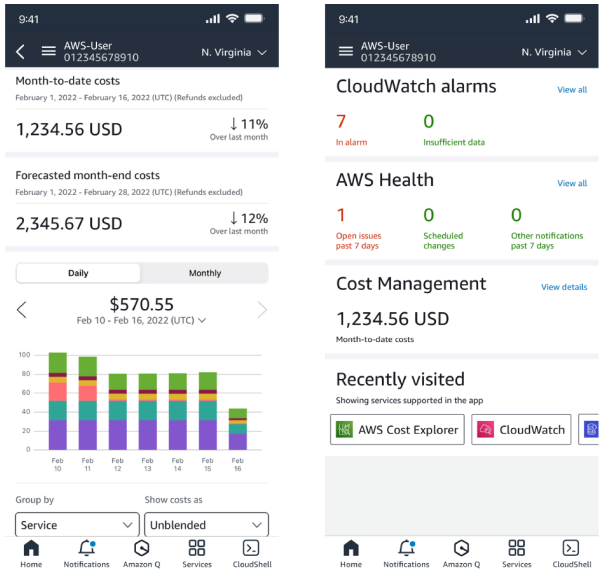


Figura 2.7: Pagina iniziale della AWS Console Mobile Application (fonte: `https://aws.amazon.com/it/console/mobile/`)

AWS CLI consente un accesso diretto alle API pubbliche dei servizi AWS. Oltre ai comandi di basso livello, equivalenti alle chiamate API, diversi servizi AWS mettono a disposizione comandi di livello superiore, che semplificano l'utilizzo delle funzionalità più complesse. Un ulteriore strumento disponibile per l'interazione con i servizi AWS é *AWS CloudShell*, una shell basata su browser accessibile direttamente dalla AWS Management Console. Le credenziali AWS utilizzate per accedere alla console vengono rese automaticamente disponibili in una nuova sessione CloudShell. L'ambiente è preconfigurato con il supporto a numerosi linguaggi di programmazione, tra cui *Node.js* e *Python*, consentendo di eseguire applicazioni senza dover installare preventivamente i relativi ambienti di esecuzione.

- Un *Software Development Kit* (SDK) è un insieme di strumenti destinati allo sviluppo di applicazioni per una determinata piattaforma. Gli SDK riuniscono in un unico ambiente tutto il necessario per lo sviluppo software, comprendendo risorse quali documentazione, tutorial, API e framework, con l'obiettivo di semplificare e velocizzare il processo di sviluppo. Generalmente, un SDK include librerie, debugger, compilatori e interpreti, profiler, code sample, strumenti di deployment e Integrated Development Environment (IDE). A tal fine, AWS fornisce SDK per numerosi linguaggi di programmazione e tecnologie, oltre a SDK specifici per l'integrazione con servizi SaaS, consentendo agli sviluppatori di utilizzare tali servizi nelle applicazioni in modo più efficiente. Tra i principali esempi figurano *AWS SDK for .NET*, *AWS SDK for Python (Boto3)*, *AWS SDK for Ruby* e *AWS WorkDocs SDK*.

2.5 Servizi AWS per l'Intelligenza Artificiale

Dopo aver analizzato le principali categorie di servizi AWS, l'attenzione viene rivolta alle soluzioni dedicate all'Intelligenza Artificiale. In tale ambito, AWS offre un'ampia gamma di servizi di IA e machine learning, oltre a un'infrastruttura appositamente progettata per supportare tali attività. Inoltre, i servizi di IA preaddestrati forniscono funzionalità pronte all'uso per applicazioni e flussi di lavoro specifici. La documentazione AWS distingue i

servizi dedicati all'Intelligenza Artificiale in diverse categorie funzionali, illustrate in Figura 2.8 e descritte di seguito.

- L'ecosistema AWS include servizi, applicazioni, strumenti e infrastrutture di supporto all'*IA generativa*, che comprendono sistemi e modelli progettati per creare nuovi contenuti, quali codice, testo, immagini, musica o altri tipi di output. Tali sistemi sono in grado di produrre contenuti sulla base delle conoscenze apprese durante la fase di addestramento sui dati disponibili. Prima di analizzare i servizi AWS dedicati all'IA generativa, è opportuno introdurre brevemente il concetto di *Foundation Model* (FM)[10], in quanto rappresenta un elemento fondamentale alla base di tali soluzioni. Questi modelli preaddestrati possono essere adattati a compiti specifici e sono disponibili in diverse tipologie:

  – *modelli di testo*, usati per elaborare e generare linguaggio naturale;
  – *modelli di immagini*, progettati per lavorare con dati visivi e supportare attività quali la generazione e l'analisi di immagini;
  – *modelli multimodali*, in grado di gestire simultaneamente differenti tipologie di dati.

  Le capacità degli FM sono generalmente correlate alle loro dimensioni; i modelli più grandi sono in grado di cogliere schemi più complessi, ma richiedono maggiori risorse computazionali. Tra i principali servizi AWS dedicati all'IA generativa rientra *Amazon Bedrock*, una soluzione progettata per fornire un accesso sicuro e di livello enterprise a Foundation Model sviluppati dalle principali aziende del settore dell'IA. Tra questi rientrano *Anthropic Claude, Cohere Command & Embed, AI21 Labs Jurassic, Meta Llama, Mistral AI, Stable Diffusion XL, Amazon Nova e Amazon Titan*. Inoltre, esso include funzionalità per la valutazione dei modelli, utili per confrontare e selezionare quelli più adatti allo specifico caso d'uso.

- Oltre ai servizi di IA generativa, AWS offre soluzioni dedicate alla *Classification AI* e alla *Predictive AI*, progettate per consentire alle aziende di creare, addestrare e implementare modelli di ML in modo più rapido. Tali servizi possono essere applicati a diverse problematiche, quali la previsione dell'abbandono dei clienti, il rilevamento delle frodi e il riconoscimento di immagini e voce.

  Tra i principali scenari di utilizzo dei servizi AWS di machine learning figurano:

  – *Casi d'uso specifici*. I servizi sono pensati per rispondere a esigenze specifiche, come la personalizzazione dei contenuti, le previsioni, il rilevamento delle anomalie e la trascrizione vocale. Inoltre, possono essere integrati nelle applicazioni senza richiedere competenze approfondite in ambito di machine learning. Alcuni esempi sono *Amazon Comprehend, Amazon Kendra, Amazon Polly, Amazon Textract, Amazon Lex, Amazon Rekognition e Amazon Transcribe*.
  – *Servizi di ML*. Questi servizi consentono di sviluppare modelli di machine learning personalizzati o flussi di lavoro che superano le funzionalità predefinite dei servizi di IA di base. Tra i principali rientrano *Amazon SageMaker AI, PyTorch on AWS, Apache MxNet e TensorFlow on AWS*.
  – *Strumenti e servizi correlati*. Questi strumenti sono progettati per accelerare le attività di deep learning nel cloud, fornendo risorse come Amazon Machine Image,

[10] I *Foundation model* sono reti neurali di grandi dimensioni basate su tecniche di deep learning. I data scientist utilizzano questi modelli come punto di partenza per sviluppare modelli di machine learning in modo più rapido ed economico.

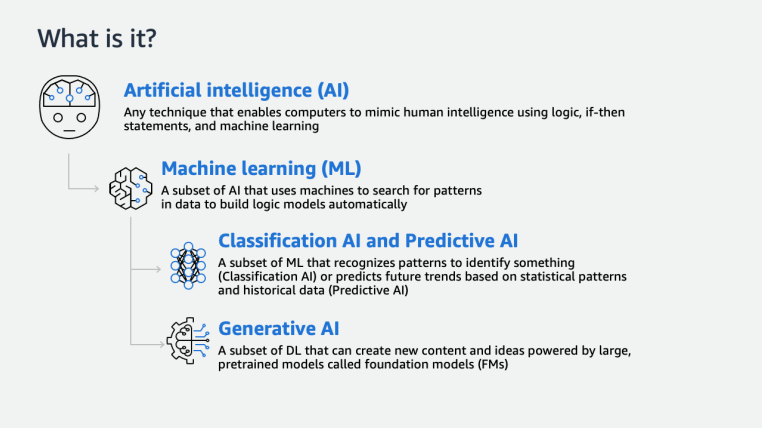


Figura 2.8: Relazione gerarchica tra Intelligenza Artificiale (AI), Machine Learning (ML), Intelligenza Artificiale Generativa, Classification AI e Predictive AI (fonte: `https://docs.aws.amazon.com`)

Docker Image e Entity Resolution. Tra i servizi disponibili figurano *AWS Deep Learning AMIs*, *AWS Deep Learning Containers* e *AWS Entity Resolution*.

Inoltre, AWS mette a disposizione hardware specializzato per le attività di addestramento e inferenza nell'ambito del machine learning ad alte prestazioni. Tra i principali servizi figurano:

- Istanze *Amazon EC2 P5*, dotate di *GPU NVIDIA H100 Tensor Core*, particolarmente indicate per applicazioni di deep learning e calcolo ad alte prestazioni.
- Istanze *Amazon EC2 G5*, che dispongono di un massimo di otto *GPU NVIDIA A10G Tensor Core* e di processori *AMD EPYC* di seconda generazione; risultano adatte a casi d'uso ad elevata intensità grafica.
- *AWS Trainium*, una famiglia di acceleratori di IA, progettati appositamente per offrire prestazioni scalabili ed efficienza dei costi nell'addestramento e nella distribuzione di modelli complessi, inclusi LLM e modelli multimodali.
- Istanze *Amazon EC2 Inf2*, basate sugli acceleratori AWS Inferentia2, progettate per fornire prestazioni elevate a un costo più basso in Amazon EC2 per le applicazioni di inferenza di deep learning e IA generativa.

2.5.1 Esempi di utilizzi di AWS per l'Intelligenza Artificiale

Concludiamo questo capitolo illustrando alcuni esempi di utilizzo dei servizi AWS per l'IA, al fine di evidenziarne le applicazioni in contesti operativi reali. I servizi AWS dedicati all'Intelligenza Artificiale sono stati integrati nei processi operativi di numerose organizzazioni. Di seguito vengono illustrati alcuni dei principali ambiti di utilizzo.

- *BMW Group* ha sviluppato una soluzione innovativa denominata *In-Console Cloud Assistant* (ICCA), sfruttando i servizi di IA generativa offerti da AWS. Questa soluzione viene analizzata con un maggiore dettaglio al fine di mostrare come i diversi servizi AWS possano essere integrati all'interno di un'architettura completa, come illustrato in Figura 2.9. La soluzione fornisce analisi e raccomandazioni in tempo reale con l'obiettivo di accelerare l'identificazione e la risoluzione dei colli di bottiglia nelle prestazioni, ottimizzare l'uso delle risorse e individuare opportunità di riduzione dei

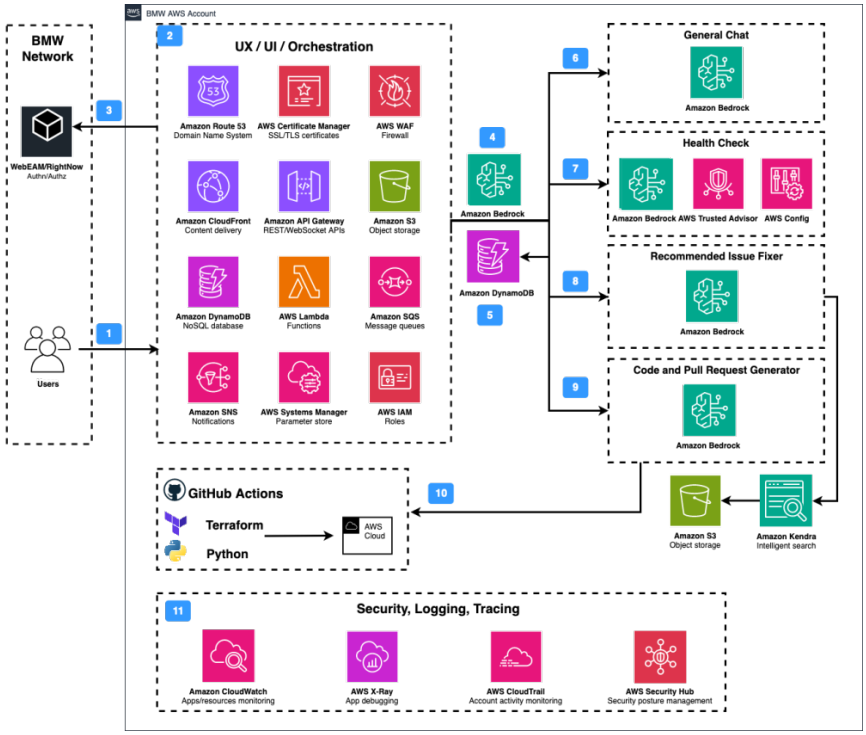


Figura 2.9: Architettura In-Console Cloud Assistant (fonte: `https://aws.amazon.com/`)

costi. A tal fine, BMW utilizza gli LLM disponibili su *Amazon Bedrock*, consentendo a ICCA di comprendere richieste formulate in linguaggio naturale e di recuperare informazioni da due fonti principali: *AWS Trusted Advisor*, che supporta l'ottimizzazione dei costi, delle prestazioni e della sicurezza dell'infrastruttura, e *AWS Config*, che analizza, verifica e valuta le configurazioni delle risorse AWS. Inoltre, ICCA può consultare la documentazione AWS e le best practice interne adottate da BMW. Il servizio offre un'interfaccia utente (UI) di tipo chatbot, che consente agli utenti di interagire con i modelli linguistici. La cronologia delle conversazioni viene archiviata tramite *Amazon DynamoDB*. Per rispondere alle domande relative ai servizi AWS e per fornire suggerimenti finalizzati all'ottimizzazione dell'infrastruttura, la soluzione sfrutta *Amazon Kendra* per indicizzare e interrogare la documentazione AWS.

- *Booking.com* utilizza l'IA generativa per migliorare i suggerimenti relativi ad alloggi e destinazioni proposti ai propri clienti. L'azienda sviluppa, addestra e distribuisce modelli di ML mediante *Amazon SageMaker*, migliorando la capacità della piattaforma di fornire raccomandazioni di viaggio basate sulle preferenze dell'utente.

- *Philips* utilizza *Amazon Bedrock* per sviluppare applicazioni di IA generativa finalizzate al potenziamento dell'elaborazione delle immagini diagnostiche digitali, alla semplificazione dei flussi di lavoro clinici e al miglioramento delle funzionalità di riconoscimento vocale. Questa soluzione permette di ridurre tempi e costi rispetto allo sviluppo e all'addestramento di Foundation Model realizzati interamente da zero.

- Il *Washington Post* integra *Amazon Polly* nel suo ambiente editoriale per offrire ai lettori un servizio di sintesi vocale, garantendo un'esperienza di ascolto coerente e di elevata qualità. Inoltre, il quotidiano statunitense utilizza *Amazon Textract* per supportare i giornalisti nel reperimento di informazioni da documenti di interesse pubblico, migliorando l'efficienza delle attività di ricerca e analisi.

Gli esempi riportati rappresentano soltanto alcune delle numerose applicazioni dei servizi AWS nell'ambito dell'Intelligenza Artificiale. Nei capitoli successivi verranno presentati specifici casi di studio finalizzati a valutarne il funzionamento, i vantaggi e le relative limitazioni.

CAPITOLO 3

Un case study relativo al riconoscimento di immagini

Questo capitolo ha l'obiettivo di illustrare alcuni casi di studio relativi al riconoscimento di immagini, con particolare riferimento alla capacità di rilevazione del testo mediante il servizio, Amazon Rekognition. In apertura viene fornita una panoramica sul funzionamento dei sistemi di riconoscimento delle immagini di Amazon Web Services, con particolare attenzione alle principali limitazioni ancora riscontrabili nei modelli e ai principali casi d'uso di tale tecnologia. Successivamente, viene approfondito il tema del riconoscimento ottico dei caratteri, poichè i casi di studio presentati nel capitolo si concentrano specificatamente su questo ambito del riconoscimento delle immagini. Viene, quindi, presentato lo strumento Amazon Rekognition, descrivendone le principali funzionalità e i possibili ambiti di applicazione. Il capitolo si conclude con l'analisi di alcuni casi di studio relativi al riconoscimento ottico dei caratteri mediante Amazon Rekognition, al fine di valutarne i principali aspetti positivi e di evidenziare eventuali problematiche.

3.1 Riconoscimento di immagini

In questa sezione viene presentato il concetto di riconoscimento delle immagini, al fine di comprendere al meglio i diversi casi di studio che verranno illustrati all'interno del capitolo. Il riconoscimento delle immagini è un ambito specifico della *computer vision* che consente a software e macchine di interpretare dati visivi. Esso può basarsi su due approcci principali:

- *Riconoscimento delle immagini con il machine learning tradizionale*. Le immagini destinate all'analisi vengono sottoposte a un processo di pre-elaborazione, che può includere la normalizzazione dei valori dei pixel, il ridimensionamento delle immagini, la conversione in scala di grigi, al fine di ridurre la complessità computazionale o per rimuovere il rumore[1] mediante tecniche di filtraggio. In tal modo, i dati risultano più coerenti e maggiormente gestibili da parte del modello di apprendimento. Successivamente, vengono selezionate le caratteristiche che rappresentano le informazioni maggiormente significative, rappresentate mediante vettori numerici che consentono al modello di confrontare e classificare i dati.

- *Riconoscimento delle immagini tramite deep learning*. A differenza dell'approccio appena descritto, i modelli di deep learning sono in grado di apprendere automaticamente le caratteristiche rilevanti direttamente dai dati visivi. In particolare, le reti neurali

[1] Con il termine *rumore* nel riconoscimento delle immagini si intende qualsiasi variazione indesiderata o causale dei pixel, ad esempio un'immagine sfocata o distorta.

convoluzionali (CNN) elaborano progressivamente i valori dei pixel dell'immagine attraverso una struttura a più livelli, individuando caratteristiche sempre più complesse, dalle informazioni locali, come bordi e texture, fino a forme ed oggetti, come illustrato nella Figura 3.1.

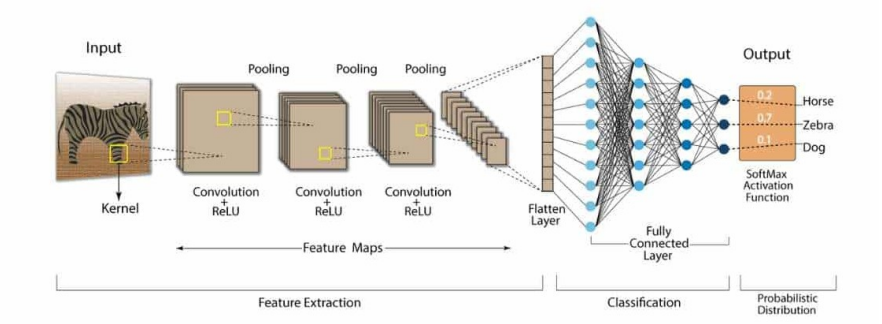


Figura 3.1: Diagramma di flusso di una CNN (fonte: `https://www.unidformazione.com/`)

Dopo aver estratto le caratteristiche, la rete può trasformare i dati in vettori unidimensionali e trasferirli a livelli completamente connessi, che integrano i pattern appresi in precedenza al fine di perfezionare il processo di classificazione e produrre una previsione finale. Tale previsione viene confrontata con il set di dati di addestramento per calcolare eventuali errori e aggiornare i parametri del modello, al fine di migliorarne le prestazioni.

Per l'addestramento dei modelli possono essere utilizzate diverse tipologie di apprendimento, già descritte nel dettaglio nella Sezione 1.3.1 e di seguito richiamate in relazione al riconoscimento delle immagini:

- *Apprendimento supervisionato*: ad esempio, per addestrare un modello a riconoscere immagini di gatti, è possibile contrassegnare migliaia di immagini con etichette quali "gatto" o "non gatto"; in tal modo, il modello apprende le principali caratteristiche dell'animale, come alcune particolarità del manto, la presenza di baffi e la forma delle orecchie.
- *Apprendimento non supervisionato*: il modello apprende a partire da dati non etichettati e identifica relazioni senza ricorrere a categorie predefinite, raggruppando le immagini sulla base di caratteristiche condivise.
- *Apprendimento autosupervisionato*: il modello apprende rappresentazioni significative partendo da dati non etichettati, generando pseudo etichette dalla struttura intrinseca dei dati.

3.1.1 Difficoltà nel riconoscimento di immagini

Le tecniche impiegate nel riconoscimento delle immagini hanno raggiunto un elevato livello di sviluppo, ma presentano ancora alcune limitazioni che possono compromettere l'affidabilità e la precisione dei risultati. Di seguito vengono riportati i principali fattori critici che incidono sulle prestazioni dei modelli, alcuni dei quali sono stati utilizzati come criteri per la selezione delle immagini analizzate nei casi di studio presentati nella Sezione 3.4.

- *Immagini degradate o oscurate*: la presenza di rumore, sfocatura, distorsioni od occlusioni può limitare la corretta identificazione degli oggetti, poichè alcune delle loro

caratteristiche risultano poco visibili, generando delle ambiguità durante il processo di classificazione.

- *Variazioni di angolo e prospettiva*: i cambiamenti nel punto di osservazione possono rappresentare un ostacolo all'identificazione degli oggetti. Ad esempio un oggetto osservato da una prospettiva insolita o ruotato rispetto agli esempi utilizzati durante l'addestramento potrebbe non essere rilevato correttamente dal modello.

- *Condizioni di illuminazione*: variazioni di luminosità, ombre o ambianti scarsamente illuminati possono influire sulle prestazioni del sistema. Ad esempio, aree eccessivamente illuminate possono causare la perdita di dettagli, mentre le ombre possono occultare caratteristiche rilevanti, inducendo il modello a interpretare erroneamente il colore, la forma o la texture di un oggetto. Tecniche di pre-elaborazione, come l'equalizzazione dell'istogramma[2], possono contribuire al miglioramento delle prestazioni in condizioni di scarsa luminosità.

- *Dimensioni e prossimità dell'oggetto*: le dimensioni dell'oggetto e la sua posizione all'interno dell'immagine possono influire sulla capacità del modello di identificarlo con precisione. Gli oggetti di piccole dimensioni possono non avere dettagli sufficienti, mentre quelli molto vicini alla fotocamera possono apparire deformati o parzialmente inquadrati dalla prospettiva. Per ovviare a tali problematiche, i modelli vengono generalmente addestrati su dataset contenenti oggetti di varie dimensioni e posizionati a diverse distanze.

- *Limitazioni nei dati di addestramento*: le capacità del modello sono strettamente legate alla qualità, alla diversificazione e alla quantità dei dati utilizzati in fase di addestramento. Ad esempio, un modello addestrato principalmente su immagini ad alta risoluzione potrebbe incontrare difficoltà nell'analizzare immagini di qualità inferiore.

- *Costo dell'etichettatura dei dati*: nell'apprendimento supervisionato, la creazione di dataset annotati richiede un significativo impegno di tempo e risorse economiche, ma esso costituisce un fattore essenziale per l'accuratezza e la capacità di generalizzazione del modello.

3.1.2 Casi d'uso di riconoscimento di immagini

L'impiego del riconoscimento delle immagini è in continua espansione e sta favorendo lo sviluppo di applicazioni sempre più sofisticate in numerosi settori. Di seguito sono riportati alcuni dei principali ambiti di applicazione di questa tecnologia.

- *Industria automobilistica*. Il riconoscimento delle immagini riveste un ruolo fondamentale nello sviluppo di veicoli a guida autonoma e semi-autonoma per rilevare con precisione la presenza di veicoli, semafori, segnaletica stradale, pedoni e altri elementi presenti nell'ambiente circostante, attraverso l'integrazione dei dati acquisiti da telecamere, sensori e sistemi LiDAR.

- *Riconoscimento facciale*. Il riconoscimento facciale trova ampio utilizzo nei sistemi di sicurezza e sorveglianza e nei dispositivi mobili per l'identificazione o la verifica dell'identità degli individui a partire dalle caratteristiche del volto.

[2] L'*equalizzazione dell'istogramma* è una tecnica di elaborazione digitale delle immagini generalmente impiegata per incrementare il contrasto globale dell'immagine.

- *Gestione dei social media*. Le piattaforme di social media impiegano tecniche di riconoscimento e analisi delle immagini per il rilevamento dei volti, l'applicazione di filtri AR[3] e la moderazione automatica di contenuti.
- *Dispositivi smart e informazioni in tempo reale*. I dispositivi indossabili, come gli occhiali smart, e alcune applicazioni per dispositivi mobili sfruttano tecniche di realtà aumentata e riconoscimento delle immagini per sovrapporre informazioni digitali agli elementi presenti nell'ambiente circostante.
- *Elettrodomestici*. Il riconoscimento delle immagini viene impiegato negli elettrodomestici intelligenti, ad esempio per il monitoraggio del contenuto dei frigoriferi, il rilevamento degli ostacoli negli aspirapolveri robotizzati, l'individuazione di persone nelle telecamere di sicurezza e l'analisi facciale negli specchi intelligenti o nei monitor per bambini. In ambito industriale, i sistemi di visione artificiale trovano impiego nell'esecuzione del controllo della qualità e nella guida dei robot durante le operazioni di scansione e recupero degli articoli, di assemblaggio delle parti e di smistamento dei materiali.
- *Diagnostica per immagini*. L'analisi delle immagini viene utilizzata in campo medico per assistere gli operatori sanitari nell'interpretazione di radiografie, risonanze magnetiche, tomografie computerizzate (TC) e altre immagini diagnostiche, contribuendo al rilevamento di malattie in fase precoce.
- *Elaborazione dei documenti*. Le organizzazioni utilizzano tecniche di riconoscimento delle immagini, spesso integrate con sistemi di *riconoscimento ottico dei caratteri*, per automatizzare il recupero e la verifica dei dati contenuti in documenti amministrativi, finanziari e sanitari.

3.2 Riconoscimento ottico dei caratteri

Dalla precedente classificazione (Sezione 3.1.2) è stato escluso esplicitamente l'ambito applicativo del *riconoscimento ottico dei caratteri* (Optical Character Recognition, OCR), che viene approfondito nel presente paragrafo, poichè i casi di studio analizzati in questo capitolo si concentrano proprio su questa specifica applicazione del riconoscimento delle immagini. L'OCR è una tecnologia che consente di convertire il testo presente all'interno di immagini, documenti digitalizzati, o acquisiti tramite scansione, in un formato leggibile ed elaborabile da un sistema informatico. I moderni sistemi di OCR possono integrare tecniche di IA e machine learning per migliorare l'accuratezza del riconoscimento, anche in presenza di caratteri manoscritti, differenti font tipografici o immagini di qualità non ottimale. Pertanto, l'OCR rappresenta una tecnologia fondamentale per la digitalizzazione dei documenti e per l'automazione dei flussi di lavoro informativi, trovando applicazione in numerosi contesti operativi che verranno illustrati nelle sezioni successive.

3.2.1 Funzionamento dell'OCR

Il funzionamento di un sistema OCR si articola in una serie di fasi che consentono di convertire un documento fisico o un'immagine contenente testo in un formato testuale digitale modificabile. Le soluzioni OCR possono essere distribuite come programmi indipendenti, API o servizi disponibili mediante piattaforme web. Le fasi tipiche di un sistema OCR sono le seguenti:

[3] I filtri di realtà aumentata (Augmented Reality, AR) rilevano i punti di riferimento del volto per posizionare correttamente elementi virtuali, quali occhiali, cappelli o altri accessori, in modo da seguire i movimenti del viso in tempo reale.

- *Acquisizione dell'immagine*. Uno scanner, una fotocamera o altro dispositivo di acquisizione acquisisce il documento e lo converte in un'immagine digitale. Successivamente, il software OCR analizza l'immagine distinguendo le aree chiare come sfondo (background) e quelle scure come testo.
- *Pre-elaborazione*. L'immagine digitale viene sottoposta a una serie di operazioni finalizzate a migliorarne la qualità e a facilitarne l'interpretazione da parte del sistema OCR. Tra le principali attività di pre-elaborazione figurano:
  - il raddrizzamento del documento scansionato, al fine di correggere eventuali problemi di allineamento dovuti alla scansione;
  - la correzione dell'illuminazione e la rimozione di bordi presenti nell'immagine;
  - l'eliminazione di righe, riquadri o altri elementi grafici che potrebbero interferire con il riconoscimento del testo;
  - il riconoscimento della lingua o del sistema di scrittura, particolarmente importante nei sistemi OCR multilingua.
- *Riconoscimento testuale*. In questa fase il sistema identifica i caratteri presenti nell'immagine. Storicamente, i due approcci principali impiegati per il riconoscimento sono i seguenti:
  - *Pattern recognition* (o pattern matching): il carattere viene riconosciuto confrontandolo con glifi precedentemente memorizzati dal sistema.
  - *Feature recognition* (rilevamento o estrazione delle caratteristiche): il carattere viene identificato attraverso l'analisi delle sue caratteristiche principali, quali linee, intersezioni, curve e cerchi. Le caratteristiche estratte vengono confrontate con quelle presenti nel modello di riferimento per individuare la corrispondenza maggiormente plausibile. Nei moderni sistemi OCR, tali approcci sono spesso integrati o sostituiti da modelli basati su reti neurali e tecniche deep learning.
- *Layout recognition*. In molti casi, il sistema analizza anche la struttura del documento per identificare e separare vari elementi che lo compongono, quali blocchi di testo, tabelle, immagini e intestazioni. Successivamente, i blocchi testuali vengono suddivisi in righe, parole e caratteri, facilitando il processo di riconoscimento.
- *Post-elaborazione*. Dopo il riconoscimento, il testo ottenuto può essere sottoposto a ulteriori controlli e correzioni, sfruttando dizionari linguistici e regole grammaticali. Infine, le informazioni raccolte vengono memorizzate in formati digitali, ad esempio, PDF modificabili. Alcuni sistemi conservano sia il documento originale sia la versione post-prodotta, garantendo una gestione più completa e tracciabile delle informazioni.

3.2.2 Tipologie di OCR

I sistemi di OCR possono essere classificati in diverse categorie sulla base del loro funzionamento. Le principali tipologie sono illustrate di seguito.

- *OCR tradizionale*. In questa tipologia di sistema, il riconoscimento avviene confrontando ciascun carattere individuato nell'immagine con un insieme di glifi memorizzati in un database di riferimento. Questo approccio presenta alcune limitazioni, poichè la sua efficacia dipende dalla capacità del sistema di rappresentare in modo esaustivo le possibili varianti di font, stile tipografico e calligrafico, che possono risultare numerose e non completamente coperte dal set di riferimento disponibile.

- *Optical Mark Recognition* (OMR). L'OMR è una tecnologia utilizzata per rilevare la presenza di marcature predefinite all'interno di documenti strutturati, quali caselle selezionate, riquadri nei sondaggi o firme su un modulo, oltre a loghi, filigrane e simboli.
- *Riconoscimento intelligente dei caratteri* (Intelligent Character Recognition, ICR). L'ICR rappresenta un'evoluzione dell'OCR tradizionale e sfrutta tecniche di machine learning e, più recentemente, di deep learning per migliorare l'accuratezza del riconoscimento. A differenza dei sistemi OCR tradizionali, l'ICR è in grado di apprendere le caratteristiche dei caratteri e di adattarsi a differenti stili di scrittura, inclusa quella manoscritta. Durante il processo di riconoscimento, il sistema analizza le caratteristiche, quali linee, curve, intersezioni e altre forme geometriche, combinando le informazioni ottenute per identificare il carattere più plausibile. Nella Tabella 3.1 vengono evidenziate le principali differenze tra l'ICR e l'OCR tradizionale, al fine di comprenderne maggiormente il funzionamento e il relativo impiego nell'analisi svolta.
- *Riconoscimento intelligente delle parole* (Intelligent Word Recognition, IWR). L'IWR può essere visto come una naturale evoluzione dell'ICR. In questo caso, il sistema non si limita al riconoscimento dei singoli caratteri, ma interpreta intere parole come unità semantiche.

Aspetto	**OCR tradizionale**	**ICR**
IA e reti neurali	Basato su modelli, esso non sfrutta queste tecnologie.	Sfrutta tali tecnologie per il recupero dei dati.
Tipologia di input	Richiede un formato specifico per l'input.	Riconosce vari formati di dati e si adatta a diverse strutture dei documenti.
Adattabilità	Ideale per applicazioni con documenti strutturati	Addestrato per adattarsi facilmente a modifiche frequenti nei documenti, ad esempio documenti commerciali.
Compatibilità con il formato di input	Limitato al testo digitale, risulta poco versatile.	Compatibile con un'ampia gamma di tipologie di input, quali immagini, moduli scritti a mano e documenti complessi.

Tabella 3.1: Principali differenze fra OCR tradizionale e ICR

3.2.3 Casi d'uso dell'OCR

Gli ambiti di applicazione dell'OCR sono numerosi e comprendono settori quali l'istruzione, la finanza, la logistica, la Pubblica Amministrazione e la sanità. Negli ultimi anni, l'evoluzione delle tecnologie di IA e di machine learning ha ampliato ulteriormente le capacità dei sistemi OCR, consentendo non solo il riconoscimento del testo, ma anche l'estrazione e la strutturazione automatica delle informazioni contenute nei documenti. Alcune soluzioni più recenti integrano modelli di IA generativa per semplificare il processo di organizzazione e interpretazione dei dati estratti. Di seguito sono riportati i principali ambiti di applicazione di tale tecnologia.

- *Digitalizzazione e gestione dei documenti*. L'applicazione più diffusa dell'OCR consiste nella conversione di documenti cartacei o immagini contenenti testo in formato digitale. I contenuti riconosciuti possono essere elaborati mediante software, quali *Microsoft Word* o *Google Docs*, oppure archiviati in sistemi di gestione documentale. Questa tecnologia trova impiego nella digitalizzazione di documenti amministrativi, cartelle cliniche, pratiche assicurative, fatture e archivi storici, contribuendo a ridurre i costi legati alla

gestione di archivi cartacei e a migliorare l'efficienza dei processi organizzativi. Anche gli appunti manoscritti possono essere convertiti in testo digitale, facilitandone la consultazione e la modifica.

- *Automazione dell'acquisizione dei dati*. L'OCR consente di automatizzare attività che, tradizionalmente, richiederebbero l'inserimento manuale dei dati. In tal modo, i dati estratti da moduli, fatture o altri documenti possono essere trasferiti direttamente nei sistemi informativi aziendali, riducendo tempi di elaborazione, costi operativi ed errori dovuti all'inserimento manuale.

- *Indicizzazione e ricerca delle informazioni*. I sistemi di OCR vengono frequentemente impiegati per rendere ricercabili grandi volumi di documenti digitalizzati. Una volta convertito il contenuto in formato digitale, i documenti possono essere indicizzati e interrogati tramite motori di ricerca, consentendo agli utenti di individuare rapidamente le informazioni desiderate senza la necessità di classificare manualmente i documenti.

- *Supporto ai processi di analisi dei dati*. L'OCR può costituire una fase preliminare fondamentale nei processi di data mining[4], consentendo la trasformazione di documenti cartacei, immagini e archivi digitalizzati in dati testuali strutturati e analizzabili. In tal modo, informazioni precedentemente non accessibili ai sistemi informatici possono essere integrate nei processi di analisi, facilitando l'individuazione di pattern, correlazioni e conoscenze utili a supportare le attività decisionali delle organizzazioni.

- *Integrazione con sistemi di Intelligenza Artificiale*. L'OCR viene spesso integrato con altre soluzioni di IA per realizzare applicazioni avanzate. Tra queste figurano, ad esempio, la scansione e la lettura delle targhe dei veicoli, il riconoscimento automatico di cartelli o segnaletica contenenti testo e l'estrazione di informazioni da documenti di identità.

- *Accessibilità*. L'OCR svolge un ruolo importante nel migliorare l'accessibilità ai contenuti digitali. Infatti, i testi riconosciuti possono essere elaborati da software di sintesi vocale o da altri strumenti di assistenza, facilitando l'accesso ai dati da parte di persone ipovedenti o con disabilità visive.

3.3 Amazon Rekognition

Per l'analisi proposta è stato utilizzato *Amazon Rekognition*, un servizio cloud di AWS dedicato all'analisi di immagini e video. La piattaforma consente l'integrazione di funzionalità avanzate di computer vision all'interno di applicazioni software mediante API progettate per diversi casi d'uso. Tale servizio è basato su tecniche di deep learning e può essere utilizzato anche in assenza di competenze specifiche nell'ambito del machine learning. Inoltre, Amazon Rekognition viene costantemente aggiornato con nuovi dati, etichette e funzionalità, ampliandone progressivamente le capacità operative. In particolare, il servizio mette a disposizione diverse funzionalità per l'analisi di immagini (Amazon Rekognition Image) e video (Amazon Rekognition Video), quali *Etichette personalizzate*, *Rilevamento delle etichette* (Figura 3.2), *Proprietà dell'immagine*, *Moderazione del contenuto*, *Ricerca dei volti*, *Confronto e analisi facciale*, *Amazon Rekognition Face Liveness*, *Riconoscimento di volti celebri* (Figura 3.3), *Rilevamento PPE* e *Rilevamento dei segmenti in video archiviati*.

[4]Con il termine *data mining* si indica l'insieme di tecniche e metodologie che hanno per oggetto l'estrazione di informazioni utili da grandi quantità di dati, attraverso metodi automatici o semiautomatici.

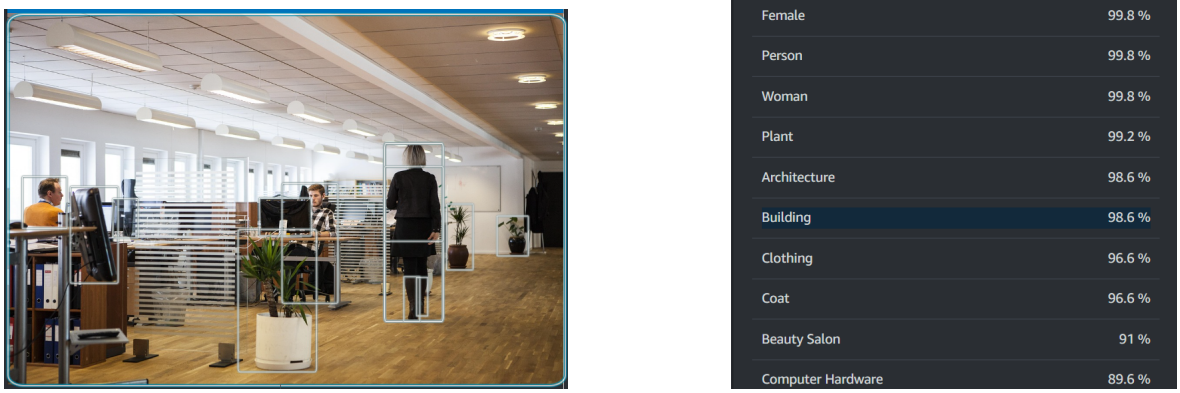


Figura 3.2: Esempio di rilevamento delle etichette con relativi risultati

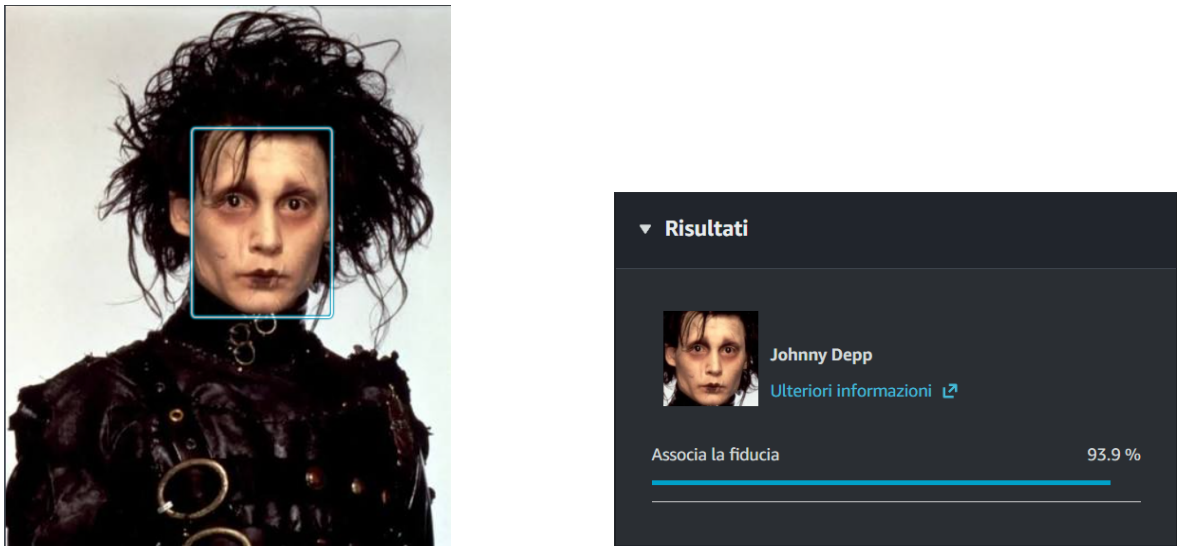


Figura 3.3: Esempio di riconoscimento di volti celebri con relativi risultati

3.3.1 Funzionamento di Amazon Rekognition

Amazon Rekognition API supporta l'analisi di tre tipologie differenti di contenuti multimediali, quali immagini, video archiviati e video in streaming. Nel presente elaborato l'attenzione è rivolta esclusivamente all'analisi delle immagini. In particolare, vengono considerate le funzionalità offerte per la categoria delle operazioni *non-storage*, per le quali il servizio non conserva alcuna informazione derivante dall'elaborazione effettuata. Amazon Rekognition consente di analizzare immagini in formato `.jpg` e `.png`. Le immagini possono essere archiviate all'interno di un *bucket* di *Amazon S3* oppure caricate direttamente dal file system locale, come avviene nei casi d'uso presentati in questo capitolo. Invece, nei casi di studio presentati nei capitoli successivi verrà utilizzato il servizio Amazon S3, illustrando le modalità di impiego dei bucket per l'archiviazione e la gestione dei contenuti da sottoporre ad analisi. Un esempio di richiesta JSON relativa all'operazione *DetectLabels*[5] mediante API è

[5]L'operazione *DetectLabels* consente di rilevare gli oggetti, i concetti e le scene presenti in un'immagine sotto forma di etichette (labels). Inoltre, essa permette di recuperare informazioni relative alle proprietà dell'immagine,

```
{
   "Image": {
      "S3Object": {
         "Bucket": "amzn-s3-demo-bucket",
         "Name": "input.jpg"
      }
   },
   "MaxLabels": 10,
   "MinConfidence": 75
}
```

Listato 3.1: Esempio di richiesta per *DetectLabels*

```
{
   {

   {
   "Labels": [
      {
         "Name": "Vehicle",
         "Confidence": 99.15271759033203,
         "Instances": [],
         "Parents": [
            {
               "Name": "Transportation"
            }
         ]
      },
      {
         "Name": "Transportation",
         "Confidence": 99.15271759033203,
         "Instances": [],
         "Parents": []
   },]
   }
   }
}
```

Listato 3.2: Esempio di output restituito da Amazon Rekognition

riportato nel Listato 3.1. In tale esempio, l'immagine di origine (`input.jpg`) viene caricata in un bucket Amazon S3 denominato `amzn-s3-demo-bucket`.

Altrimenti, le operazioni di analisi possono essere effettuate direttamente tramite la console del servizio Amazon Rekognition. L'output restituito è, generalmente, espresso in formato JSON e presenta una struttura analoga a quella riportata nel Listato 3.2. Una descrizione dettagliata dei parametri ottenuti dall'analisi e del loro significato verrà fornita nelle sezioni dedicate ai casi di studio.

3.4 Case study relativi all'individuazione di testo nelle immagini

La funzionalità di rilevamento del testo offerta da Amazon Rekognition è stata volutamente esclusa dalla panoramica precedente degli strumenti disponibili, poichè costituisce l'oggetto principale dei casi di studio presentati nelle sezioni successive. A tal fine, sono state analizzate immagini caratterizzate da un livello di complessità progressivamente crescente, partendo da scenari in cui il testo risulta chiaramente visibile fino ad arrivare a situazioni più complesse, contraddistinte dalla presenza di particolari condizioni visive o testi caratterizzati da una minore leggibilità. Per valutare il comportamento di Amazon Rekognition, il servizio è stato utilizzato direttamente dalla console, che dispone di tre sezioni accessibili dal pannello laterale:

- *risultati*: mostra gli elementi individuati nell'immagine;
- *richiesta*: visualizza la richiesta inviata al servizio sotto forma di codice JSON;

quali nitidezza, luminosità e contrasto.

- *risposta*: visualizza i risultati restituiti dall'analisi sotto forma di codice JSON.

Per approfondire l'analisi, è stata presa in esame anche la sezione *Risposta*. Questa permette di esaminare la struttura dei dati ottenuti dall'operazione *TextDetection*, che restituisce un array denominato *TextDetections*. Ciascun elemento di tale array rappresenta una riga o una parola individuata nell'immagine e contiene diversi attributi, tra cui:

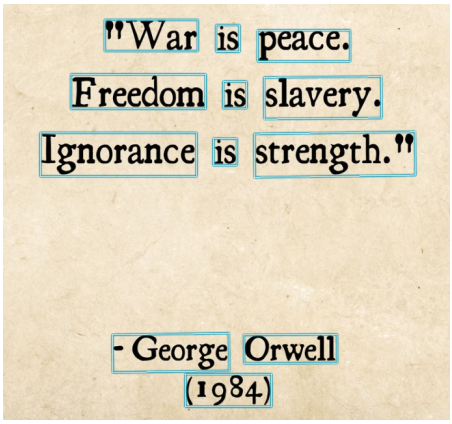
- il testo rilevato (*DetectedText*);
- la relazione che intercorre tra parole e righe (*Id and ParentId*);
- la posizione del testo all'interno dell'immagine *bounding box* (*Geometry*);
- il livello di confidenza associato sia al testo rilevato sia al relativo riquadro di delimitazione (*Confidence*);
- la tipologia dell'elemento identificato (*Type*).

I parametri sopra descritti sono stati presi in considerazione nella valutazione delle capacità del servizio. Poichè l'analisi è stata condotta direttamente da console, i parametri disponibili sono stati mantenuti invariati, al fine di osservare il comportamento di Amazon Rekognition nelle condizioni di utilizzo predefinite. Tutte le immagini sono state analizzate utilizzando la regione predefinita del servizio, corrispondente a *Stati Uniti (Virginia settentrionale)- us-east-1*, salvo diversa indicazione. Nei casi di studio presentati nelle sezioni successive, un riconoscimento è stato considerato corretto quando tutti gli elementi testuali presenti nell'immagine sono individuati senza errori. In particolare, non devono verificarsi sostituzioni o confusioni tra caratteri, associazioni errate tra parole e rispettive righe o omissioni dei segni di punteggiatura. Di conseguenza, la valutazione ha tenuto conto non solo della capacità del sistema di riconoscere correttamente le parole presenti nell'immagine, ma anche di preservarne la disposizione e la struttura all'interno del testo. Inoltre, per ciascun caso di studio, vengono illustrate le motivazioni alla base della selezione dell'immagine, l'ipotesi relativa al comportamento del sistema, i risultati ottenuti e le osservazioni emerse dalla loro analisi. Infine, si sottolinea che, per alcune immagini, l'analisi è stata ripetuta al fine di verificare l'eventuale presenza di variazioni nei risultati. Tuttavia, non sono emersi cambiamenti in nessuna delle ripetizioni effettuate.

3.4.1 Immagini con testo chiaramente leggibile

In questa sezione vengono analizzate immagini contenenti scritte chiaramente visibili e facilmente leggibili, che non dovrebbero, quindi, costituire un ostacolo alla corretta rilevazione da parte del sistema. Sono state selezionate immagini contenenti iscrizioni sia in lingua inglese sia in lingua italiana, al fine di verificare la capacità del sistema di riconoscere testi appartenenti a lingue differenti. Negli esempi riportati di seguito, le due lingue sono presenti in immagini distinte; invece, nelle sezioni successive verrà esaminato anche il caso di immagini contenenti più idiomi contemporaneamente.

La prima immagine sottoposta ad analisi è quella mostrata in Figura 3.4. Essa riporta una citazione tratta dall'opera *"1984"* di George Orwell, redatta in lingua inglese, che recita: *"War is peace. Freedom is slavery. Ignorance is strength."*. Al di sotto della citazione sono riportati il nome dell'autore e, tra parentesi, il titolo dell'opera da cui essa è tratta. L'immagine è stata scelta come punto di partenza dell'analisi, poichè presenta caratteristiche favorevoli al riconoscimento del testo: la scritta è chiaramente visibile, il font risulta facilmente leggibile e il colore dei caratteri garantisce un elevato contrasto rispetto allo sfondo chiaro. Pertanto,



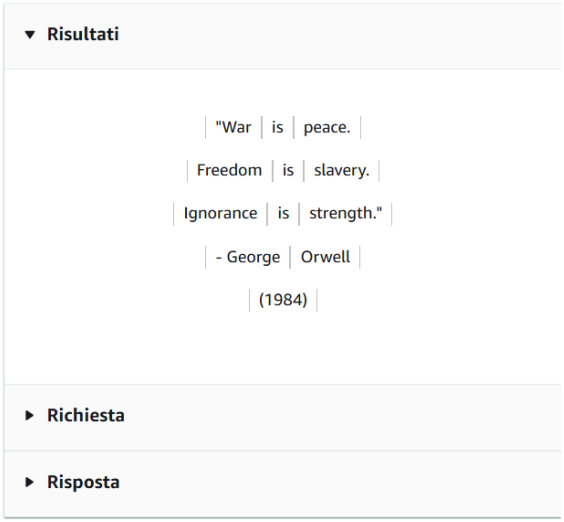


Figura 3.4: Rilevamento del testo nell'immagine contenente una citazione di *George Orwell* e relativi risultati

si ipotizza che lo strumento sia in grado di rilevare correttamente tutti gli elementi presenti nell'immagine.

Il servizio ha riconosciuto correttamente tutti gli elementi testuali, restituendo anche la punteggiatura in modo accurato, comprese le virgolette, i tre punti che separano le diverse proposizioni, il trattino e le parentesi. Un ulteriore aspetto rilevante riguarda la corretta distinzione tra caratteri maiuscoli e minuscoli, che vengono riportati senza errori di interpretazione. È possibile osservare come il testo restituito mantenga una struttura coerente con la disposizione originale presente nell'immagine. Tuttavia, gli elementi di punteggiatura non vengono riconosciuti come entità autonome, ma risultano associati alla parola alla quale sono adiacenti. Infine, posizionando il cursore sul riquadro di delimitazione (bounding box) associato a un elemento di testo individuato, viene evidenziata la corrispondente voce all'interno della sezione *Risultati*, facilitando l'associazione tra testo rilevato e la sua posizione nell'immagine.

L'immagine mostrata in Figura 3.5 contiene una citazione in lingua italiana che recita: *"Gli unici modi per fuggire dalla vita sono la pazzia e l'ironia."* Al di sotto della frase compare il nome dell'autore: *"Luigi Pirandello"*, rappresentato con caratteri differenti e con una dimensione del font inferiore rispetto a quella utilizzata per la citazione.

Come evidenziato dai risultati riportati in Figura 3.5, il servizio è stato in grado di riconoscere correttamente tutti gli elementi presenti nell'immagine, come ipotizzato. In particolare, il sistema ha identificato senza errori il passaggio dallo stampato minuscolo allo stampato maiuscolo, mantenendo la corretta formattazione del testo restituito. Analogamente, la differenza di colore tra i caratteri, che passano dal nero al giallo, non ha influenzato il processo di riconoscimento. La lieve variazione nella dimensione dei caratteri e il diverso allineamento degli elementi testuali non hanno rappresentato un ostacolo per la corretta individuazione del contenuto. Inoltre, sono stati rilevati correttamente anche i segni di punteggiatura, tra cui le virgolette poste nella parte superiore sinistra dell'immagine, nonostante fossero distanziate dal resto del testo e caratterizzate da un colore diverso, l'apostrofo presente nella parola *l'ironia* e il punto al termine della frase.

L'immagine mostrata in Figura 3.6 contiene una frase in lingua inglese composta dalle espressioni *"have a beautiful day"*, *"good"* e *"morning"*, ciascuna rappresentata mediante uno stile grafico differente. In particolare, la parola *good* presenta un font diverso rispetto alla frase iniziale, mentre *morning* si discosta ulteriormente per formato e caratteristiche stilistiche.



Figura 3.5: Rilevamento del testo nell'immagine contenente una citazione di *Luigi Pirandello* e relativi risultati

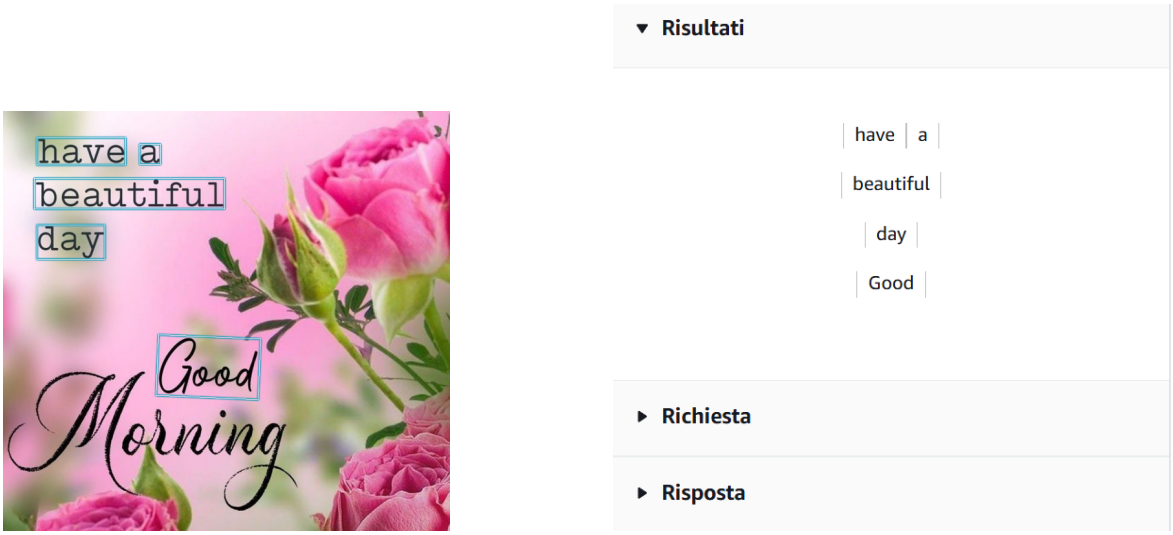


Figura 3.6: Rilevamento del testo nell'immagine contenente una scritta in lingua inglese ben visibile e relativi risultati

L'analisi effettuata da Amazon Rekognition mostra come le prime due porzioni di testo siano state correttamente riconosciute. Al contrario, l'ultima parola, ossia *"morning"*, non è stata rilevata dal sistema. Tale risultato è particolarmente significativo, poichè l'immagine era stata inizialmente considerata un caso semplice di rilevamento, non presentando caratteristiche che facessero presumere una particolare difficoltà per il sistema. Pertanto, la mancata individuazione di *morning* evidenzia come anche testi apparentemente ben visibili possano presentare elementi in grado di influire sull'accuratezza del riconoscimento. Il comportamento potrebbe essere attribuito al font utilizzato, che rende i caratteri meno facilmente distinguibili rispetto a font più convenzionali. Un'ulteriore motivazione potrebbe risiedere nella presenza di contorni poco definiti, che riducono il contrasto tra il testo e gli elementi circostanti dell'immagine, rendendo più difficile l'individuazione delle singole lettere. Nonostante tale criticità, il sistema ha continuato a rilevare correttamente l'uso di caratteri maiuscoli e minuscoli e a preservare la struttura del testo, compresi gli a capo presenti nell'immagine originale.

3.4.2 Immagini con testo leggibile

In questa sezione vengono illustrati alcuni casi di studio in cui il testo contenuto nelle immagini risulta visibile, ma presenta caratteristiche particolari che potrebbero influire sulla precisione del processo di rilevamento da parte del servizio.

La prima immagine analizzata in questa sezione (Figura 3.7) raffigura una pagina di libro contenente la celebre poesia *"L'Infinito"* di Giacomo Leopardi. L'immagine è stata selezionata con l'obiettivo di verificare se lo strumento di Amazon sia in grado di rilevare correttamente particolari elementi tipografici, quali apici e pedici, frequentemente presenti in diverse tipologie di documenti.

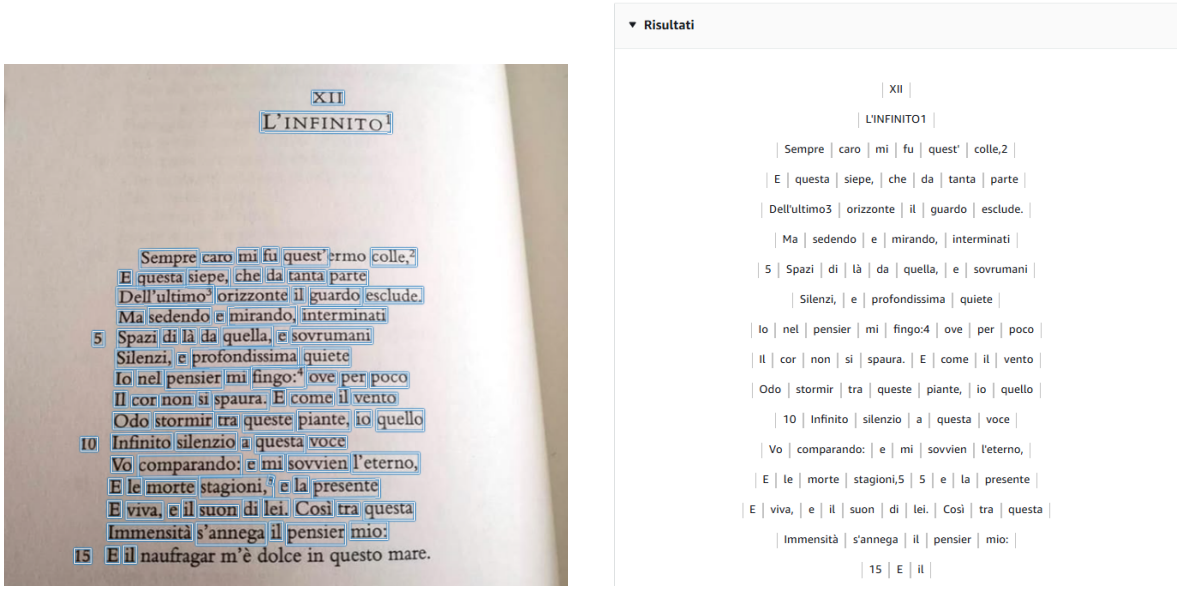


Figura 3.7: Rilevamento del testo nell'immagine contenente una pagina di libro con il componimento *L'infinito* e relativi risultati

Dall'analisi mostrata in Figura 3.7 emerge come Amazon Rekognition sia stato in grado di rilevare correttamente la maggior parte del contenuto presente nell'immagine. Tuttavia, alcune parole non sono state individuate, contrariamente a quanto previsto. In particolare, la parola *ermo* presente nel primo verso del componimento, così come gran parte dell'ultima riga della pagina. Quest'ultima imprecisione potrebbe essere ricondotta a diversi fattori. Una possibile spiegazione riguarda il limite massimo di parole rilevabili in una singola immagine, come specificato all'interno della documentazione ufficiale di AWS, che indica un limite massimo di circa 100 parole. Inoltre, la porzione di testo riportata nella pagina risulta particolarmente densa e caratterizzata da una quantità notevole di elementi ravvicinati, condizione che può rendere più complessa l'identificazione accurata di tutte le parole presenti. Oltretutto, occorre considerare che l'immagine rappresenta una fotografia di una pagina di un libro e non una scansione digitale. Di conseguenza, i caratteri non si trovano su una superficie perfettamente piana, ma seguono la naturale curvatura della pagina, condizione che può influire sulla capacità del sistema di distinguere alcuni caratteri. A ciò si aggiunge una distribuzione non uniforme dell'illuminazione, che genera alcune zone d'ombra potenzialmente in grado di ostacolare il rilevamento. Per quanto riguarda lo scopo principale dell'analisi, lo strumento si rivela in grado di individuare correttamente i numeri utilizzati come apici. Tuttavia, tali elementi tipografici non vengono interpretati come una categoria tipografica distinta. In alcuni casi il numero viene estrapolato come elemento singolo, mentre in altri risulta associato alla parola alla quale è adiacente. Ad esempio, il numero *5* presente nel verso *E le morte stagioni* viene identificato sia come parte della parola sia come elemento

indipendente. Questa osservazione trova ulteriormente supporto nelle modalità di visualizzazione dei risultati di Amazon Rekognition. Come visibile in Figura 3.7, il numero *5* risulta incluso contemporaneamente in due riquadri distinti, a indicare la doppia rilevazione effettuata dal servizio. Invece, nessuna problematica è stata riscontrata nell'individuare i numeri utilizzati per la numerazione dei versi, che sono stati riportati in modo coerente con quanto presente nell'immagine originale. Infine, è stata effettuata un'ulteriore prova modificando la regione AWS utilizzata, passando dalla configurazione predefinita alla regione *Europa(Irlanda)*. L'analisi della medesima immagine ha prodotto risultati identici a quelli illustrati in precedenza, senza evidenziare alcuna variazione nel comportamento del sistema. Tale risultato suggerisce che, nelle condizioni valutate, il processo di riconoscimento non risente della regione selezionata. La documentazione AWS specifica che versioni diverse dei modelli non sono necessariamente compatibili tra loro e possono produrre risultati differenti.

Per il successivo caso di studio, l'intento principale è volto a comprendere la capacità del servizio di individuare parole i cui caratteri non siano disposti su un'unica riga. L'immagine riportata in Figura 3.8 presenta una scritta in lingua inglese, *Good Morning*.

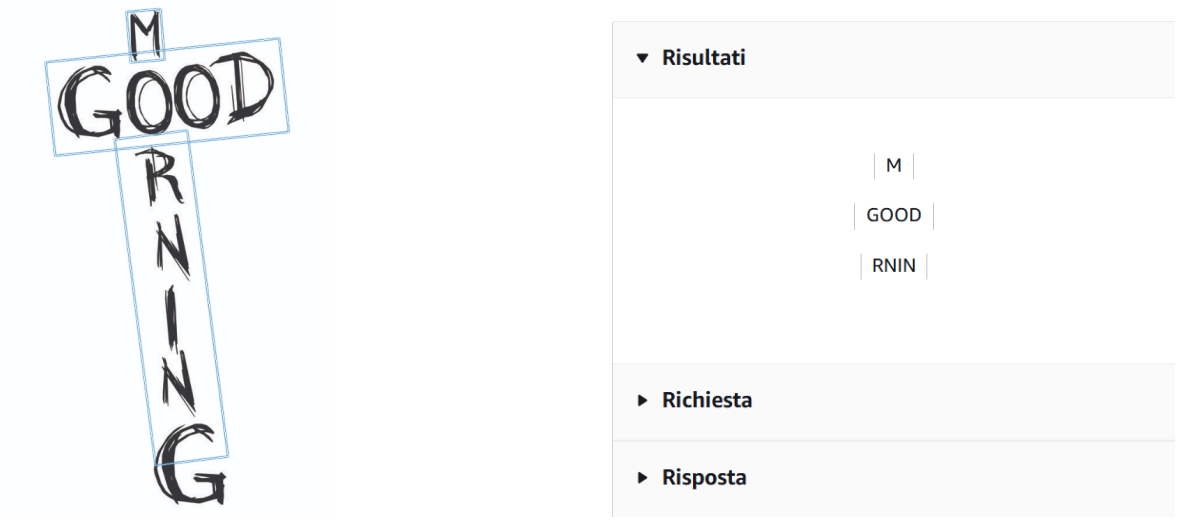
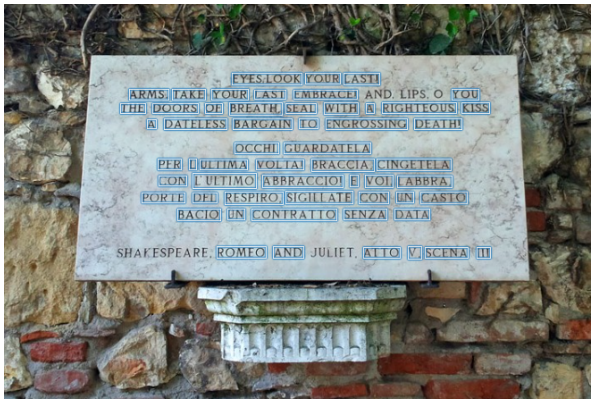


Figura 3.8: Rilevamento del testo nell'immagine contenente una scritta in lingua inglese con caratteri disposti verticalmente e relativi risultati

La particolarità di tale immagine risiede nella disposizione delle due parole: *Good* è collocata orizzontalmente, mentre *Morning* è disposta verticalmente. Inoltre, la prima lettera *o* di entrambe le parole è condivisa dalle due scritte. Avanzando un'ipotesi, si può ritenere che il servizio incontri difficoltà nella rilevazione, poichè la scritta non presenta uno stile semplice, bensì una forma caratterizzata da contorni delle lettere non perfettamente definiti. Nonostante ciò, il testo risulta ben visibile e non è mascherato da altri elementi. Inoltre, il quantitativo di caratteri risulta limitato rispetto agli esempi precedenti ed il colore, trattandosi di una scritta nera su sfondo bianco, dovrebbe semplificarne l'individuazione. Le deduzioni formulate nell'ipotesi hanno trovato riscontro nel risultato dell'analisi. Infatti, lo strumento non ha riscontrato impedimenti nell'individuare la scritta *Good*, di cui ha riconosciuto anche l'uso delle lettere maiuscole. La parola *morning*, invece, non è stata riportata correttamente e si riscontrano le seguenti incongruenze. Innanzitutto, la lettera *m* è stata riconosciuta correttamente, ma è stata separata dal resto dei caratteri. Pertanto, la lettera *o*, condivisa graficamente dalle due parole, è stata associata esclusivamente a *good* e non a *morning*. A dispetto delle aspettative, i restanti caratteri, nonostante la scritta fosse disposta in verticale, sono stati raggruppati in un'unica bounding box. Infine, l'ultima lettera, ovvero la *g*, non è stata localizzata; ciò potrebbe essere dovuto alla maggiore dimensione del carattere rispetto

ai precedenti oppure al font utilizzato. Dalla porzione di codice analizzata si può notare che il grado di confidenza con cui Amazon Rekognition afferma di aver rilevato le parole *good* e *m* è all'incirca del 98%, mentre per quanto riguarda l'elemento *RNIN*, è pari solo al 70%. In conclusione, è possibile affermare che il servizio sia stato in grado di identificare scritte non disposte in orizzontale, come di consuetudine, ma abbia riscontrato difficoltà nel riconoscere una particolare disposizione grafica.

Per la successiva trattazione si è optato per l'immagine riportata nella Figura 3.9, raffigurante la targa commemorativa dedicata all'autore inglese William Shakespeare, sulla quale è riportato un celebre passo tratto dall'Atto V, Scena III della tragedia *"Romeo e Giulietta"*. La targa, situata presso la presunta Tomba di Giulietta, nell'ex convento di San Francesco al Corso, a Verona, presenta il testo citato sia in lingua inglese sia in italiano. L'intento è, infatti, verificare se lo strumento AWS sia in grado di gestire l'identificazione di lingue differenti, caratterizzate dal medesimo alfabeto, all'interno di una stessa immagine, ipotizzandone il corretto riconoscimento.



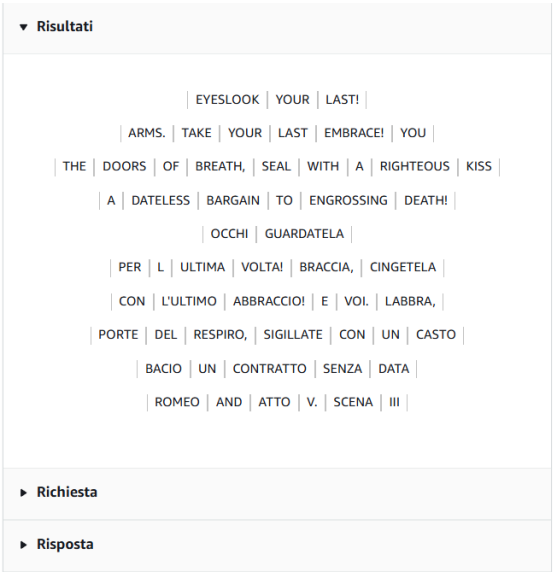


Figura 3.9: Rilevamento del testo nell'immagine contenente la targa commemorativa dedicata all'autore *William Shakespeare* e relativi risultati

Al di sotto del testo estratto dell'opera sono riportati il nome dell'autore, il titolo dell'opera, l'atto e la scena da cui è tratto. Dalla Figura 3.9 è possibile notare come non siano state rilevate particolari difficoltà nell'individuazione di parole appartenenti a lingue differenti; tuttavia, l'analisi non risulta esente da imprecisioni. In particolare, sono stati riscontrati errori nell'identificazione della punteggiatura, nell'associazione di alcune parole all'interno delle bounding box e nell'omissione di diversi termini presenti nell'immagine, quali *and, lips, o, Shakespeare* e *Juliet*. Avanzando un'ipotesi, è probabile che il servizio abbia incontrato maggiori difficoltà a causa della lunghezza del testo rispetto agli esempi precedenti e del maggior numero di righe in cui è strutturato. Inoltre, poichè il testo non è realizzato tramite un programma digitale, alcune parole presentano caratteri poco marcati o disomogeneità dovute al materiale di cui è costituita la targa, oltre che all'usura del tempo a cui la stessa risulta essere sottoposta. A ciò si aggiungono ulteriori fattori: la qualità della fotografia potrebbe rendere più complessa l'identificazione del testo; il colore delle venature che caratterizzano la targa è molto simile a quello delle scritte e, pertanto, potrebbe determinare una perdita di informazione; infine, le ridotte dimensioni degli elementi di punteggiatura potrebbero aumentare il livello di inesattezze nel riconoscimento. Inoltre, si può ipotizzare che lo strumento non effettui una vera e propria analisi semantica successiva al riconoscimento, finalizzata a

comprendere il significato delle parole identificate, ma si limiti a individuare i caratteri e a raggrupparli in parole e righe sulla base della loro disposizione nell'immagine. Pertanto, il sistema risulta in grado di rilevare parole appartenenti a lingue differenti, anche all'interno della stessa immagine.

Nel prossimo esempio, è stata selezionata un'immagine che potesse mettere in evidenza le capacità del servizio di identificare scritte caratterizzate da uno stile peculiare e non disposte su un unico piano, ma contraddistinte da una certa prospettiva. La Figura 3.10 raffigura una scala, sulla quale, su ogni gradino è riportato un verso della poesia di Giovanni Pascoli dedicata alla città di Matera, situata presso i Sassi di Matera.





Figura 3.10: Rilevamento del testo contenuto in un'immagine raffigurante una poesia di *Giovanni Pascoli*, con ciascun verso disposto su un gradino di una scala, e relativi risultati

Sulla base dei risultati ottenuti nei casi precedenti, si ipotizza che il servizio possa avere difficoltà nell'individuare tutti gli elementi presenti nell'immagine. In particolare, i primi versi del componimento potrebbero risultare più difficili da identificare a causa dell'illuminazione non omogenea e del font utilizzato per realizzare la scritta. Invece, il sistema è stato perfettamente in grado di individuare tutte le parole, riportando senza errori persino la punteggiatura, inclusi gli accenti. Inoltre, dal codice JSON prodotto dall'analisi è possibile osservare come le righe situate più in alto nell'immagine presentino un livello di confidenza maggiore nella rilevazione rispetto a quelle finali. Ciò risulta particolarmente interessante considerando che, da un punto di vista prospettico, queste ultime risultano maggiormente evidenti.

3.4.3 Immagini con testo presente su cartelli stradali

Di seguito verranno proposti diversi casi di studio con immagini contenenti cartelli stradali, compresi quelli caratterizzati da lingue differenti da quelle italiana e inglese trattate finora, al fine di comprovarne le capacità di riconoscimento in presenza di alfabeti diversi da quello latino.

La prima immagine di questa trattazione è riportata in Figura 3.11. Al suo interno sono presenti diversi cartelli che forniscono indicazioni stradali relative a una nota zona di New York. Il caso di studio mira a valutare la capacità dello strumento di individuare correttamente testo presente sui cartelli stradali, considerando le difficoltà derivanti dalle loro dimensioni

e dalla prospettiva di ripresa. Si ipotizza, pertanto, che non tutti gli elementi vengano identificati correttamente.

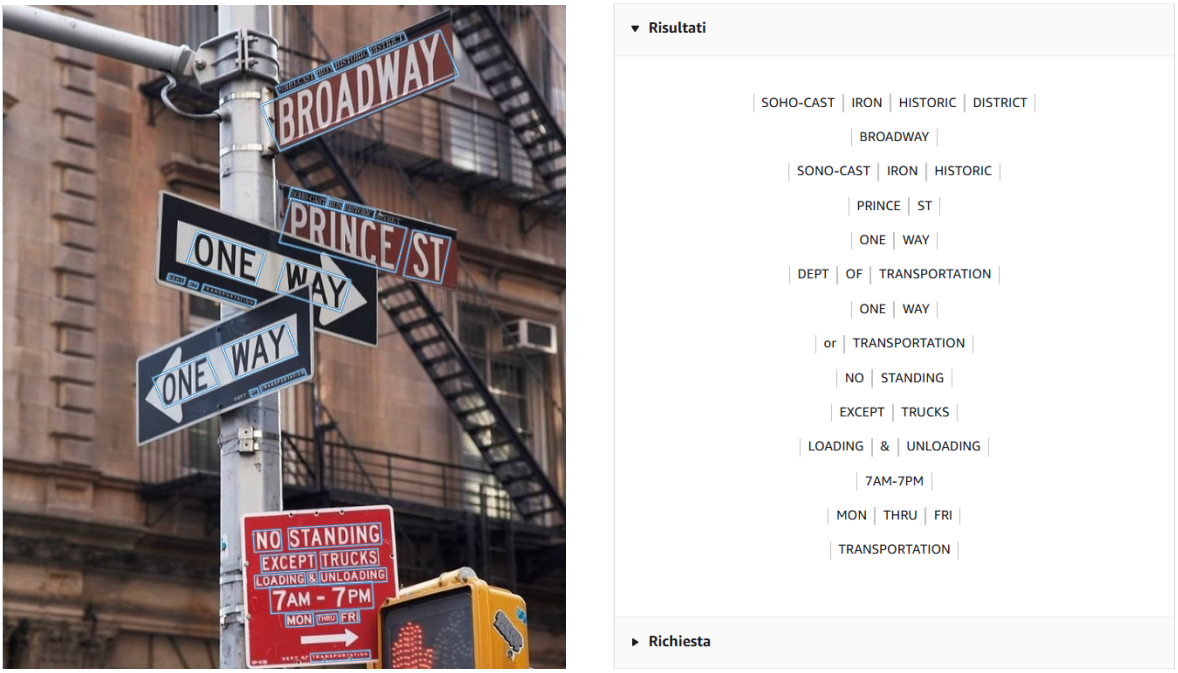


Figura 3.11: Individuazione del testo nell'immagine contenente segnaletica stradale newyorkese e relativi risultati

Nel primo cartello, situato in alto a destra, tutti gli elementi testuali sono stati correttamente individuati, compresi quelli apparentemente meno visibili. Nel cartello immediatamente sottostante, invece, la parola *DISTRICT* non è stata riportata, probabilmente a causa dei segni di usura presenti sul cartello, che non ne garantiscono una lettura precisa. Nel terzo cartello, procedendo verso il basso, sono state individuate tutte le parole, sebbene la parola *WAY* risulti leggermente coperta da un altro cartello. Anche le parole di dimensioni minori sono state rilevate correttamente. Nel cartello stradale successivo, la parola *DEPT* non è stata riconosciuta, probabilmente a causa delle sue dimensioni ridotte e della posizione che il cartello assume rispetto al punto di vista dello scatto. Inoltre, nello stesso cartello, la parola *OF* è stata sostituita con la parola *or*, in linea con le difficoltà ipotizzate in precedenza. Infine, per quanto riguarda il cartello posto nella parte inferiore della figura, tutti gli elementi rilevati riportano correttamente il contenuto della segnaletica. Altri elementi, invece, non sono stati rilevati e, pertanto, non risultano direttamente visualizzabili nella sezione *Risultati* dell'interfaccia. Analizzando il codice JSON prodotto d'analisi, è possibile notare che alcuni di essi sono stati individuati, ma con un basso grado di confidenza; ne è un esempio l'elemento *IP-430*.

Proseguendo la trattazione attraverso immagini contenenti cartelli stradali, la scelta è ricaduta su un'immagine (Figura 3.12) nella quale sono presenti indicazioni in alfabeto arabo.

Pertanto, si vuole comprendere se il servizio sia in grado di gestire alfabeti differenti da quello latino. Dall'analisi emerge che Amazon Rekognition è stato in grado di rilevare correttamente una parte della scritta in arabo; tuttavia, la prima parola non è stata riconosciuta. Le altre due righe, rispettivamente in francese e in inglese, sono state entrambe individuate correttamente, senza riportare errori di alcuna tipologia. Inoltre, è opportuno sottolineare che il servizio rileva la scritta in lingua araba seguendo il relativo ordine di lettura, ovvero da destra verso sinistra. Analizzando il codice di output, si può notare che, in realtà, l'intera riga è stata individuata, ma con un livello di confidenza dell'88%.

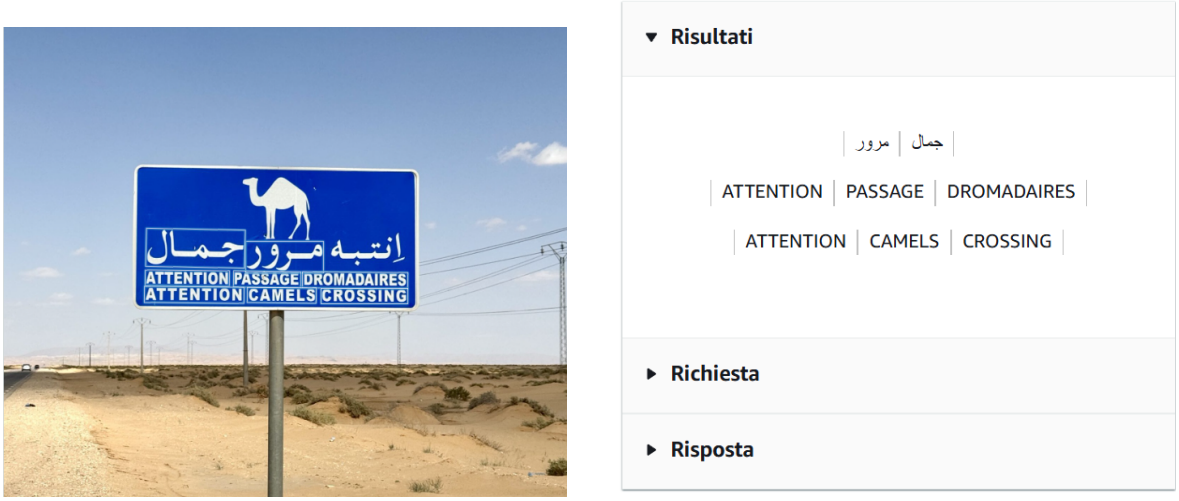


Figura 3.12: Individuazione del testo nell'immagine contenente segnaletica stradale in alfabeto arabo e relativi risultati

Nel successivo caso di studio si riporta un'immagine con un cartello stradale in primo piano, sul quale è presente una scritta in alfabeto cirillico e, sottostante ad essa, una scritta in lingua inglese. L'immagine contenuta nella Figura 3.13 è stata selezionata con l'intento di osservare il comportamento del servizio nei confronti di alfabeti differenti da quello latino, in modo analogo al precedente, e in particolare nei confronti dell'alfabeto cirillico.



Figura 3.13: Rilevamento del testo nell'immagine contenente cartello stradale in alfabeto cirillico e relativi risultati

Inoltre, essendo l'immagine scattata, presumibilmente, in condizioni di scarsa luminosità, la qualità della fotografia risulta compromessa. Si vuole evidenziare, pertanto, la presenza di un riflesso sulla segnaletica, dovuto a un'illuminazione artificiale, al fine di verificare se lo strumento incontri difficoltà in presenza di disturbi luminosi. Dalla Figura 3.13 si può evincere che il servizio non ha avuto problemi nel riconoscere gli elementi in alfabeto cirillico, tutti riportati senza errori. Analoga osservazione può essere fatta per la riga sottostante, che è stata rilevata correttamente. Quindi, la presenza del riflesso non ha costituito un ostacolo. Osservando il codice riportato nella sezione *Risposta* dell'interfaccia, si può notare che il livello di confidenza associato al riconoscimento degli elementi è superiore al 98%.

Per terminare questa analisi sulla rilevazione di alfabeti differenti da quello latino, è stata scelta un'immagine che mostra della segnaletica in lingua giapponese. In tal modo, si vuole

verificare la capacità del servizio di gestire l'individuazione di lingue basate su sistemi di scrittura logografici.



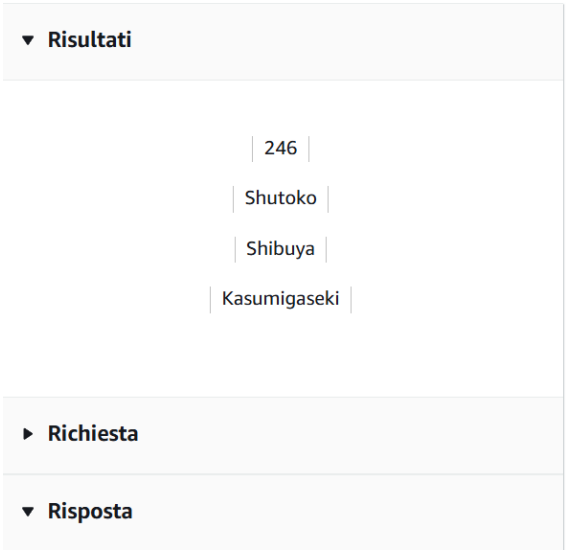


Figura 3.14: Individuazione del testo nell'immagine contenente indicazioni stradali in lingua giapponese e relativi risultati

Dal risultato riportato nella Figura 3.14 è possibile osservare che sono stati rilevati correttamente gli elementi in rōmaji[6]. Anche i numeri sono stati individuati senza alcuna difficoltà. Al contrario, tutti gli altri caratteri non sono stati rilevati. Inoltre, anche osservando il codice, le uniche parole riportate sono quelle visibili nella sezione *Risultati*. Non si è verificato, quindi, neppure quanto osservato in altri casi, ovvero che un elemento venga rilevato, ma associato a un livello di confidenza molto basso. In questo caso, il servizio non è stato in grado di individuare i caratteri giapponesi, nonostante questi siano posizionati nell'immagine in modo analogo a quelli presenti negli esempi precedenti.

3.4.4 Immagini con testo presente su insegne

In questa sezione viene valutata la capacità del servizio di individuare elementi testuali all'interno di immagini contenenti insegne, le quali possono presentare particolari caratteristiche grafiche.

In questo primo esempio, l'immagine riportata nella Figura 3.15 è stata selezionata in quanto raffigura l'insegna di una nota catena di ristoranti tematici, situata a Copenhagen. Formulando un'ipotesi, una prima possibile difficoltà che potrebbe essere riscontrata dallo strumento è rappresentata dalla presenza della scritta situata all'estrema sinistra dell'immagine (*ROCK SHOP*), poichè presenta dimensioni minori rispetto agli altri elementi ed è anche più difficilmente individuabile dal punto di vista prospettico, essendo fortemente inclinata. Una situazione analoga si presenta per gli altri due elementi non disposti su un'unica linea orizzontale. Tuttavia, rispetto alla scritta appena analizzata, i caratteri presentano dimensioni maggiori e le relative insegne sono situate in primo piano.

Dalla Figura 3.15 emergono alcune problematiche nell'individuazione di diversi elementi testuali presenti all'interno dell'immagine. Infatti, la scritta *Hard Rock* contenuta all'interno dell'insegna principale non è riportata nella sezione *Risultati*. Analogamente, per il testo presente sul lato destro dell'immagine è stata rilevata soltanto la parola *Hard*, mentre le restanti sono state tralasciate. Osservando nuovamente la sezione *Risposta*, è possibile notare

[6]Il rōmaji è il sistema di scrittura giapponese che utilizza i caratteri dell'alfabeto latino.



Risultati

CAFE

Hard

ROCK | SHOP

Richiesta

Risposta

Figura 3.15: Rilevamento del testo nell'immagine contenente l'insegna dell'Hard Rock Cafe di Copenhagen e relativi risultati

che la scritta *Hard Rock* contenuta nell'insegna di forma circolare è stata rilevata come un'unica parola, con un grado di confidenza del 64%. In modo analogo, anche la riga posta sulla destra è stata rilevata con un livello di confidenza del 75%; tuttavia, le ultime due parole, *Rock* e *Cafe*, sono state individuate come se costituissero un'unica parola. Tale risultato potrebbe essere dovuto al punto di osservazione dell'insegna, che non permette di distinguere con chiarezza gli spazi tra un carattere e l'altro.

L'immagine riportata nella Figura 3.16 è stata proposta con l'intento di verificare se il servizio sia in grado di individuare scritte presenti su insegne luminose. Infatti, la figura mostra l'insegna di benvenuto della città di Las Vegas. Inoltre, si vuole valutare la capacità del sistema di gestire immagini che raffigurano scene riprese durante le ore notturne. In tal caso, si ipotizza che il servizio possa avere problemi nel rilevamento di una scritta luminosa.



Risultati

E

TO | Fabulous

LAS | VEGAS

NEVADA

YESCO

Richiesta

Risposta

Figura 3.16: Individuazione del testo nell'immagine contenente un'insegna luminosa e relativi risultati

Dal risultato si può osservare che il servizio non è stato in grado di rilevare totalmente l'elemento luminoso *WELCOME*, del quale è stata individuata soltanto la lettera *E*. Le restanti parole sono state tutte rilevate correttamente, nonostante presentino stili e dimensioni

differenti. Pertanto, il risultato è coerente con l'ipotesi formulata, in quanto l'analisi dello strumento ha evidenziato difficoltà nella rilevazione di scritte luminose.

3.4.5 Immagini con testo presente su locandine di film

In questa sezione si vuole analizzare la capacità del servizio di rilevare testo all'interno di immagini in cui sono presenti numerosi elementi, come nel caso delle locandine cinematografiche. Inoltre, lo scopo è osservare se il sistema sia in grado di cogliere il maggior numero possibile di elementi testuali all'interno di figure in cui il testo può facilmente confondersi con lo sfondo, spesso caratterizzato da una grafica particolarmente ricca. Tale circostanza può rendere più difficoltoso il corretto riconoscimento. Partendo da quest'ultima considerazione, l'immagine oggetto di discussione è la locandina del film *Inception*, riportata nella Figura 3.17. Il contenuto della locandina comprende i nomi degli attori principali, il nome del regista, il titolo dell'opera e altre informazioni presenti nella parte inferiore del poster.

Considerata la quantità di elementi testuali presenti, è ragionevole ipotizzare che il servizio non sia in grado di individuarli tutti correttamente. Infatti, come è possibile constatare dai risultati, alcune parole non sono state rilevate, probabilmente a causa del basso contrasto con lo sfondo e delle ridotte dimensioni del testo. Inoltre, in diversi casi, gli elementi sono stati assegnati a righe differenti rispetto a quelle di appartenenza, generando una ricostruzione non corretta della struttura del testo, come evidenziato dall'analisi del codice prodotto dall'elaborazione. Le righe centrali della locandina sono state riconosciute correttamente, con poche imprecisioni relative alla punteggiatura. Le maggiori difficoltà si sono concentrate nelle righe caratterizzate da elementi di dimensioni differenti e da una disposizione non lineare, che hanno portato all'omissione di alcune parole e all'inserimento di elementi appartenenti a righe diverse. Tale risultato potrebbe essere stato influenzato dalla presenza, nella parte bassa della locandina, di diversi elementi grafici, quali loghi che contengono, a loro volta, parole e che possono contribuire a generare disordine nel riconoscimento. Quest'ultimo aspetto potrebbe essere dovuto alla vicinanza delle varie componenti testuali, che vengono, quindi, mescolate tra loro, non garantendo un'appropriata ricostruzione delle righe. Un ulteriore fattore potrebbe essere rappresentato dal quantitativo massimo di elementi identificabili, come già emerso dall'analisi della pagina di libro contenente il componimento *L'infinito* (Sezione 3.4.2).

3.4.6 Immagini con testo presente su graffiti

Per la sezione corrente sono state scelte alcune immagini che potessero mettere in evidenza le capacità del servizio di individuare elementi testuali in contesti non convenzionali. Per tale motivo, le immagini proposte raffigurano scritte riportate all'interno di pitture murali.

La Figura 3.18 riporta la scritta *Atomic* realizzata sotto forma di pittura murale. Si potrebbe supporre che il servizio non sia in grado di rilevare alcun elemento di testo.

Il testo non risulta coperto da altri elementi che possano ostacolarne la corretta identificazione; inoltre, non contiene caratteri di piccole dimensioni che possano costituire un ulteriore fattore di disturbo. La scritta è in primo piano e la luminosità dell'immagine la rende perfettamente leggibile. Nonostante queste condizioni favorevoli, lo strumento non è stato in grado di riportare il contenuto della scritta: nessuna parola è stata, infatti, evidenziata. Da ciò emerge una difficoltà del servizio nell'analisi di elementi grafici che possono contenere al loro interno parti di testo.

Per il secondo esempio, la scelta è ricaduta sull'immagine riportata nella Figura 3.19, che presenta la scritta REVOLUTION. La particolarità di quest'ultima consiste nella presenza, al suo interno, di un'ulteriore parola, *LOVE*, i cui caratteri sono, però, invertiti rispetto al loro normale senso di scrittura. Con molta probabilità, il servizio sarà in grado di individuare le







Figura 3.17: Rilevamento del testo nell'immagine contenente la locandina del film *Inception* e relativi risultati

lettere scritte nel modo convenzionale, ma non quelle invertite, perdendo, pertanto, sia il significato complessivo della parola *REVOLUTION* sia quello della scritta in essa racchiusa.

Come è possibile osservare dalla Figura 3.19, lo strumento ha correttamente rilevato la lettera *R* e l'insieme di caratteri *UTION*, come ipotizzato, riportandoli, però, su due righe differenti. La parola *LOVE*, invece, è stata interpretata dal servizio mediante caratteri appartenenti all'alfabeto cirillico, trasponendola su un'ulteriore riga. Inoltre, osservando il codice, si può notare che il livello di confidenza associato al riconoscimento di quest'ultimo elemento è pari a circa il 95%. Pertanto, un livello di confidenza elevato può essere associato anche a un'identificazione errata del testo e, viceversa, un livello di confidenza inferiore non implica necessariamente un riconoscimento scorretto, come già osservato negli esempi precedenti. Si evidenzia, inoltre, che nella documentazione di Amazon viene specificato come gli elementi che rendono un testo irriconoscibile vengono, comunque, inclusi nei risultati del rilevamento, ad esempio C@t in luogo di Cat.

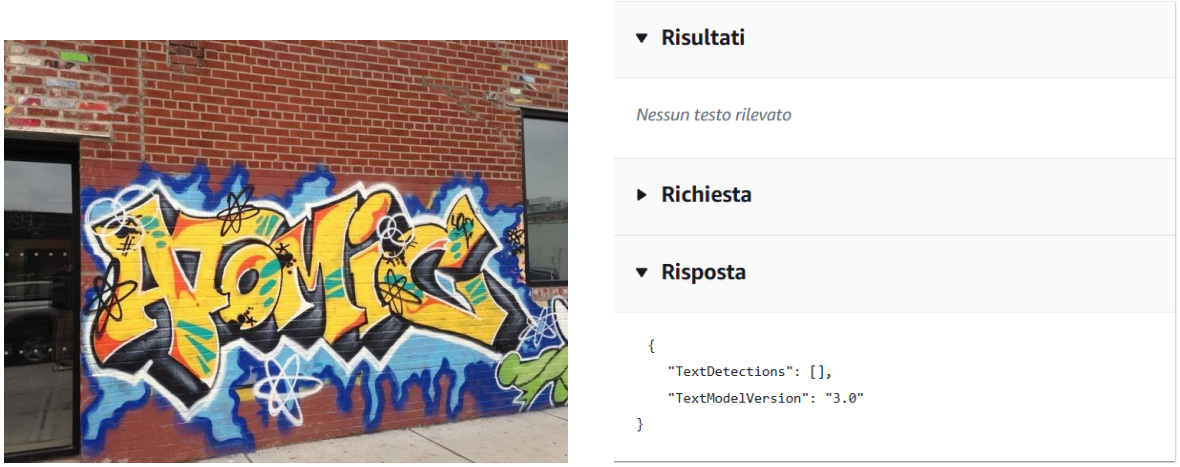


Figura 3.18: Individuazione del testo nell'immagine contenente una pittura murale e relativi risultati

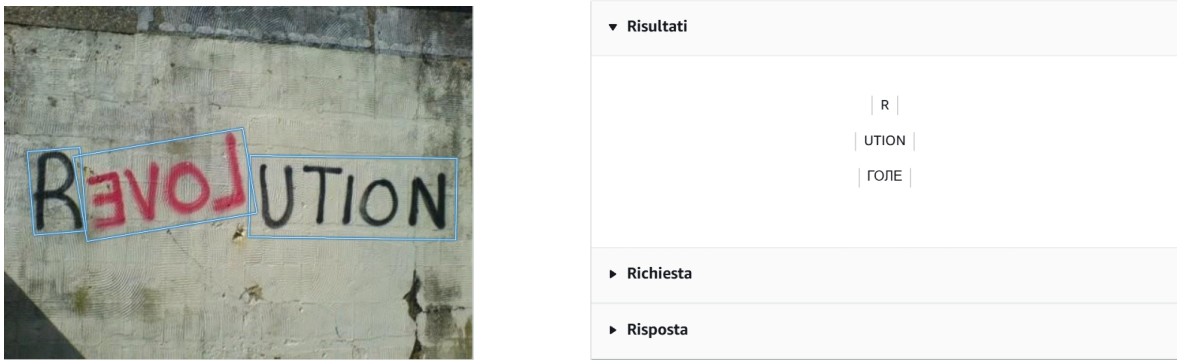


Figura 3.19: Individuazione del testo nell'immagine contenente una pittura murale con caratteri disposti in ordine inverso rispetto al normale senso di lettura e relativi risultati

3.4.7 Immagini con testo caratterizzato da aspetti particolari

In questa sezione vengono riportati una serie di casi particolari, con i quali si conclude la trattazione relativa al riconoscimento del testo da parte di Amazon Rekognition all'interno di immagini.

Per questo primo caso di studio è stata selezionata l'immagine riportata in Figura 3.20, in cui il testo è costituito da palloncini a forma di lettere che compongono la scritta *HAPPY BIRTHDAY*. Inoltre, nell'angolo in basso a destra, è presente il numero *18*. Si ipotizza che il sistema sia in grado di rilevare correttamente entrambi gli elementi testuali, poichè, nonostante la particolare conformazione della scritta, l'immagine non presenta ulteriori caratteristiche che possano ostacolarne l'identificazione.

Delle due parole presenti, soltanto la scritta *HAPPY* è stata riportata nella sezione *Risultati*, mentre la seconda è stata rilevata dal sistema, come si può evincere dalla sezione *Risposta*, ma con un livello di confidenza pari al 56%. Per quanto riguarda il numero *18*, quest'ultimo non è stato rilevato. Tale risultato potrebbe essere dovuto alla presenza di altri palloncini che lo coprono parzialmente e al fatto che una sua porzione non sia completamente contenuta nell'immagine.

Con l'immagine presentata di seguito, si vuole osservare la capacità del sistema di rilevare scritte ruotate. Inizialmente, l'immagine viene fornita nel suo normale senso di lettura, come riportato nella Figura 3.21a; successivamente, viene ruotata di 90° in senso antiorario, come mostrato nella Figura 3.21b, e infine di ulteriori 90°, sempre in senso antiorario, raggiungendo una rotazione complessiva di 180°, come riportato in Figura 3.21c.

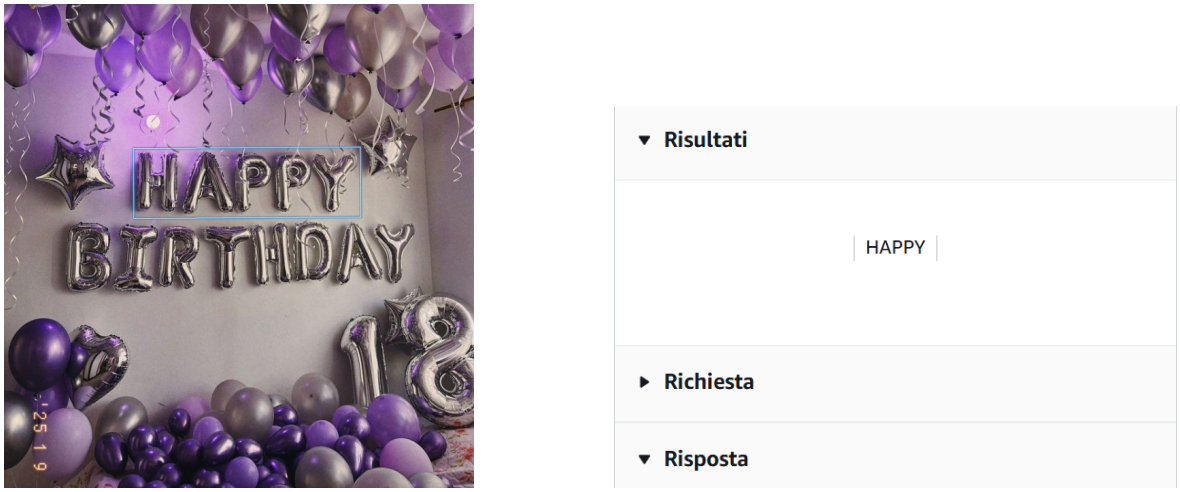
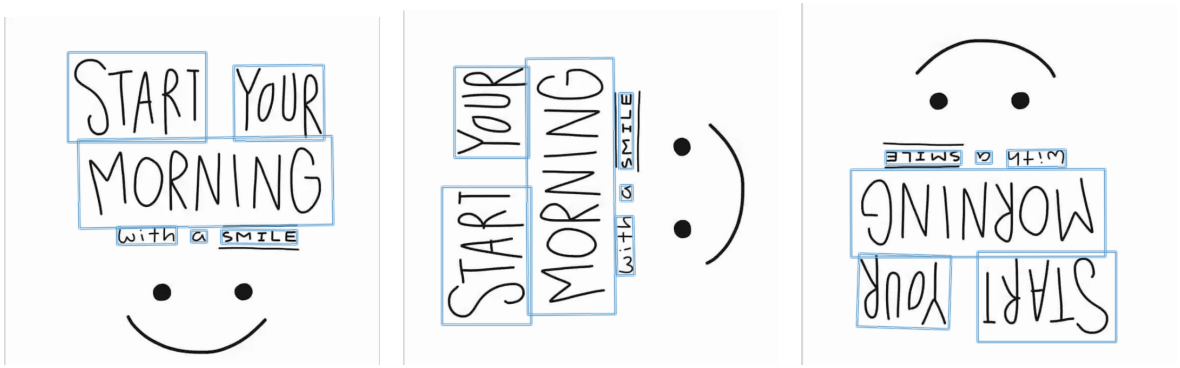


Figura 3.20: Rilevamento del testo nell'immagine contenente una scritta composta da palloncini e relativi risultati



(a) Testo non ruotato **(b)** Testo ruotato di 90° antiorario **(c)** Testo ruotato di 180° antiorario

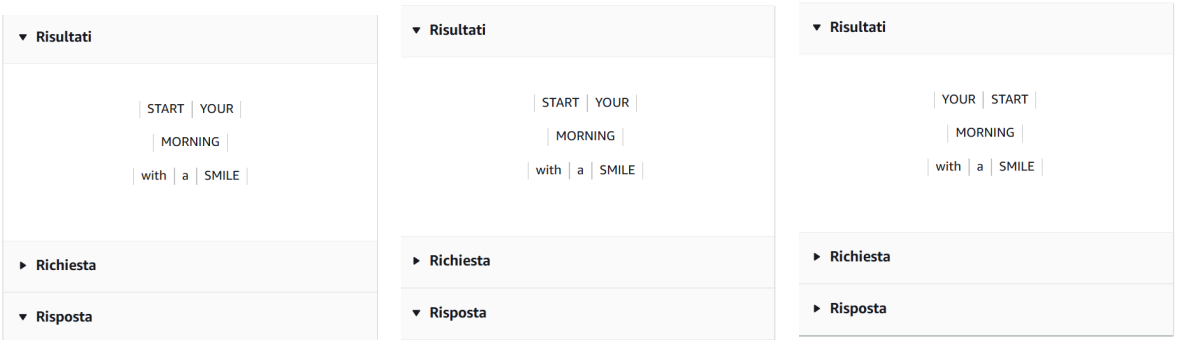
Figura 3.21: Confronto delle prestazioni di Amazon Rekognition nel riconoscimento del testo al variare del suo orientamento

Nel suo normale senso di lettura, tutti gli elementi presenti sono stati individuati correttamente. Anche nel caso in cui le scritte siano ruotate di 90° in senso antiorario, il servizio non ha avuto difficoltà nell'individuarle correttamente. Infine, nel caso in cui l'immagine sia ruotata di 180° rispetto alla posizione iniziale, tutti gli elementi sono stati rilevati correttamente. L'unica osservazione riguarda l'ordine di lettura delle parole nella prima riga, che sono state riportate in modo invertito, come mostrato in Figura 3.22c

L'immagine successiva, mostrata nella Figura 3.23, presenta una scritta distorta alla quale è stato aggiunto un effetto di dissolvenza. Tale immagine è stata scelta per comprovare la capacità del sistema di rilevare elementi testuali caratterizzati da distorsione o altri effetti grafici che potrebbero comprometterne il riconoscimento. L'ipotesi iniziale è che il servizio non sia in grado di rilevare tali elementi.

Come è possibile osservare dal risultato riportato nella Figura 3.23, il servizio non ha rilevato alcun elemento testuale, confermando l'ipotesi formulata.

L'immagine conclusiva dei case study è riportata nella Figura 3.24. Essa contiene una parola composta da altre scritte e si vuole osservare se la scritta complessiva venga rilevata oltre ai singoli elementi testuali che la compongono. Infatti, si ipotizza che solo alcuni degli elementi che compongono la parola *LIFE* vengano rilevati, mentre altri non vengano riconosciuti a causa della loro vicinanza o della loro disposizione; inoltre, si ritiene possibile



(a) Testo non ruotato **(b)** Testo ruotato di 90° antiorario **(c)** Testo ruotato di 180° antiorario

Figura 3.22: Confronto dei risultati relativi alla Figura 3.21



Figura 3.23: Rilevamento del testo nell'immagine contenente una scritta distorta con effetto di dissolvenza e relativi risultati

che la parola *LIFE* nel suo complesso non venga identificata.

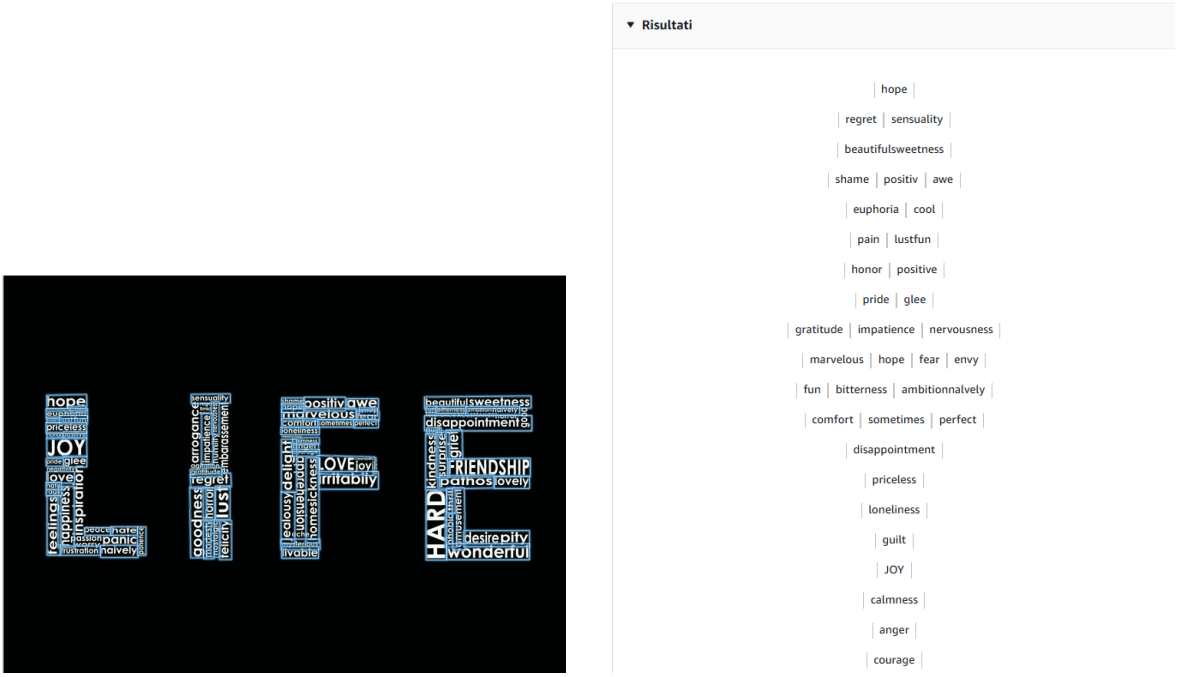


Figura 3.24: Rilevamento degli elementi testuali che compongono la parola LIFE e relativi risultati

Il risultato mostra che il servizio è stato in grado di rilevare quasi tutti i singoli elementi correttamente, comprese le parole ruotate di 90° rispetto all'asse orizzontale. Tuttavia, non ha rilevato la parola *LIFE*, all'interno della quale sono contenute le altre scritte, come ipotizzato. Questi ultimi esempi mettono in evidenza alcune delle principali difficoltà ancora riscontrate dai sistemi di riconoscimento del testo all'interno delle immagini.

CAPITOLO 4

Un case study relativo allo Speech-to-Text

In questo capitolo vengono presentati alcuni casi di studio relativi alla tecnologia Speech-to-Text, realizzati mediante il servizio Amazon Transcribe. Prima di procedere alla descrizione dei casi di studio, viene fornita un'introduzione generale nella tecnologia di riconoscimento automatico del parlato, con particolare riferimento al suo funzionamento. Successivamente, vengono illustrate le principali limitazioni che caratterizzano tali sistemi e che sono tuttora oggetto di analisi e approfondimento. Tali aspetti costituiscono, inoltre, la base per la selezione delle tracce audio impiegate nei casi di studio. Vengono, quindi, presentati i principali casi d'uso della tecnologia Speech-to-Text. A conclusione della parte introduttiva, viene presentato il servizio Amazon Transcribe, al fine di fornire gli elementi necessari alla comprensione dei successivi casi di studio. In tale contesto, viene, inoltre, approfondito il servizio Amazon S3, utilizzato per la gestione delle tracce audio impiegate nelle sperimentazioni. A conclusione del capitolo, vengono descritti i casi di studio relativi alla trascrizione del linguaggio parlato, realizzati con l'obiettivo di valutare le prestazioni del servizio, evidenziandone gli aspetti positivi e le eventuali problematiche emerse dall'analisi dei risultati.

4.1 Introduzione allo Speech-to-Text

Nella sezione corrente viene introdotto il concetto di *Speech-to-Text*, al fine di fornire le conoscenze necessarie alla comprensione dei casi di studio presentati nel capitolo. Con il termine *Speech-to-Text* si indica il processo di conversione del linguaggio parlato in una trascrizione testuale. Tale tecnologia è basata su tecniche di Intelligenza Artificiale e, in particolare, sul *riconoscimento vocale automatico* (Automatic Speech Recognition, ASR). I sistemi di riconoscimento vocale automatico sono spesso disponibili come servizi cloud, che permettono di elaborare automaticamente l'input sonoro e produrre la relativa trascrizione testuale.

4.1.1 Funzionamento dello Speech-to-Text

Generalmente, un sistema di *Speech-to-Text* converte un segnale vocale in una sequenza testuale attraverso una serie di fasi, come mostrato nella Figura 4.1. Tali fasi comprendono:

- *Input vocale e pre-elaborazione*: il segnale audio contenente il parlato viene acquisito, generalmente tramite un microfono, e può essere sottoposto a operazioni preliminari finalizzate a migliorarne la qualità e, di conseguenza, l'accuratezza del riconoscimento. Tali operazioni possono includere, a seconda del sistema adoperato, la rimozione del rumore di fondo e delle frequenze irrilevanti, la segmentazione della traccia audio in

frammenti di più semplice elaborazione, la stabilizzazione del livello del volume e la conversione del file audio nel formato richiesto. Il segnale ottenuto viene, quindi, fornito al sistema di riconoscimento.

- *Analisi ed estrazione delle caratteristiche*: il segnale audio viene analizzato ed elaborato al fine di estrarre le informazioni rilevanti per il riconoscimento del parlato.
- *Riconoscimento e decodifica*: un modello acustico o, nei sistemi più recenti, un modello *end-to-end*, interpreta la rappresentazione del segnale e determina la sequenza di parole (o token) più probabile, sfruttando, ove previsto dall'architettura, anche informazioni di tipo linguistico.
- *Output*: il testo finale viene formattato, includendo la punteggiatura necessaria a renderne più agevole la lettura.

Le fasi descritte rappresentano gli elementi comuni ai sistemi di *Speech-to-Text*, i quali possono, tuttavia, essere implementati mediante architetture differenti.

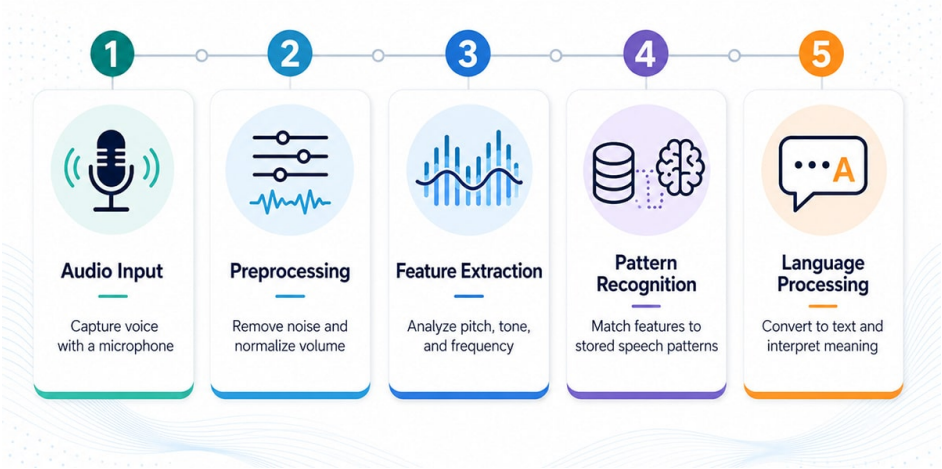


Figura 4.1: Schema illustrativo del processo di riconoscimento vocale automatico (`https://it.shaip.com/`)

Una possibile distinzione può essere effettuata tra approcci tradizionali, caratterizzati dalla presenza di componenti specifici per l'analisi acustica e linguistica, e approcci moderni, basati su modelli neurali *end-to-end*. Sebbene entrambi abbiano l'obiettivo di convertire il segnale in una sequenza testuale, essi differiscono nelle modalità con cui l'informazione linguistica viene rappresentata, come illustrato di seguito.

- *Approccio basato su unità fonetiche*. Nei sistemi tradizionali, il segnale audio viene convertito in una rappresentazione delle sue caratteristiche acustiche. Il modello di riconoscimento sfrutta tali informazioni al fine di individuare unità linguistiche elementari, quali i fonemi[1]. Le sequenze di unità fonetiche vengono utilizzate, insieme alle informazioni fornite da un modello linguistico, per determinare la sequenza di termini più probabile. In tale approccio, le componenti dedicate alla modellazione acustica e linguistica e alla decodifica sono distinte.
- *Approccio neurale end-to-end*. Nei moderni sistemi basati su tecniche di *deep learning*, il segnale viene convertito in una rappresentazione numerica delle sue caratteristiche

[1] I fonemi costituiscono le più piccole unità sonore distintive di una lingua e permettono di distinguere parole con significati differenti. Le principali classi di fonemi sono le vocali e le consonanti.

nel tempo e in frequenza, quale uno spettrogramma[2]. Tale rappresentazione viene elaborata da parte di un modello neurale che apprende direttamente la relazione fra il segnale acustico e la sequenza testuale. Un esempio è costituito dalle architetture *encoder-decoder*, nelle quali l'encoder produce la rappresentazione dell'input e il decoder (essenzialmente un modello linguistico condizionale) genera progressivamente token della trascrizione ad ogni timestamp, utilizzando sia le informazioni provenienti dall'encoder sia il contesto della sequenza generata in precedenza.

Nel primo approccio, il riconoscimento avviene attraverso una sequenza di componenti specializzate, mentre nei modelli *end-to-end* la conversione dal segnale audio al testo viene appresa in modo maggiormente integrato.

4.1.2 Difficoltà nel riconoscimento del linguaggio parlato

Nonostante i notevoli miglioramenti raggiunti negli ultimi anni, i sistemi di Speech-to-Text presentano ancora alcune limitazioni che possono rendere più complesso il riconoscimento accurato del parlato. Tali aspetti sono stati utilizzati come criteri per selezionare le tracce audio sottoposte ad analisi nei casi di studio presentati nel capitolo. Le principali difficoltà sono elencate di seguito:

- *Dimensione del vocabolario*. Le conversazioni e i contenuti multimediali sottoposti ad analisi possono, spesso, contenere un quantitativo di vocaboli molto elevato, inclusi termini poco frequenti o specifici di determinati contesti. Un vocabolario più ampio può rendere più complesso il riconoscimento automatico rispetto a contesti caratterizzati da un insieme di vocaboli più limitato.

- *Diarizzazione*. In presenza di più parlanti, il sistema deve essere in grado di distinguere i diversi interventi vocali. Tale aspetto risulta particolarmente rilevante, ad esempio, nella trascrizione di riunioni.

- *Canale e rumore*. La presenza di rumore di fondo può rendere più complesso il processo di trascrizione, in particolare quando la registrazione è disturbata o quando i parlanti si trovano a una distanza significativa dal dispositivo di acquisizione dell'audio.

- *Accento e caratteristiche del parlante*. L'identificazione del parlato può risultare più accurata quando le caratteristiche linguistiche del parlante, quali l'accento o la varietà linguistica utilizzata, sono adeguatamente rappresentate nei dati utilizzati per l'addestramento del sistema. Analogamente, il riconoscimento del parlato infantile può risultare più complesso nel caso in cui il modello sia stato addestrato prevalentemente o esclusivamente su registrazioni di parlanti adulti.

- *Volume*. Un volume troppo basso, ad esempio quello del parlato sussurrato, può rendere più difficile il riconoscimento dei vocaboli. Analogamente, variazioni significative dell'intensità vocale, ad esempio un parlato con un volume particolarmente elevato, possono influire negativamente sull'individuazione dei vocaboli.

- *Velocità.* Una velocità del parlante eccessivamente elevata o, viceversa, particolarmente ridotta può compromettere il riconoscimento accurato del parlato da parte del sistema.

[2]Uno spettrogramma rappresenta la distribuzione delle frequenze presenti nelle forme d'onda in funzione del tempo.

4.1.3 Casi d'uso dello Speech-to-Text

Lo Speech-to-Text trova applicazione in diversi contesti. Tra i principali si possono individuare i seguenti:

- *Call center e agenti*. Gli assistenti vocali dotati di tecnologia IA possono utilizzare sistemi di Speech-to-Text per trascrivere automaticamente le interazioni con i clienti, indirizzare le chiamate sulla base delle esigenze espresse, ricavare informazioni dalle conversazioni con i clienti ed eseguire sentiment analysis. Nel contesto dei call center dedicati al servizio clienti, il sistema può gestire le domande più semplici dei clienti e indirizzare le richieste più complesse agli agenti umani.
- *Servizi di trascrizione*. I sistemi di Speech-to-Text possono essere integrati all'interno di software per consentire la trascrizione automatica di riunioni online o webinar, la generazione di verbali e la creazione di sottotitoli e disascalie per i contenuti video. Alcune applicazioni sono progettate per adattarsi a contesti specifici, quali quello sanitario, legale e formativo.
- *Riconoscimento vocale*. Integrato con tecniche di elaborazione del linguaggio naturale, il riconoscimento vocale consente di interpretare l'input vocale dell'utente e tradurlo in comandi eseguibili. Questa tecnologia può essere impiegata, ad esempio, per effettuare telefonate o controllare dispositivi intelligenti presenti all'interno di un'abitazione domotica, quali luci e termostati, attraverso chatbot e assistenti digitali, come *Alexa*, *Google Assistant* e *Cortana*.
- *Accessibilità*. Le persone con disabilità possono utilizzare sistemi di riconoscimento vocale per interagire con computer e smartphone mediante comandi vocali. Tali strumenti consentono, quindi, di comporre messaggi, digitare testo, prendere appunti e scrivere e-mail, senza dover ricorrere necessariamente all'interazione tramite tastiera.
- *Monitoraggio dei contenuti*. I sistemi che integrano l'IA possono esaminare il contenuto delle trascrizioni ottenute da video e registrazioni audio al fine di individuare eventuale materiale inappropriato. In questo contesto, la tecnologia può supportare le attività di moderazione dei contenuti, segnalando agli operatori gli elementi che richiedono un'ulteriore verifica.

4.2 Amazon Transcribe

In questa sezione viene presentato *Amazon Transcribe*, il servizio di riconoscimento vocale automatico utilizzato nei casi di studio descritti nella Sezione 4.3. Il servizio sfrutta modelli di apprendimento automatico per convertire il parlato in testo e può essere utilizzato sia autonomamente che integrato in applicazioni software. Esso supporta numerose lingue e offre funzionalità specifiche per diversi ambiti, tra cui le soluzioni dedicate al settore medico, come *Amazon Transcribe Medical* e *AWS HealthScribe*. L'analisi seguente si concentra esclusivamente sul servizio standard, con particolare attenzione ad alcune delle funzionalità da esso messe a disposizione.

4.2.1 Funzionamento di Amazon Transcribe

Amazon Transcribe permette di convertire il parlato contenuto in un input audio in testo attraverso due modalità principali: *batch* e *streaming*. La modalità *batch* permette di elaborare file audio precedentemente caricati all'interno di un bucket Amazon S3, mentre la modalità

streaming consente di elaborare il flusso audio in tempo reale e restituire progressivamente la trascrizione testuale. Nei casi di studio presentati viene adoperata la modalità batch. Infatti, i file sottoposti ad analisi sono stati precedentemente caricati in un opportuno bucket S3. Il procedimento relativo al caricamento e alla gestione dei file viene illustrato nella Sezione 4.2.2. Il servizio supporta diversi formati di file audio per l'elaborazione in modalità batch, quali AMR, FLAC, WAV, Web4, MP4 e MP3. La Figura 4.2 mostra la sezione *Processi di trascrizione*, accessibile dalla console di Amazon Transcribe, da cui è possibile creare un nuovo processo di trascrizione mediante l'opzione *Crea processo*, nonchè visualizzare quelli precedentemente avviati.

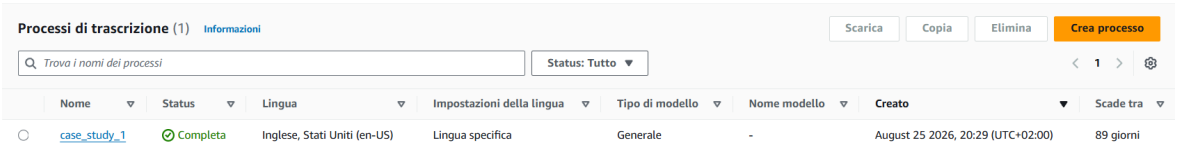


Figura 4.2: Sezione *Processi di trascrizione* della console di Amazon Transcribe

Selezionando un determinato processo è possibile visualizzarne i dettagli. In particolare, la Figura 4.3 mostra la sezione *Dettaglio processo*, nella quale sono riportate le informazioni relative all'input utilizzato per la trascrizione. Le ulteriori impostazioni disponibili per la creazione di un processo sono descritte nella Sezione 4.2.1.

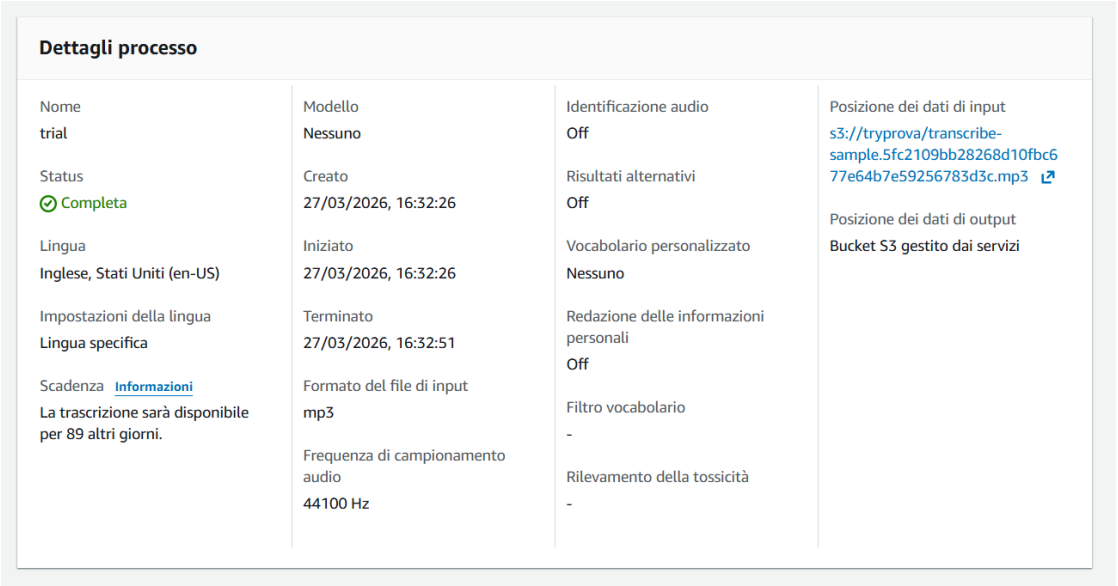


Figura 4.3: Sezione *Dettagli processo*

Al di sotto della sezione dedicata ai dettagli del processo è disponibile un'anteprima della trascrizione, come mostrato in Figura 4.4. Attraverso il comando *Scarica* è possibile salvare una copia locale del file JSON contenente l'output prodotto oppure, qualora siano stati richiesti, i file dei sottotitoli nel formato WebVTT (`.vtt`) e SubRip (`.srt`). Nella sezione *Testo* viene visualizzata la trascrizione completa, comprensiva della punteggiatura introdotta automaticamente sulla base del parlato rilevato. Selezionando le singole parole identificate, è possibile visualizzare l'intervallo temporale in cui ciascuna di esse è presente nella traccia audio, espresso mediante il tempo di inizio e di fine in secondi a partire dall'inizio della registrazione. Inoltre, per ciascun elemento riconosciuto, viene riportato il grado di confidenza. Invece, la sezione *Sottotitoli* consente di visualizzare un'anteprima dei sottotitoli generati.

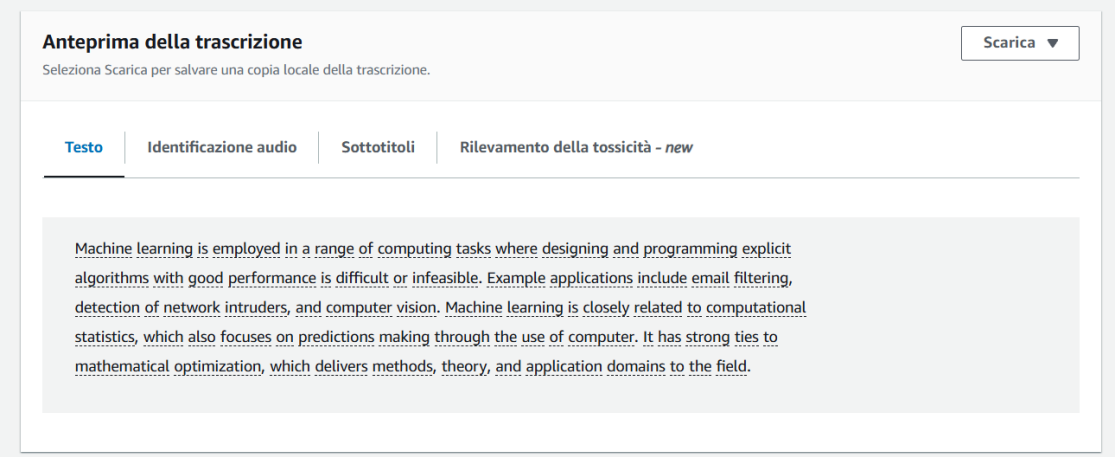


Figura 4.4: Sezione *Anteprima della trascrizione*

Analogamente a quanto illustrato nel capitolo relativo al servizio Amazon Rekognition, Amazon Transcribe mette a disposizione API che consentono di interagire con il servizio. Nel Listato 4.1 viene riportato un esempio di richiesta API per l'avvio di un processo di trascrizione.

```
{
   "TranscriptionJobName": "case_study_1",
   "LanguageCode": "en-US",
   "MediaSampleRateHertz": 44100,
   "MediaFormat": "mp3",
   "Media": {
      "MediaFileUri": "s3://tryprova/transcribe-sample.mp3"
   }
}
```

Listato 4.1: Esempio di richiesta per una trascrizione

Invece, nel Listato 4.2 viene riportato un esempio dell'output in formato JSON atteso per una richiesta di trascrizione. In modalità batch, l'output prevede le sezioni *transcripts*, contenente la trascrizione completa in un blocco di testo unico; *items*, relativa ai singoli elementi riconosciuti; *audio_segments*, contenente un discorso con pause o interruzioni minime, in cui ciascun *audio segment* comprende gli identificatori degli elementi appartenti al segmento. Per ciascun elemento vengono, inoltre, indicate, quando disponibili, informazioni quali gli intervalli temporali, il grado di confidenza e la tipologia, che può essere *pronunciation* o *punctuation*.

Creazione di un processo di trascrizione

In questa sezione vengono illustrate le principali impostazioni disponibili per la creazione di un processo di trascrizione, con particolare attenzione alle funzionalità utilizzate nei casi di studio descritti nel capitolo. L'obiettivo è fornire una descrizione della metodologia adottata per la configurazione delle diverse trascrizioni. Quindi, la console consente di personalizzare l'analisi tramite una serie di funzionalità, come mostrato in Figura 4.5.

Tra le impostazioni principali figurano le seguenti:

- *Nome*: consente di specificare il nome da assegnare al processo di trascrizione.
- *Impostazioni della lingua*: consente di specificare la lingua da utilizzare per la trascrizione. In particolare, sono disponibili le seguenti modalità:

```
{
{
   "jobName": "my-first-transcription-job",
   "accountId": "111122223333",
   "results": {
      "transcripts": [
         {
            "transcript": "Welcome to Amazon Transcribe."
         }
      ],
      "items": [
         {
            "id": 0,
            "start_time": "0.64",
            "end_time": "1.09",
            "alternatives": [
               {
                  "confidence": "1.0",
                  "content": "Welcome"
               }
            ],
            "type": "pronunciation"
         },
   ....
         {
            "id": 4,
            "alternatives": [
               {
                  "confidence": "0.0",
                  "content": "."
               }
            ],
            "type": "punctuation"
         }
      ],
      "audio_segments": [
         {
            "id": 0,
            "transcript": "Welcome to Amazon Transcribe.",
            "start_time": "0.64",
            "end_time": "2.56",
            "items": [
               0,
               1,
               2,
               3,
               4
            ]
         }
      ]
   },
   "status": "COMPLETED"
}
}
```


Listato 4.2: Esempio di output restituito da Amazon Transcribe

  - *Lingua specifica*: consente di selezionare la lingua parlata nella traccia audio.
  - *Identificazione automatica della lingua*: consente al servizio di identificare automaticamente la lingua prevalente nella registrazione audio. È possibile indicare una o più lingue presumibilmente presenti nella traccia, restringendo, così, l'insieme delle lingue considerate dal sistema.
  - *Identificazione automatica di più lingue*: consente di identificare la presenza di più lingue all'interno della stessa traccia audio, con la possibilità di specificare preventivamente l'elenco delle lingue che possono essere presenti.

  Nel caso della rilevazione automatica, viene riportato il grado di confidenza riposto nell'identificazione della lingua, espresso mediante un valore compreso fra 0 e 1. Inoltre, le impostazioni relative alla lingua influenzano la formattazione della trascrizione, inclusi l'utilizzo della punteggiatura, delle lettere minuscole e maiuscole e la rappresentazione dei numeri. Tali caratteristiche dipendono dalle relative regole di formattazione supportate dal servizio per la lingua selezionata e contribuiscono a spiegare alcuni dei risultati ottenuti nelle trascrizioni dei casi di studio trattati.

- *Tipo di modello*: consente di selezionare il modello da utilizzare per la trascrizione.

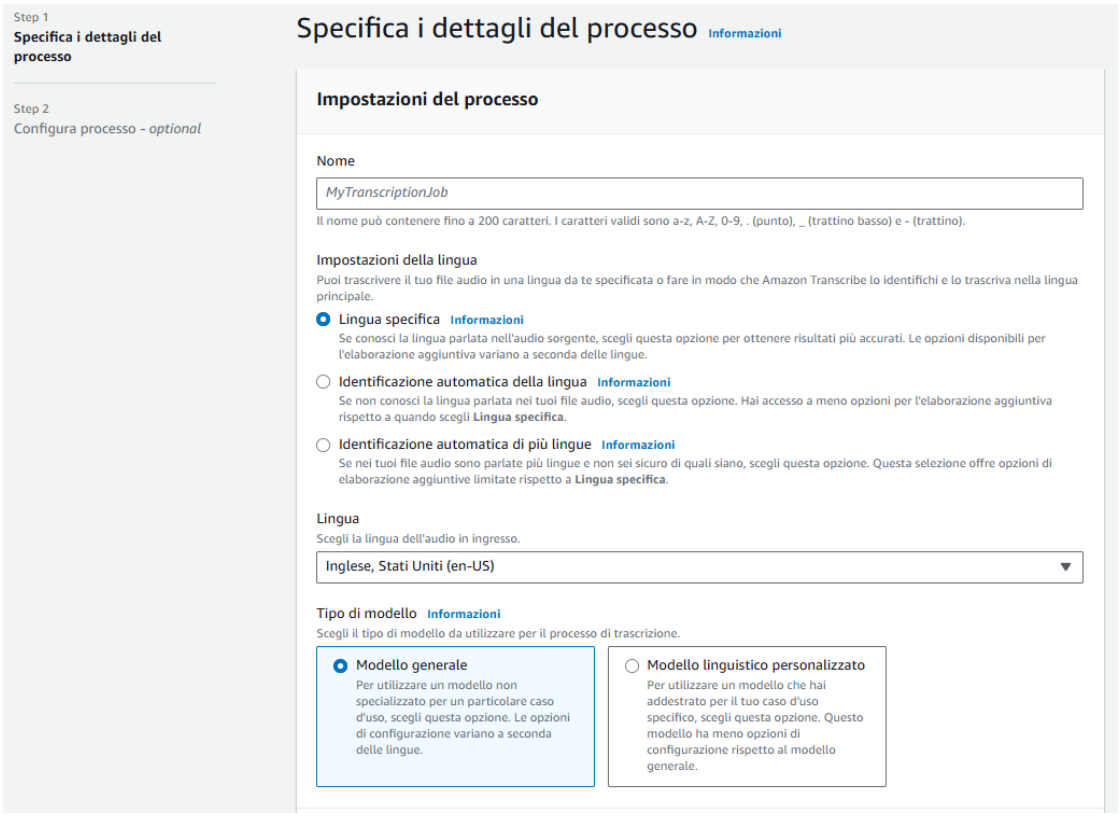


Figura 4.5: Schermata per la selezione delle impostazioni del processo

Sono disponibili un *modello generico*, adatto a scenari di utilizzo generali, e un *modello linguistico personalizzato*, che consente di utilizzare un modello addestrato per migliorare il riconoscimento di terminologia specifica di un determinato dominio.

La configurazione del processo può essere completata mediante la selezione di ulteriori funzionalità opzionali, come illustrato in Figura 4.6. Tra quelle analizzate nei casi di studio figurano le seguenti:

- *Identificazione audio*: abilitando la funzione *Suddivisione di chi parla*, è possibile identificare i parlanti presenti nella traccia audio e associare a ciascuno di essi le porzioni di testo corrispondenti. È, inoltre, possibile specificare il numero di parlati previsto, al fine di ottenere una maggiore accuratezza dei risultati. Le informazioni relative ai parlanti vengono riportate nell'output della trascrizione attraverso le etichette associate agli speaker.

- *Rilevamento delle tossicità*: questa funzionalità consente di analizzare le conversazioni per individuare contenuti potenzialmente tossici, utilizzando sia le informazioni testuali sia i segnali audio, tra cui l'intonazione e il tono della voce. Il servizio classifica i contenuti in diverse categorie, tra cui volgarità, incitamento all'odio, riferimenti sessuali espliciti, linguaggio offensivo, violenza, minacce, molestie e abusi. Per ciascun segmento della conversazione viene assegnato un punteggio di confidenza compreso tra 0 e 1, che indica la probabilità di appartenenza a ciascuna categoria, con valori più elevati corrispondenti a una maggiore probabilità. I risultati sono visualizzabili nella sezione *Rilevamento delle tossicità* dell'*Anteprima della trascrizione*.

- *Filtraggio del vocabolario*: questa funzionalità permette di specificare un insieme di parole da modificare nell'output della trascrizione. Per ciascun termine è possibile applicare una delle seguenti modalità di visualizzazione:
  - *mask*: sostituisce la parola specificata con tre asterischi $***$;
  - *tag*: associa un'etichetta alla parola senza modificarne il contenuto;
  - *remove*: elimina la parola dall'output, lasciando uno spazio nella relativa posizione.

  Tale funzionalità può essere utilizzata, ad esempio, per filtrare contenuti considerati inappropriati o informazioni che non si desidera riportare nella trascrizione.
- *Vocabolario personalizzato*: questa funzionalità consente di migliorare il riconoscimento e la formattazione di termini specifici che possono risultare difficili da identificare per un modello generico, come termini tecnici, nomi di marchi e acronimi. Il vocabolario può essere creato tramite la console di gestione AWS oppure mediante un file caricato su Amazon Transcribe o su un *bucket* di Amazon S3. Il file presenta una struttura tabellare in cui la colonna *Phrase*, obbligatoria, specifica i termini da riconoscere, mentre la colonna facoltativa *DisplayAs* consente di definirne la modalità di visualizzazione nell'output della trascrizione. Le colonne *SoundsLike* e *IPA* non risultano, invece, supportate nella configurazione considerata.

4.2.2 Funzionamento di Amazon S3

In questa sezione viene illustrato il funzionamento di Amazon S3, fornendo una panoramica del suo utilizzo nei case study descritti nel presente capitolo e delle modalità con cui verrà impiegato nel caso di studio presentato nel capitolo successivo. Amazon S3, già presentato nella Sezione 2.4, consente di archiviare un *oggetto*, costituito da un file e i relativi metadati, all'interno di *bucket*, ovvero contenitori utilizzati per l'archiviazione e la gestione degli oggetti. In particolare, esso supporta quattro tipologie di bucket, ovvero *bucket per uso generico*, *bucket di directory*, *bucket di tabelle* e *bucket vettoriali*.

Nella presente trattazione sono stati utilizzati bucket per uso generico, adatti alla maggior parte dei casi d'uso e dei modelli di accesso. Durante la creazione di un bucket è necessario specificare la regione AWS e configurare le relative impostazioni di accesso. Le policy consentono di definire quali servizi possono accedere alle risorse contenute nel bucket e possono essere gestite tramite *AWS Identity and Access Management*. In particolare, ogni oggetto archiviato in Amazon S3, così come il bucket che lo contiene, è identificabile mediante un URI, che consente di fare riferimento alla risorsa all'interno di altri servizi AWS. Tale funzionalità è stata sfruttata nei casi di studio mediante la creazione del bucket `case-study-transcribe`, nel quale sono state caricate le tracce audio successivamente elaborate da Amazon Transcribe, come mostrato in Figura 4.7.

4.3 Case study relativi alla trascrizione del linguaggio parlato

In questa sezione viene analizzata la capacità di Amazon Transcribe di convertire il linguaggio parlato contenuto nelle tracce audio in testo. Oltre alla trascrizione, il servizio restituisce informazioni aggiuntive relative al contenuto riconosciuto, tra cui i punteggi di affidabilità e i timestamp associati alle parole o ai segni di punteggiatura individuati. Per tutti i casi di studio è stato utilizzato il formato audio MP3. L'analisi è stata condotta aumentando progressivamente il grado di complessità del parlato contenuto nella traccia audio, partendo da situazioni in cui il parlato risulta facilmente identificabile fino ad arrivare

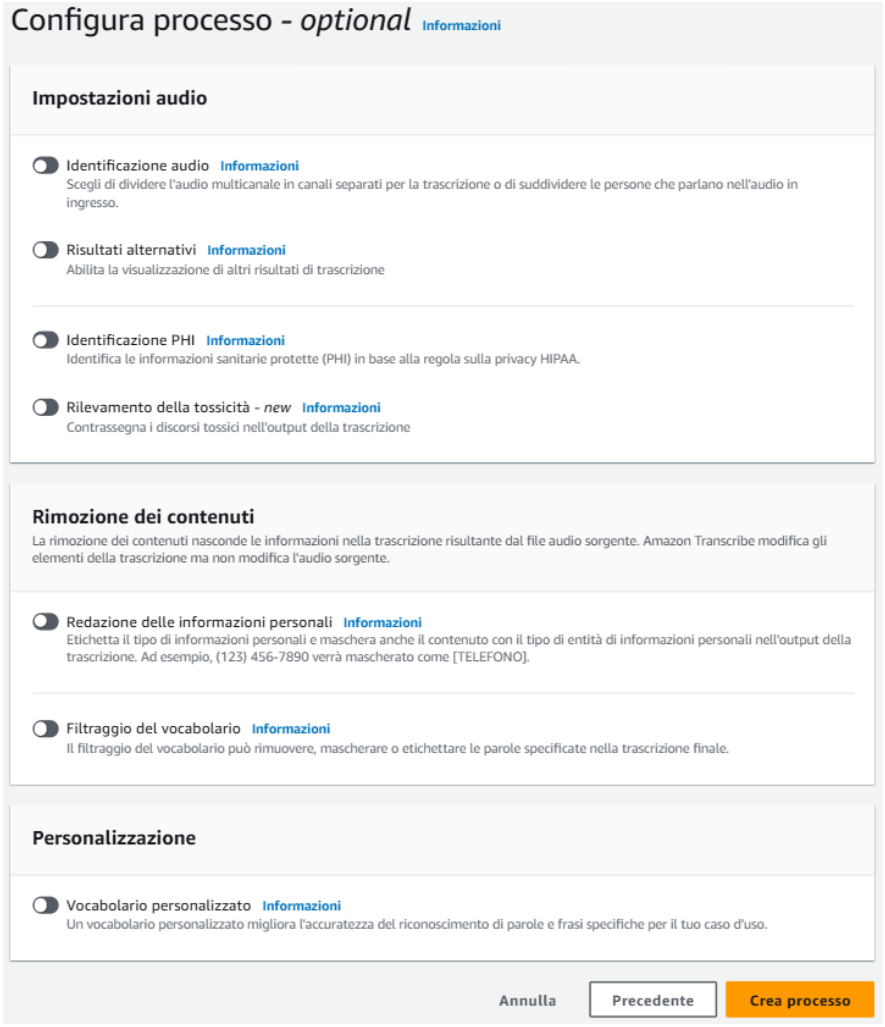


Figura 4.6: Schermata relativa alla configurazione di processo e alle funzionalità opzionabili

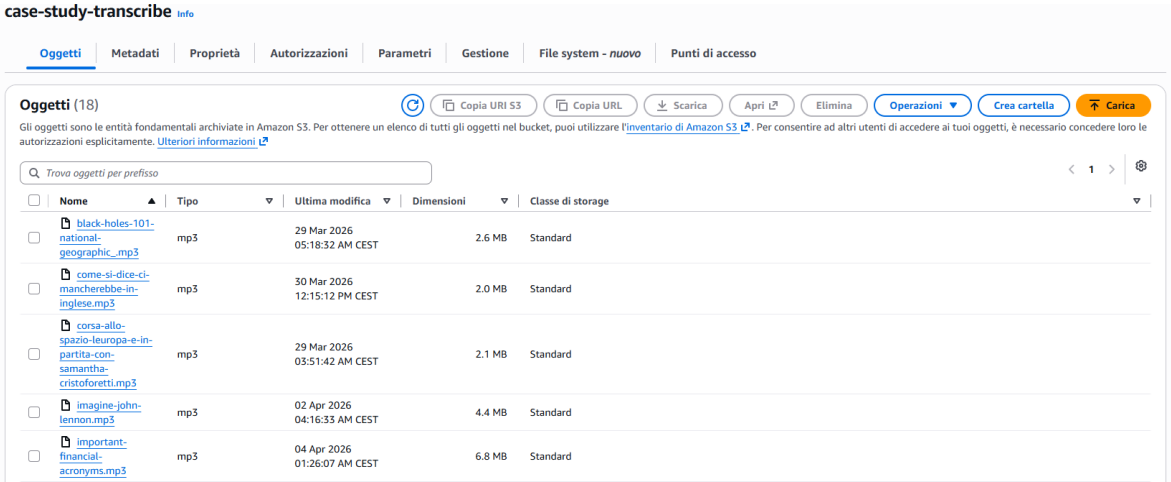


Figura 4.7: Schermata di gestione degli oggetti contenuti in un bucket Amazon S3

a condizioni caratterizzate da una maggiore complessità. Oltre ai risultati visualizzati nella sezione *Anteprima della trascrizione*, è stato esaminato il file JSON prodotto dal processo di trascrizione. I principali parametri restituiti sono descritti nella Sezione 4.2.1, mentre ulteriori

specifiche relative alle singole funzionalità vengono fornite, ove necessario, nella descrizione dei relativi casi di studio. Poichè le tracce audio sono state analizzate mediante la console di Amazon Transcribe, i valori dei parametri del servizio sono stati mantenuti invariati rispetto alla configurazione predefinita. Questa modalità di analisi consente di valutare il servizio in condizioni d'uso analoghe a quelle adottate nei casi di studio relativi ad Amazon Rekognition, presentati nella Sezione 3.4. Tutte le tracce audio sono state elaborate utilizzando la regione predefinita del servizio, corrispondente a *Stati Uniti (Virginia settentrionale)- us-east-1*.

4.3.1 Criteri di analisi e organizzazione dei risultati

Per i casi di studio descritti nelle sezioni successive, una trascrizione è stata considerata corretta quando il contenuto della traccia audio è stato riportato coerentemente nel testo prodotto dal servizio, senza evidenziare comportamenti anomali. In particolare, sono stati considerati la corretta individuazione dei termini pronunciati, la loro collocazione temporale rispetto alla traccia audio e l'inserimento coerente della punteggiatura. Inoltre, qualora fosse stata specificata una o più lingue, è stato valutato il corretto riconoscimento delle lingue presenti nella registrazione. Per ciascun caso di studio vengono illustrate le motivazioni alla base della selezione della traccia audio, l'ipotesi formulata sul comportamento atteso del sistema, i risultati ottenuti e le osservazioni emerse dalla loro analisi. In particolare, delle trascrizioni prodotte dal servizio sono state riportate le parti ritenute maggiormente significative ai fini dell'analisi. Sono, inoltre, stati evidenziati gli elementi maggiormente rilevanti emersi dall'analisi, distinguendo tra errori di trascrizione e risultati particolarmente positivi, al fine di facilitarne l'individuazione e il confronto con il contenuto della traccia audio. A tale scopo sono state adottate le seguenti convenzioni:

- in *rosso*, sono evidenziati gli errori di trascrizione, ad esempio, le sostituzioni;
- in *giallo*, sono evidenziati gli elementi di formattazione (inclusa la punteggiatura) introdotti dal servizio che riducono la leggibilità della trascrizione, nonché le scelte di formattazione che si discostano dal comportamento atteso;
- in *verde*, invece, gli elementi della trascrizione che hanno mostrato risultati particolarmente positivi.

In alcuni casi, la stessa traccia è stata sottoposta a più analisi per osservare eventuali variazioni nei risultati. Quando le impostazioni sono state mantenute invariate, gli output ottenuti sono risultati coincidenti; invece, quando sono state modificate una o più opzioni, è stato effettuato un confronto tra gli output prodotti, evidenziando le differenze riscontrate. Infine, per consentire, ove necessario, l'ascolto delle tracce audio utilizzate come input nei case study, è stato inserito un QR code che permette di accedere alla relativa registrazione.

4.3.2 Trascrizione in condizioni standard

In questa sezione vengono presentati alcuni casi di studio caratterizzati da condizioni di parlato favorevoli al riconoscimento automatico. In particolare, vengono proposte tracce in lingua inglese e italiana, ciascuna contenuta in una singola registrazione, con l'obiettivo principale di verificare la capacità di trascrivere correttamente più idiomi.

Trascrizione di un monologo in lingua inglese

La trattazione inizia con una traccia audio in lingua inglese, con pronuncia americana (Figura 4.8), della durata di circa 30 secondi. La registrazione contiene un monologo relativo

al machine learning, con alcuni esempi di applicazione di tale tecnologia nelle manzioni quotidiane svolte tramite il computer. La traccia presenta un unico parlante, una voce chiaramente identificabile e l'assenza di rumore di fondo. Considerate tali condizioni, si ipotizza che il servizio sia in grado di riconoscere correttamente il parlato e di produrre una trascrizione coerente con il contenuto della registrazione.



Figura 4.8: QR code per l'accesso alla traccia audio in lingua inglese (fonte: `https://docs.aws.amazon.com/`)

La Trascrizione 4.1 riporta l'output prodotto dal servizio, la cui analisi evidenzia il corretto rilevamento di tutte le parole pronunciate.

> Machine learning is employed in a range of computing tasks where designing and programming explicit algorithms with good performance is difficult or infeasible. Example applications include email filtering, detection of network intruders, and computer vision. Machine learning is closely related to computational statistics, which also focuses on predictions making through the use of computer. It has strong ties to mathematical optimization, which delivers methods, theory, and application domains to the field.

Trascrizione 4.1: Trascrizione della traccia audio in lingua inglese

Anche la punteggiatura introdotta risulta coerente con le pause contenute nel monologo. Pertanto, il sistema ha prodotto un risultato conforme alle aspettative.

Trascrizione di un dialogo in lingua inglese

Il secondo caso di studio riguarda una conversazione in lingua inglese, con pronuncia britannica, caratterizzata dalla presenza di due parlanti (Figura 4.9). Anche in questo caso la lingua è stata specificata esplicitamente, mentre le funzionalità di identificazione automatica della lingua vengono analizzate nei casi di studio successivi. L'audio contiene una conversazione relativa alla visita di nuovi luoghi in Gran Bretagna. Le voci dei due interlocutori sono chiaramente distinguibili e non sono presenti rumori di fondo. Per questo caso di studio è stata abilitata la funzionalità *Identificazione audio* con l'opzione *Suddivisione di chi parla*, al fine di comprovare la capacità del servizio di distinguere i due parlanti presenti nella registrazione. In queste condizioni, si ipotizza che il servizio sia in grado di identificare correttamente le parole contenute nel dialogo e di attribuire ciascuna porzione di testo al relativo interlocutore, poiché i parlanti si alternano senza sovrapporsi.

Dalla trascrizione emerge che le parole pronunciate sono state riportate correttamente e che la punteggiatura introdotta risulta coerente con le pause presenti nella conversazione, analogamente a quanto riscontrato nella prova precedente. Selezionando le singole parole nella sezione *Testo* della console si osserva il relativo grado di confidenza, risultato pari al 100% per quasi tutti i vocaboli riconosciuti. Come mostrato nella Trascrizione 4.2, essa risulta



Figura 4.9: QR code per l'accesso alla traccia audio in lingua inglese caratterizzata dalla presenza di due parlanti (fonte: `https://www.bbc.co.uk/`)

correttamente suddivisa tra gli oratori, con le porzioni di testo attribuite ai rispettivi interlocutori. Anche in questo caso, trattandosi di una traccia audio priva di particolari elementi di complessità, il servizio ha prodotto risultati coerenti con le aspettative, confermandone l'affidabilità nella trascrizione.

Persona che parla 0:
Let's start with the UK.
Persona che parla 1:
OK. If I visit somewhere in the UK I would usually want to go to the countryside. That's because I live in a city. I live in London, but I'm from the countryside. So sometimes I crave going out into a more rural area where I can see more green space. Um, there's more wildlife, um, It's almost like you need to escape the, the busy hustle and bustle of the city.
Persona che parla 0:
Yeah, you know, I, I completely agree with you. Cities in the UK are great. There are some great cities, but I have visited quite a lot of them. And after a while, the cities are all similar. Whereas the countryside, you can do different activities. It's really nice to see new places. I am A few years ago, I went to Kent and I had never been to Kent before because I'm from the north of England. And Kent is in the south and it was amazing. They had different kinds of buildings and like architecture on the farms. We don't have those same kind of roofs in the farms in the in the north.

Trascrizione 4.2: Trascrizione della traccia audio in lingua inglese caratterizzata dalla presenza di due parlanti

Trascrizione di un monologo in lingua italiana

L'ultima prova svolta relativa a condizioni di parlato favorevoli riguarda una traccia audio in lingua italiana, estratta da un'intervista a Samantha Cristoforetti per la rivista *Focus* e incentrata sul tema dell'esplorazione spaziale (Figura 4.10).



Figura 4.10: QR code per l'accesso alla traccia audio in lingua italiana (fonte: `https://www.focus.it/`)

La registrazione è stata selezionata per verificare la capacità di Amazon Transcribe di elaborare correttamente una lingua diversa dall'inglese. La traccia contiene un unico parlante e presenta lievi rumori di sottofondo, che si ipotizza non compromettano significativamente la corretta identificazione dei vocaboli. Infatti, dai risultati (Trascrizione 4.3) è possibile osservare che il servizio è stato in grado di riconoscere correttamente le parole presenti e di inserire la punteggiatura in modo appropriato. Inoltre, è stato osservato che i segnali di esitazione e i riempitivi, come ''ehm'', ''mah'' e ''uhm'', nonché le ripetizioni di parole presenti nella registrazione, vengono mantenuti nella trascrizione. Questo risultato suggerisce l'assenza di una fase di post-elaborazione specificamente finalizzata alla rimozione di tali elementi, che vengono, pertanto, conservati nel testo trascritto.

essere pronti, la ISS, la stazione spaziale [...] e astronauti a bordo. um Si avvia insomma alla sua alla sua naturale fine di di vita, [...] 2030 e del 2031 verrà fatta rientrare in sicurezza naturalmente su- sulla terra, [...] andrà distrutta. um E la la visione che venga sostituita nell'orbita bassa [...] che potranno essere delle, o delle aziende puramente americane, ma [...], delle no- con delle aziende,

Trascrizione 4.3: Trascrizione della traccia audio in lingua italiana

4.3.3 Gestione di contenuti multilingua

I casi di studio di questa sezione sono finalizzati a valutare il comportamento del servizio in presenza di più lingue contenute nella stessa traccia audio. Per ciascuna situazione vengono confrontati i risultati ottenuti specificando esplicitamente le lingue presenti nell'audio con quelli ottenuti lasciando al servizio il compito di identificarle automaticamente.

Trascrizione in presenza di alternanza tra lingua italiana e inglese

In questo caso di studio sono state analizzate due tracce audio caratterizzate dall'alternanza tra italiano e inglese all'interno dello stesso discorso. L'obiettivo è verificare la capacità del servizio di rilevare correttamente il contenuto in entrambe le lingue. Nella prima prova, le lingue specificate sono l'inglese, con pronuncia britannica, e l'italiano, mentre nella seconda il rilevamento automatico è stato affidato al servizio (Figure 4.11a e 4.11b). Il confronto tra i due risultati consente di valutare se la disponibilità preventiva di tali informazioni favorisca il riconoscimento del parlato.

(a) (fonte: `https://www.youtube.com/watch?v=dQzM0fljQ7E`)

(b) (fonte:`https://www.youtube.com/watch?v=SpgXRnk7xd4`)

Figura 4.11: QR code per l'accesso alle tracce audio caratterizzate da alternanza di lingua inglese e italiana

Prima di procedere con l'analisi dei risultati, è possibile formulare alcune ipotesi relative al comportamento atteso per il servizio nei due casi considerati. Nel primo caso, ci si attende che il servizio sia in grado di trascrivere correttamente sia le parti in italiano sia quelle in inglese, sebbene possano emergere alcune difficoltà nelle porzioni in italiano a causa della pronuncia del parlante. Nel secondo caso, invece, si può ipotizzare che il modello riconosca entrambe le lingue, ma che la trascrizione presenti un numero maggiore di errori.

Risultati con indicazione manuale delle lingue Dalla Trascrizione 4.4 si evince che il servizio non è stato in grado di rilevare correttamente tutti i vocaboli presenti nelle due lingue, producendo diversi errori.

> lo sapevo Però non complicata la vita perché è anche è uguale in italiano. Per esempio, Can you? The phone or Paulette, can you answer? The door. Qui non stiamo dando una soluzione, informazione, ma qui è sempre rispondere al telefono oppure alla porta. Va bene. Però voglio aggiungere qui per rispondere all'apporto o al telefono, possiamo anche usare, Modira, can you get the phone or can you get the Significa anche rispondere. Io so cosa state pensando, ma che significa prendere con. Se io dico can you get the phone, nel senso mi puoi portare il telefono. Anche però se il telefono sta squillando o se c'è qualcuno che sta usando alla porta proprio adesso, se si dice can you get the phone, è ovvio dal contesto che stiamo usando come rispondere.

Trascrizione 4.4: Trascrizione di una traccia audio caratterizzata dall'alternanza fra italiano e inglese

Tali errori sono riconducibili principalmente alla pronuncia, alla rapida alternanza tra italiano e inglese e alla presenza di pause nella traccia audio. Alcune parole sono state erroneamente riconosciute o sostituite (*complicate* → *complicata*, *oppure* → *or*, *alla porta* → *all'apporto*, *bussando* → *usando*). Altri termini, come *answer e door*, sono state completamente omessi. In corrispondenza di alcuni termini, il basso grado di confidenza, come per *Paulette*, evidenzia l'incertezza della trascrizione. Sono, inoltre, presenti problemi nella gestione della punteggiatura e nell'interruzione delle frasi, oltre a diversi errori nelle ultime battute del discorso, probabilmente causati dalla velocità con cui vengono alternate le due lingue. Nel complesso, le principali incongruenze si concentrano nei punti di passaggio da una lingua all'altra.

Risultati con identificazione automatica delle lingue Nel secondo caso, il servizio ha identificato come lingue presenti nella registrazione l'inglese, nella variante *Inglese, WL (en-WL)*, e l'italiano. Dall'analisi della Trascrizione 4.5 si osserva che alcune parti di discorso non sono state rilevate, mentre altre sono state trascritte in modo errato.

Pertanto, anche in questa prova sono stati riscontrati diversi errori, riconducibili principalmente alla pronuncia del parlante e al rapido alternarsi delle lingue italiana e inglese. Tra gli esempi più significativi si osservano la sostituzione di *viene* con *vino*, di *oppure* con *or a* e *Oh Porter*, di *però questo un po' più* con *peroque* e di *quindi in* con *con i in*. Sono, inoltre, presenti imprecisioni nella punteggiatura, spesso non coerente con le pause e con la struttura delle frasi. Confrontando con le informazioni ricavate dal caso precedente, si osserva che gli errori si concentrano in larga parte nelle medesime porzioni della registrazione. Di conseguenza, fornire l'elenco delle lingue presenti non ha prodotto un miglioramento significativo della trascrizione, che risulta complessivamente simile a quella ottenuta mediante rilevamento automatico.

Okay, il primo contesto in cui ci mancherebbe vino usato in italiano è una risposta ad un ringraziamento. Per esempio, grazie per avermi accompagnata a casa. Quindi in inglese, thanks. For taking me home, or a, thanks for accompanying me home, peroque. Formale. Quindi in italiano si risponde, ci mancherebbe. In inglese abbiamo tre opzioni diverse. Noi possiamo dire, Don't mention it. Any time. Oh Porter, no. Problem. E anche come italiano noi possiamo anche usare più di un'opzione. Per esempio, sei stato davvero gentile con i in inglese. You've been so kind. Quindi in italiano si rispondono figurati, ci mancherebbe. Quindi un paio di opzioni.

Trascrizione 4.5: Trascrizione di una traccia audio caratterizzata dall'alternanza fra inglese e italiano con rilevamento automatico della lingua

Trascrizione di audio multilingue

Successivamente, sono state sottoposte ad analisi due tracce audio, contenenti più di due idiomi, al fine di valutare la capacità del servizio di gestire registrazioni caratterizzate dalla presenza di diverse lingue. La prima traccia audio (Figura 4.12a) contiene olandese, inglese, francese, italiano e spagnolo. La seconda (Figura 4.12b) contiene spagnolo, portoghese, tedesco e una breve parte in italiano. In entrambi i casi, la presenza di numerose lingue, i relativi cambiamenti nel corso della registrazione e la pronuncia adottata dal parlante costituiscono condizioni potenzialmente critiche per il riconoscimento dei vocaboli.

(a) Olandese, inglese, francese, italiano e spagnolo

(b) Spagnolo, portoghese, tedesco e italiano

Figura 4.12: QR code per l'accesso alle tracce audio caratterizzate da multilinguismo (fonte: `https://it.babbel.com/`)

Pertanto, è plausibile ipotizzare che il servizio sia in grado di identificare correttamente le lingue presenti nel discorso, sebbene possano verificarsi alcuni errori nel riconoscimento di specifici vocaboli.

Risultati con indicazione manuale delle lingue Dalla Trascrizione 4.6 è possibile osservare che il servizio ha riconosciuto correttamente le parole pronunciate nelle diverse lingue, senza evidenziare errori significativi. Anche la punteggiatura risulta coerente con le pause presenti nella registrazione e il grado di confidenza associato ai singoli vocaboli risulta generalmente elevato.

La buona accuratezza del risultato può essere ricondotta, almeno in parte, alla netta separazione dei diversi segmenti linguistici all'interno della registrazione. I passaggi in ciascuna lingua non presentano, infatti, significative sovrapposizioni con vocaboli appartenenti alle altre lingue.

> Ik ga jullie dus nu vertellen hoe ik deze zeven talen geleerd heb. Of ja, waarom. Ik ben natuurlijk met Nederlands begonnen. Want ja, [...] in in Holland it's kind of a normal thing you. There is no way you can escape English. [...] I speak very good English. Sorry, sorry can I ask a question mais c'était avec le français à 11 ans que j'ai compris mais en fait ça [...] j'ai un peu parlé français bon, c'est un peu [...] me dis putain, ça te change la vie quand tu parles une autre langue. E e allora ho pensato che questa cosa, infatti era, cioè la cosa che volevo scappare [...] ho fatto tre mesi lì e ho pensato, mannaggia cioè è un un'altra cioè. U una otra parte de de mí misma. No, que [...] total y y no. Y yo pensé pero para mí, un español no es la misma cosa que en italiano o un francés.

Trascrizione 4.6: Trascrizione di una traccia audio contenente più lingue specificate esplicitamente

Risultati con identificazione automatica delle lingue Nella sezione *Dettagli processo* sono state identificate le seguenti lingue: portoghese, nella variante Portoghese, PT (pt-PT), tedesco e spagnolo, nella variante Spagnolo, Stati Uniti (es-US). Pertanto, il servizio ha riconosciuto correttamente le principali lingue presenti nella registrazione. Tuttavia, per lo spagnolo è stata selezionata la variante statunitense, anziché quella associata alla pronuncia spagnola. L'italiano non è stato, invece, identificato, probabilmente anche a causa della sua presenza limitata a una breve frase inserita all'interno della porzione in spagnolo. Tale frase, inoltre, è stata riportata in modo errato, probabilmente in conseguenza del mancato riconoscimento della lingua e del rapido passaggio tra lo spagnolo e l'italiano.

> La había un poco estudiada en casa, pero a ver, no era Roma, ¿eh? É Thiago l'espresso era completamente outro no para mim e E então, para mim foi e era uma coisa, eu pensei se então funcionou mais ou [...] , não tenho o taco, é isso e e nada assim. Sempre menos medo assim de falar. Und ich glaube, so habe ich dann mit Deutsch dann richtig angefangen, weil dann bin ich nach Deutschland [...] und. Ja, das. Das kann ich ja nur mitgeben. Gracias por ver mi vídeo. Si tienes una pregunta, un comentario, lo puedes dejar aquí abajo.

Trascrizione 4.7: Trascrizione di una traccia audio contente più lingue rilevate automaticamente

A differenza del caso precedente, la trascrizione presenta alcune imprecisioni. L'output risulta complessivamente meno accurato rispetto alla traccia in cui le lingue sono state esplicitate. Tale differenza suggerisce che, nelle condizioni analizzate, la disponibilità delle informazioni relative alle lingue presenti nella registrazione possa aver favorito il riconoscimento dei vocaboli.

4.3.4 Analisi di condizioni acustiche particolari

All'interno di questa sezione vengono riportati alcuni esempi in cui la presenza di particolari condizioni nella traccia audio può rendere più complessa per il servizio la corretta identificazione dei vocaboli.

Trascrizione di audio con più parlanti

In questa sottosezione si intende valutare la capacità del servizio di rilevare più di due parlanti all'interno di un'unica traccia audio in lingua inglese (Figura 4.13). Inoltre, si vuole verificare se la presenza di più interlocutori possa compromettere la corretta identificazione

dei vocaboli. A tale scopo, è stata selezionata una traccia contenente un dibattito politico trasmesso dall'emittente statunitense CNN.



Figura 4.13: QR code per l'accesso alla traccia audio contenente più parlanti (fonte: `https://www.youtube.com/watch?v=r8_eI-stGQc`)

Tra i parametri di trascrizione è stata, quindi, selezionata l'opzione *Identificazione audio*, impostando un numero massimo di 5 interlocutori. Pertanto, ci si attende che il servizio sia in grado di identificare correttamente i vocaboli pronunciati durante il dibattito, sebbene possano verificarsi alcuni errori nelle parti caratterizzate da sovrapposizioni tra le voci dei diversi interlocutori o da un volume di voce ridotto.

Risultati con rilevazione automatica della lingua La lingua identificata è l'inglese, nello specifico Inglese, Stati Uniti (en-US), con un grado di confidenza pari a 0,70. Il servizio non ha, quindi, riscontrato particolari difficoltà nell'identificazione della lingua. Dalla Trascrizione 4.8, che non presenta la separazione tra i diversi interlocutori, si osserva che il modello ha rilevato correttamente la maggior parte dei vocaboli pronunciati durante il dibattito. In presenza di sovrapposizioni tra più parlanti, il servizio riporta principalmente la voce che riesce a distinguere maggiormente, con conseguente omissione di alcune parti della conversazione. Anche la punteggiatura risulta complessivamente coerente, sia in corrispondenza delle pause dei singoli interlocutori sia nei momenti di passaggio da un parlante all'altro. Alcune delle imprecisioni osservate possono essere ricondotte proprio alla gestione delle sovrapposizioni tra le voci. In particolare, dopo *what I'm saying is that we have* viene inserito un punto di sospensione e viene omessa la parola *become*. Un ulteriore esempio si osserva nel momento in cui un interlocutore interrompe il parlante corrente. In questo caso, il modello privilegia la voce dell'interlocutore che interviene, omettendo parte del parlato precedente, sebbene questo risulti comprensibile nella traccia audio. Un esempio è rappresentato dalla frase *it's not just about that, but that is part of it*.

That's just say this one's unconventional, Joe. You specialize in this stuff, buddy. You, you run, you run races. [...] Never. This is not, Joe. I'm not, I'm not listen, but you, but. For you to say [...] He, he's doing it for a career too. Explain what you mean when you say Joe. Well, Joe, Joe's a political consultant. [...] Of course, tons, right? So how is that? So what does that have to do with the division talking about the United States who is the president. [...] to be healed and brought together. [...] I think it's horrible for Republicans to do it or Democrats. I'm not saying that Republicans have a monopoly on racism and bigotry. What I'm saying is that we have. numb to some of the very real challenges [...] he didn't invent racism or discrimination. I'll be the first one to say that.[...] All I'm asking is that we don't make it only that the United States of America is it's [...] enslave Africans? Did this country take away the land of the Native Americans? Did this country have Jim Crow and black codes? That's not in my imagination. And, and still there are articles [...]

Trascrizione 4.8: Trascrizione di una traccia audio contenente numerosi parlanti

Osservando la Trascrizione 4.9 con la suddivisione del discorso in base ai parlanti emergono numerosi errori. Infatti, già nelle prime battute del dibattito diverse frasi non vengono attribuite correttamente all'interlocutore corrispondente. Dunque, da questo punto di vista, la trascrizione appare frammentata e presenta numerose incoerenze rispetto all'effettiva successione degli interventi. In particolare, alcune porzioni conclusive delle frasi vengono associate all'interlocutore successivo. Complessivamente, il servizio individua 4 interlocutori. In realtà, nel dibattito è presente un quinto partecipante, i cui interventi non vengono correttamente rilevati, ma alcuni dei vocaboli da lui pronunciati vengono erroneamente attribuiti agli altri parlanti. A differenza di quanto osservato nella Sezione 4.3.2, in cui i due parlanti si alternano in modo distinto e pronunciano porzioni di discorso relativamente lunghe, senza determinare particolari difficoltà per il servizio, in questa prova la presenza di un numero maggiore di interlocutori, caratterizzati da turni di parola più brevi, rapidi e parzialmente sovrapposti, ha reso più complessa la corretta trascrizione della conversazione.

Risultati con specificazione manuale della lingua Per completare il caso di studio corrente, è stata eseguita un'ulteriore elaborazione utilizzando la stessa traccia audio, questa volta specificando manualmente la lingua, ovvero Inglese, Stati Uniti (en-US). Inoltre, considerato il contenuto della traccia, è stata utilizzata un'ulteriore funzionalità messa a disposizione dal servizio, *Rilevamento di linguaggio tossico*. In tal caso, nessuna delle categorie considerate ha superato il valore di 0,5, come osservabile nella Figura 4.14. Tale risultato è ulteriormente evidenziato dai colori utilizzati dal servizio per segnalare le parole che potrebbero presentare caratteristiche riconducibili a un linguaggio potenzialmente tossico. Pertanto, il sistema ha ritenuto che gli elementi presenti nel discorso non fossero sufficientemente tossici da richiedere l'assegnazione di un'etichetta corrispondente a un grado di tossicità più elevato, in linea con la valutazione attesa. Nel complesso, la funzionalità si è, quindi, rivelata utile per l'impiego in piattaforme di social media o di gioco online, consentendo di identificare contenuti potenzialmente tossici.

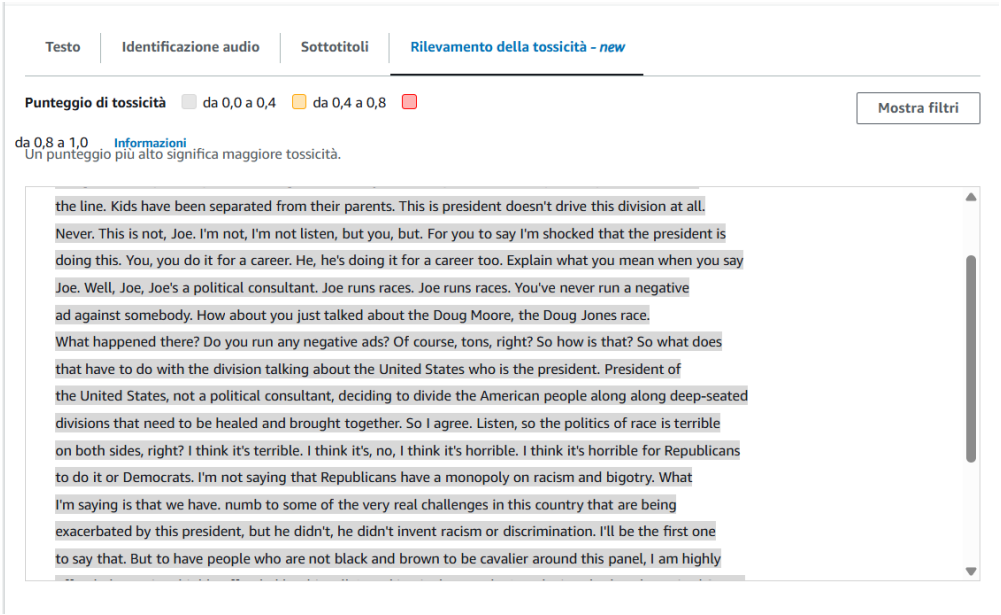


Figura 4.14: Output restituito dal servizio per l'individuazione di linguaggio potenzialmente tossico nella traccia audio

Infine, la specificazione manuale della lingua non ha prodotto un miglioramento del risultato, poiché sono stati riscontrati gli stessi errori già osservati nella prova precedente.

Persona che parla 0:
That's just say this one's unconventional, Joe. You specialize in this stuff,
Persona che parla 1:
buddy. You, you run, you run
Persona che parla 0:
races. That's what about, man. You're right there with you did it.
Persona che parla 1:
You did it,
Persona che parla 2:
Joe. That's what you do for a living messed up. You spin for a living. This is really messed up. This is not a spin. People's lives are on the line. Kids have been separated from their parents. This is
Persona che parla 0:
president doesn't drive this division at all. Never. This is not, Joe.
Persona che parla 1:
I'm not, I'm not listen, but you, but. For you to say I'm shocked that the president is doing this. You, you do it for a
Persona che parla 2:
career. He, he's doing it for a career too.
Persona che parla 0:
Explain what you mean when you say Joe.
Persona che parla 1:
Well, Joe, Joe's a political consultant. Joe runs races. Joe runs races. [...] How about you just talked about the Doug Moore, the Doug Jones race. What happened there? Do you run any negative ads? Of course, tons, right? So how is that? So
Persona che parla 0:
what does that have to do with the division talking about the [...], not a political consultant, deciding to divide the American people along along deep-seated divisions that need to be healed and brought together.
Persona che parla 1:
So I agree. Listen, so the politics of race is terrible on both sides, right? I think it's terrible. I think it's, no, I think it's horrible. I think it's horrible for Republicans to do it
Persona che parla 0:
or
Persona che parla 2:
Democrats. I'm not saying that Republicans have a monopoly on racism and bigotry. What I'm [...] to be cavalier around this panel, I am highly offended. I am just highly offended by this.
Persona che parla 3:
All I'm asking is that we don't make it only that the United States of America is it's not just about that, but that is part of it.
Persona che parla 2:
Was this country, did this country enslave Africans? [...] and we all have to care about it.

Trascrizione 4.9: Trascrizione con diarizzazione di una traccia audio contenente numerosi parlanti

Influenza del rumore di fondo sulla trascrizione

In questo esempio, l'analisi è stata condotta utilizzando una traccia audio caratterizzata dalla presenza di rumore di sottofondo, con l'obiettivo di verificare la capacità del servizio di estrarre correttamente il parlato. La traccia selezionata è tratta da una serie di interviste realizzate nel 1979 nella zona di Wall Street, a New York (Figura 4.15), per raccogliere le opinioni delle persone riguardo a economia e sviluppo tecnologico. La registrazione,

presumibilmente realizzata mediante supporti analogici, presenta una qualità audio non particolarmente elevata. Inoltre, l'ambiente circostante contribuisce alla presenza di un rumore di sottofondo che potrebbe interferire con la corretta rilevazione dei dialoghi.



Figura 4.15: QR code per l'accesso alla traccia audio contenente rumore di sottofondo (fonte: `https://www.youtube.com/watch?v=jCjuXt7CiIY`)

Nelle impostazioni della trascrizione è stata selezionata l'opzione per il rilevamento automatico della lingua, indicando tra le lingue disponibili l'inglese con pronuncia americana e l'inglese con pronuncia britannica. In aggiunta, è stata attivata la funzionalità per la suddivisione degli interlocutori, indicando un numero massimo di 3 parlanti. Dalla sezione *Dettagli di processo* è emerso che la lingua rilevata è l'inglese, nello specifico Inglese, Stati Uniti (en-US), con un grado di confidenza pari a 0,99. Dalla Trascrizione 4.10 è possibile osservare che il servizio è stato in grado di rilevare correttamente la quasi totalità dei vocaboli presenti nelle conversazioni. Quindi, il modello ha dimostrato di essere in grado di separare il rumore di sottofondo dalle voci degli interlocutori, anche nei casi in cui queste ultime non risultano chiaramente distinguibili. Tale comportamento potrebbe essere riconducibile alla pronuncia generalmente più scandita delle parole da parte dei parlanti, che ne facilita il riconoscimento da parte del servizio. Anche la punteggiatura è stata inserita in modo coerente con i momenti di silenzio presenti nell'audio. Tuttavia, sono presenti alcune imprecisioni. Ad esempio, la domanda dell'intervistatore *do you have a lot of paperwork?* viene trascritta come *A lot of people*. Tale errore potrebbe essere riconducibile alla distanza del parlante dal dispositivo utilizzato per la registrazione e alla presenza di rumori di fondo. Invece, la suddivisione dei parlanti risulta significativamente meno accurata, analogamente a quanto osservato nel caso precedente. Infatti, le prime domande e risposte vengono attribuite alla medesima persona. In particolare, la voce dell'intervistatore non viene identificata come un interlocutore distinto; le sue frasi risultano frammentate e distribuite tra gli altri parlanti. Successivamente, è stata eseguita una seconda prova, aumentando a 7 il numero di parlanti. Tuttavia, tale modifica non ha prodotto alcun miglioramento nella trascrizione, che è risultata identica a quella ottenuta nella prova precedente.

> What do you do? What's your job? I'm a fixed income salesman. [...] The first thing that comes to your head, remember you're gonna repeat exactly what, uh, I asked you, OK? Um, OK, computer. Programming. I repeat the word computer programming computer tape. Computer tape You're uh Bureaucracy Washington. Uh, bureaucracy, red tape. You guys move a little closer together. We're not married. No, you guys are doing great, really. [...] , regulations that they impose upon us. A lot of people Do I have a lot of paperwork? Yes. Do I have a lot of paperwork? Yes, I do. What do you think of the corporations? They're devious. What do you think about the information society. What do I think about the information site? [...] It's good and bad in both, I think.

Trascrizione 4.10: Trascrizione di una traccia audio caratterizzata da rumori di sottofondo

Influenza della velocità di riproduzione

Per questa prova è stata selezionata una traccia audio in lingua inglese caratterizzata da un parlato comprensibile, con la presenza di effetti sonori e musica di sottofondo, che potrebbero incidere sulla corretta identificazione dei vocaboli. La registrazione è estratta da un video di National Geographic riguardo ai buchi neri ed accessibile mediante il QR code riportato nella Figura 4.16. Nelle impostazioni della richiesta è stata esplicitata la lingua inglese, nello specifico Inglese, Stati Uniti (en-US).

Figura 4.16: QR code per l'accesso alla traccia audio in lingua inglese (fonte: `https://education.nationalgeographic.org`)

Riproduzione a velocità normale Poichè, dalla Trascrizione 4.11 non emergono errori di identificazione, la traccia è stata utilizzata come riferimento per valutare l'influenza della velocità del parlato sulla qualità dell'output prodotto dal servizio. Per tale motivo, sono stati creati altri due processi di trascrizione. Nel primo, la velocità dell'audio è stata ridotta a 0,5x; nel secondo, essa è stata aumentata a 1,5x. L'obiettivo è osservare eventuali differenze nei risultati ottenuti al variare di tale parametro.

> Stellar black holes form when massive stars die and collapse. They're roughly 10 to 20 times the mass [...] holes in the Milky Way alone. [...] Sagittarius A, the supermassive black hole at the center of the Milky Way, has a mass of roughly 4 million suns and has a diameter about the distance between the Earth and our sun. [...] This includes accretion [...] black holes.

Trascrizione 4.11: Trascrizione a velocità di riproduzione normale

Riproduzione a velocità ridotta La riduzione della velocità di riproduzione non ha evidenziato differenze rispetto al caso di riproduzione a velocità normale.

Riproduzione a velocità aumentata Anche l'aumento della velocità di riproduzione non ha prodotto variazioni nei risultati; la trascrizione è rimasta invariata e non sono stati commessi errori nell'identificazione dei vocaboli. Inoltre, la punteggiatura è stata inserita in modo coerente e identico in tutte e tre le prove. Infine, dalle trascrizioni prodotte dal servizio si evince che alcuni nomi propri vengono correttamente riportati con la lettera iniziale maiuscola, come nel caso di *Milky Way*. Anche la presenza di un linguaggio di carattere scientifico non ha costituito un ostacolo alla corretta trascrizione della traccia.

4.3.5 Influenza degli accenti sulla trascrizione

Per questa sezione sono state selezionate alcune tracce audio in lingua inglese caratterizzate dalla presenza di accenti differenti per comprendere se il servizio sia in grado di identificare correttamente i vocaboli anche in presenza di pronunce differenti.

Nel primo esempio viene analizzata la capacità del servizio di trascrivere una traccia in lingua inglese caratterizzata da accento indiano (Figura 4.17a). L'audio è tratto da una serie di interviste condotte con persone di origine indiana e incentrate proprio sulla tematica della diffusione dell'inglese in India. La particolare pronuncia di alcuni vocaboli potrebbe rendere più complessa la corretta identificazione da parte del modello; inoltre, la presenza di rumori di sottofondo potrebbe influire ulteriormente sulla qualità della trascrizione. Nelle impostazioni relative alla lingua non è stata selezionata una specifica variante, lasciando al servizio il compito di identificarla automaticamente. Considerata la disponibilità, tra le lingue selezionabili, dell'inglese con accento indiano, ci si attende che il servizio sia comunque in grado di produrre una trascrizione accurata.



(a) Indiano (fonte: `https://www.youtube.com/watch?v=dJgoTcyrFZ4`)



(b) Scozzese (fonte: `http://youtube.com/watch?v=73uATsa8y5Y`)

Figura 4.17: QR code per l'accesso alle tracce audio caratterizzate da accento particolare

La lingua identificata automaticamente dal servizio è l'inglese, precisamente Inglese, Regno Unito (en-GB), con un grado di confidenza pari a 0,74. Pertanto, il servizio non ha identificato la specifica pronuncia indiana. Dall'analisi della Trascrizione 4.12 è possibile formulare una serie di osservazioni. Tutte le parole contenute nell'audio sono state opportunamente rilevate, senza errori significativi. Tuttavia, la punteggiatura risulta poco coerente e appare frammentata tra i diversi interventi degli interlocutori.

> So why does everybody speak English as opposed to the native Indian languages in India, I guess speaking English is a trend or maybe a It is like a show off or something. [...] So it's school's responsibility to teach me English. [...] If you speak fluent English, do people treat you better? Yeah, in India? Yes, I have seen that. Yeah, I guess so. Yeah. When you just come across someone, you have a sort of mindset towards them. As soon as they start speaking, your mindset changes. Like, OK, they can speak English. Yeah, OK, they must be like cool, I guess.

Trascrizione 4.12: Trascrizione di una traccia audio in lingua inglese con accento indiano

A integrazione del caso precedente, è stata analizzata una traccia audio in lingua inglese tratta da un'intervista a una persona scozzese, incentrata sul film *"Braveheart"* e sulla figura storica del condottiero scozzese William Wallace (Figura 4.17b). Tra le diverse varianti della lingua inglese selezionabili nelle impostazioni del processo è presente anche quella scozzese. Si ipotizza, pertanto, che la trascrizione possa risultare sufficientemente accurata. Analoga-

mente al caso precedente, non è stata specificata manualmente alcuna lingua, devolvendo al servizio il compito di effettuare automaticamente l'identificazione. Nella traccia sono, inoltre, presenti alcuni disturbi che potrebbero rendere più complessa l'elaborazione da parte del servizio. Nelle impostazioni della richiesta di trascrizione è stata selezionata la suddivisione dei parlanti, indicando un numero massimo di 2 interlocutori. Dalla sezione *Dettagli processo* è possibile osservare che il servizio ha identificato la lingua come Inglese, AB (en-AB), con un grado di confidenza pari a 0,67. Nella Trascrizione 4.13 sono presenti pochi errori nell'identificazione di vocaboli, che risultano, nel complesso, correttamente riportati. Tuttavia, sono state riscontrate alcune incongruenze nella suddivisione degli interventi tra i due interlocutori, come già emerso nei casi precedenti. Anche la punteggiatura non risulta completamente accurata; infatti, sono presenti alcuni punti di sospensione in corrispondenza di frasi che non risultano effettivamente complete.

> Aye, in the mountain, man, I was, he said, Kevin said he snowboarded off this mountain with his son. I've done it. [...] There's actually 3, but we can't see the 3rd 1. It's just directly behind the other one.[...] tended it, you know, one of the things I saw and, and, uh, [...] they joined up with. Wallace, and [...] like. William Wallace, Braveheart. What do you think of the movie? Fantastic. [...] Like, I, I, I know it. I know it probably wasn't exactly historically accurate, but it was a killer movie, wasn't it? It's one of my favourites. Really good, fantastic. Watch it at least once a year when you want a buzzy patriotism. That's what you do. You get out Wallace and you watch the whole lot. Fantastic. Like it really is.

Trascrizione 4.13: Trascrizione di una traccia audio in lingua inglese con accento scozzese

Nell'ultimo esempio analizzato (Figura 4.18) è presente una traccia audio in lingua inglese caratterizzata da un forte accento russo. L'obiettivo è verificare se il servizio sia in grado di identificare correttamente i vocaboli nonostante la particolare pronuncia del parlante.

Figura 4.18: QR code per l'accesso alla traccia audio in lingua inglese con accento russo

L'analisi con Amazon Transcribe è stata condotta attraverso due prove distinte. Nella prima è stata utilizzata l'identificazione automatica della lingua, al fine di verificare la capacità del servizio di individuare correttamente l'idioma presente nella traccia; nella seconda è stata specificata manualmente la lingua, con l'obiettivo di verificare l'eventuale miglioramento della trascrizione.

Risultati con identificazione automatica della lingua In questo caso si ipotizza che il servizio possa commettere alcuni errori nella trascrizione a causa della pronuncia del parlante, considerando, inoltre, che nell'elenco delle lingue selezionabili non è presente una specifica variante dell'inglese con accento russo. Dalla sezione *Dettagli processo* è possibile notare che il servizio non è stato in grado di rilevare correttamente la lingua. Infatti, la lingua identificata è il Russo, RU (ru-RU), con un grado di confidenza pari a 0,68. Tale risultato può essere ricondotto alla forte influenza dell'accento russo sulla pronuncia del parlante.

Nonostante la lingua identificata non sia corretta, la trascrizione è stata, comunque, prodotta in inglese e le parole effettivamente rilevate risultano, generalmente, corrette. Tuttavia, sono presenti numerose omissioni, che rendono il risultato complessivamente molto impreciso, come rilevabile nella Trascrizione 4.14.

> Night at the servant went to the door and the young woman stood outside dressed in the heavy rain. The prince happened to be walking past, but when heard the voice of the young woman in his tracks. She got lost in the rain and the light from your Would you be kind enough to shelter for the night and this must be quickly ran up and excused the Of course you are welcome to stay he said respectfully allow me to take your.

Trascrizione 4.14: Trascrizione di una traccia audio in lingua inglese con accento russo mediante rilevazione automatica della lingua

L'analisi del file JSON associato alla trascrizione mostra, inoltre, che, tra le possibili lingue rilevate, il russo presenta il punteggio maggiore, seguito dall'ucraino, dall'inglese con accento indiano, dall'inglese con accento americano e, infine, dal bielorusso. In aggiunta a ciò, tutte le parole presenti nella trascrizione sono caratterizzate da un grado di confidenza molto basso.

Risultati con indicazione manuale della lingua Come è possibile osservare dalla Trascrizione 4.15, tutte le parole presenti nel discorso sono state correttamente rilevate da parte del modello, con un grado di confidenza significativamente maggiore rispetto all'analisi precedente.

> That night during a terrible thunderstorm, there was a knock at the door. A servant [...] but when he heard the voice of the young woman, he stopped in his tracks. She explained to the servant, My driver got lost in the rain, and we saw the light from your castle. Would [...] for the night? She heard the voice of an angel, and the prince knew immediately that this must be his princess. [...] Of course you are welcome to stay, good lady, he said, bowing respectfully. Allow, allow me to take your coat.

Trascrizione 4.15: Trascrizione di una traccia audio in lingua inglese con accento russo mediante indicazione manuale della lingua

Il confronto tra le due prove evidenzia una differenza significativa rispetto ad altri casi di studio, nei quali l'esecuzione di più richieste ha prodotto risultati sostanzialmente identici. In questo caso, la selezione esplicita della lingua ha determinato un miglioramento evidente della qualità della trascrizione.

4.3.6 Gestione del linguaggio specialistico

In questa sezione viene approfondita un'ulteriore funzionalità messa a disposizione da Amazon Transcribe, ovvero la possibilità di creare un *Vocabolario personalizzato*. A tale scopo sono state selezionate alcune tracce audio caratterizzate dalla presenza di linguaggio settoriale, ad esempio, relativo al settore dell'ingegneria informatica, oppure al linguaggio giuridico e contrattuale. Per ciascun caso viene inizialmente eseguita una trascrizione senza l'utilizzo di un vocabolario personalizzato, allo scopo di comprendere la capacità del servizio di riconoscere i termini appartenenti a uno specifico settore. Successivamente, viene creato

un vocabolario personalizzato, con l'obiettivo di verificare l'eventuale miglioramento della trascrizione. La lingua utilizzata per lo svolgimento delle prove è l'inglese.

Linguaggio tecnico nel settore dei microcontrollori

In questo primo esempio è stata selezionata una traccia caratterizzata da un marcato linguaggio settoriale (Figura 4.19), relativa a un'introduzione ai microcontrollori della famiglia STM32 dell'azienda STMicroelectronics NV. Infatti, la traccia contiene numerosi vocaboli tipici del settore, tra cui sigle relative a protocolli di comunicazione e nomi di componenti specifici. Pertanto, in assenza di un vocabolario personalizzato, si ipotizza che il modello possa incontrare difficoltà nel riconoscimento di alcuni dei termini presenti, in particolare di quelli propri dell'azienda e delle relative tecnologie.



Figura 4.19: QR code per l'accesso alla traccia audio caratterizzata da linguaggio tecnico (fonte: `https://youtu.be/u7hwHh_-7_g?si=vlmbixb9FhIerf5K`)

Risultati senza vocabolario personalizzato L'analisi della trascrizione, riportata nella Trascrizione 4.16 evidenzia che il modello è stato in grado di riconoscere correttamente soltanto alcuni dei termini maggiormente utilizzati in questo settore, tra cui *USB*, *ADCs*, *MCUs*, *DMA* e *CPU*.

Infatti, emergono numerosi inesattezze nel riconoscimento di termini tecnici, sigle e nomi propri relativi ai microcontrollori e ai protocolli di comunicazione. Molti degli errori riguardano la formattazione e l'uso di lettere maiuscole e minuscole e trattini, mentre altri consistono nel riconoscimento errato di termini quali *Arduino*, *STM32s*, *UART*, *SPI*, *CAN*, *ZigBee*, *LoRa*, *ST-LINK* e *STM32CubeIDE*. Alcune espressioni tecniche sono state, inoltre, interpretate sulla base della loro pronuncia anziché essere trascritte nella forma corretta, come nel caso di *I square C*, utilizzato al posto di *I2C*. Ciò evidenzia una delle principali difficoltà riscontrate nella corretta resa grafica di acronimi, marchi e nomi tecnici. È, invece, da notare il corretto utilizzo dei numeri affiancati alle lettere nella composizione di determinate sigle. In alcuni casi, tuttavia, sono state utilizzate lettere maiuscole anche quando, apparentemente, non risultavano necessarie.

Risultati con vocabolario personalizzato Il vocabolario personalizzato realizzato per questo caso di studio è riportato nella Tabella 4.1, secondo le specifiche illustrate nella Sezione 4.2.1 per la creazione di un nuovo vocabolario.

Per verificare il corretto riconoscimento dei termini inseriti nel vocabolario, è stata utilizzata la modalità di trascrizione in streaming, selezionando il vocabolario precedentemente creato nelle impostazioni della trascrizione e impostando come lingua l'inglese, nello specifico Inglese, Stati Uniti. Dalla Trascrizione 4.17 è possibile constatare che non sono stati ottenuti miglioramenti significativi, poiché numerosi termini sono stati nuovamente identificati nello stesso modo riscontrato nella Trascrizione 4.16.

Nonostante ciò, nel secondo tentativo, alcuni termini tecnici hanno mostrato un miglioramento, come *ARM Cortex*, *SPI* e *LoRa*; tuttavia, persistono numerosi errori di riconoscimento.

Have you ever wondered why so many serious projects around the world use STM 32 microcontrollers? These chips made by ST Micro Electronics, are part of a huge family of 32-bit arm cortex and processors. [...], and most of them run on just 3.3 volts, [...] Unlike, Simpler boards like the classic Ardino, SEM 32s can run at tens or even hundreds of megahertz and pack in a huge range of peripherals such as multiple UR, SBI, and ISquareC interfaces, high resolution timers, fast ADCs, USB, C, and even DMA for moving data in the background without loading the CPU. There are many SCM 32 families, each aimed at a different purpose. At the entry level, you have the F0 or GZ0 series, perfect for cost sensitive projects. The F1 and F3 series are more capable, with the F3 line focusing on better analog performance. For high-end processing, the F4, F7, and H7 families over powerful cores and, Extra memory. If low power is your priority, the L series chips can run for years on a coin cell, and if you need wireless, the WB series includes Bluetooth and Zigby, while the WL series integrates Sujigger's radios for Lara and similar protocols. [...] such as an SD link to upload code. [...] with a separate SDin programmer. These work fine, [...] in STM 32 cubiDE. [...], which include modern SD-Link debuggers, we'll try it in, [...]. Programming SCM 32s is usually done in SCM 32 QIDE, a free tool from SD that combines code editing, compiling, flashing, and. Bugging. Inside, it is QMX, which lets you visually configure your chip's pins, clock settings, and peripherals. Once you are set up, the tool automatically generates C or C++ code using hull drivers, [...] a basic Ardio setup. But the real reason people choose STM 32 is [...] tiny local sports to a powerful high-end MCUs.

Trascrizione 4.16: Trascrizione di una traccia audio in lingua inglese contenente linguaggio tecnico in assenza di vocabolario personalizzato

Phrase	SoundsLike	IPA	DisplayAs
S.T.M.-thirty-two			STM32
S.T.-Microelectronics			STMicroelectronics
ARM			ARM
Cortex			Cortex
Arduino			Arduino
S.T.M.-thirty-two-s			STM32s
UART			UART
USART			USART
S.P.I.			SPI
I.-square-C.			I2C
CAN			CAN
F.-zero			F0
G.-zero			G0
L.-series			L-series
Zig-Bee			ZigBee
sub-giga-hertz			sub-gigahertz
Lora			LoRa
S.T.-Link			ST-LINK
S.T.M.-thirty-two-cube-ide			STM32CubeIDE
HAL			HAL
ide			IDE
Cube-M.X.			CubeMX
ports			ports

Tabella 4.1: Vocabolario personalizzato per il riconoscimento dei termini tecnici

Inoltre, l'inserimento di *CAN* nel vocabolario ha causato falsi riconoscimenti ogni volta che viene pronunciato il verbo inglese *can*, senza, peraltro, migliorare l'identificazione dell'acronimo nel contesto corretto. La sigla *H7*, precedentemente riconosciuta correttamente, è

Have you ever wondered why so many serious projects around the world use STM 32 microcontrollers? These chips made by ST Micro Electronics, are part of a huge family of 32-bit ARM Cortex and processors. [...] Unlike, Simpler boards like the classic Ardio, SEM 32s CAN run [...] multiple UR, SPI and IquareC interfaces, high resolution timers, fast ADCs, USB, C, and even [...] There are many SCM 32 families, [...] F0 or GZ0 series, [...] and HA7 families over powerful cores and, Extra memory. [...], the LS chips CAN run [...] Bluetooth and Zigby, while the WL series integrates Sujigger's radios for LoRa and [...] an SD Link to upload code. [...] separate Stilink programmer. [...] in STM 32 cubideE. Instead, [...] modern SD-link debuggers, [...] Programming SCM 32s is usually done in SCM 32 QDE, a free tool from SD that combines code editing, compiling, flashing, and. Bugging. Inside, it is QMX, which [...] using hull drivers, so you CAN start coding right away. You CAN step [...] basic Ardio setup. But the real reason people choose STM 32 is [...] local sports to a powerful high-end MCUs.

Trascrizione 4.17: Trascrizione di una traccia audio in lingua inglese contenente linguaggio tecnico in presenza di vocabolario personalizzato

stata trascritta erroneamente. Dunque, in tale contesto, la creazione del vocabolario non ha prodotto un miglioramento complessivo della trascrizione e, per alcuni termini, ha persino determinato un peggioramento della qualità del risultato.

Linguaggio nel settore finanziario

In questo caso di studio viene analizzata una traccia audio in lingua inglese relativa all'importanza dell'utilizzo di una terminologia finanziaria corretta (Figura 4.20).

Figura 4.20: QR code per l'accesso alla traccia audio caratterizzata da linguaggio finanziario (fonte: `https://youtu.be/DApfoOmMBFw?si=IOISIZouXdmdg7Fo`)

Analogamente al caso precedente, si ipotizza che, in assenza di un vocabolario personalizzato, il servizio possa incontrare difficoltà nel riconoscimento di sigle, acronimi e altri elementi caratteristici del settore finanziario.

Risultati senza vocabolario personalizzato Nelle opzioni della richiesta di trascrizione, a differenza del caso precedente, è stata selezionata la modalità di identificazione automatica della lingua, senza fornire al servizio alcuna informazione preliminare. Dai risultati è emerso che la lingua rilevata è l'inglese, nello specifico Inglese, Stati Uniti (en-US), con un grado di confidenza pari a 0,62.

Il risultato prodotto dal servizio, riportato nella Trascrizione 4.18, risulta complessivamente accurato e presenta soltanto alcune imprecisioni, principalmente riconducibili al riconoscimento della terminologia specifica del settore. Anche la presenza della musica di sottofondo non ha costituito un ostacolo alla corretta rilevazione dei contenuti. Gli errori riscontrati riguardano principalmente il riconoscimento di acronimi finanziari. In particolare, *GAAP* è stato trascritto come *gap*, mentre *IFRS* è stato identificato in modi differenti, come

There's a couple of acronyms on this list, so let me just talk through, uh, one, the Financial Accounting Standards Board, FASB, you will [...] principles that organizations use for their. Financial reporting IFERS, or the International Financial Reporting Standard, is FASB's peer in the international community. Outside of the US organizations use the IFERS or the International Financial Reporting Standards for governing their financial reporting. There are nuanced differences between FASB and. Gap and IFE's accounting, um, but that's the, [...] The first bucket of costs that we look at are the cost of goods sold, or COGS is the common acronym for this. [...] or services ready for market. and in manufacturing organizations, [...] the first profit metric you get on the PNL is often referred to as gross profit. [...] The next bucket of costs that we look at is often referred to as SGNA. It's an acronym for the Selling General and Administrative expenses. You might [...] functions HR, finance, legal, IT, audit, the corporate executives, all [...] overhead in the organization's business expenses.

Trascrizione 4.18: Trascrizione di una traccia audio in lingua inglese caratterizzata da linguaggio finanziario in assenza di vocabolario personalizzato

IFE e *IFERS*. Al contrario, *COGS* è stato riconosciuto correttamente. Gli acronimi contenenti il simbolo *&*, quali *P&L* e *SG&A*, sono, invece, stati trascritti rispettivamente come *PNL* e *SGNA*.

Risultati con vocabolario personalizzato Analogamente al caso precedente, è stato costruito un vocabolario personalizzato contenente i termini finanziari presenti nella traccia; esso è riportato nella Tabella 4.2.

Phrase	SoundsLike	IPA	DisplayAs
fasbee			FASB
gap			GAAP
ifers			IFRS
cogs			COGS
P.-and-L.			P&L
S.G.-and-A.			SG&A

Tabella 4.2: Vocabolario personalizzato con termini finanziari

Tra gli elementi riconosciuti correttamente dal servizio rientra l'acronimo *GAAP*. Invece, l'acronimo *IFRS* non è stato riportato correttamente ed è stato identificato come *IFersers*. Gli altri due termini, *P&L* e *SG&A*, che non erano stati rilevati correttamente nella precedente trascrizione, continuano a essere riportati nella medesima forma. Pertanto, il vocabolario si è rivelato efficace nella correzione di un solo termine tra quelli inseriti, senza determinare un miglioramento significativo della qualità complessiva della trascrizione.

4.3.7 Trascrizione di contenuti musicali

All'interno di questa sezione vengono analizzate due prove sperimentali che fanno uso di contenuti musicali, con l'obiettivo di valutare la capacità del servizio di rilevare il parlato e il testo cantato in presenza di musica di sottofondo e di pronunce che possono rendere più complesso il riconoscimento dei vocaboli. A tale scopo, sono state selezionate due canzoni, la prima è *Imagine* di John Lennon (1971) (Figura 4.21a), mentre la seconda è *Heal the World* di Michael Jackson (1991) (Figura 4.21b).

Per ciascuna traccia viene inizialmente utilizzata l'identificazione automatica della lingua e, successivamente, viene specificata manualmente la lingua, al fine di confrontare i risultati

(a) *Imagine* **(b)** *Heal the World*

Figura 4.21: QR code per l'accesso alle tracce audio caratterizzate da contenuti musicali

ottenuti con le due modalità. Per la prima canzone si ipotizza che il servizio sia in grado di identificare correttamente la lingua, ma possa incontrare difficoltà nel riconoscimento dei vocaboli a causa della presenza della componente musicale e delle particolari modalità di pronuncia proprie del canto. Per la seconda canzone, invece, la presenza di una porzione iniziale parlata, precedente alla parte cantata, consente di formulare l'ipotesi che il modello possa riconoscere correttamente il parlato, nonostante la presenza della musica di sottofondo, ma possa incontrare maggiori difficoltà nella trascrizione della parte cantata.

Risultati con identificazione automatica della lingua Per la prima canzone, la lingua identificata dal servizio è l'inglese, precisamente Inglese, Stati Uniti (en-US), con un grado di confidenza pari a 0,74, sebbene la variante più appropriata per la traccia sia l'inglese britannico. La Trascrizione 4.19 evidenzia che il servizio non è stato in grado di rilevare tutte le parole presenti nella canzone. Come ipotizzato, tale risultato potrebbe essere dovuto sia alle modalità di pronuncia proprie del canto, sia alla componente strumentale, che costituisce un'ulteriore fonte di disturbo per il riconoscimento dei vocaboli. Specificamente, il modello ha riconosciuto correttamente alcuni vocaboli, come *If* e *sky*, mentre altri, come *love*, sono stati identificati erroneamente o non risultano presenti nel testo. La corretta rilevazione di *sky* nel verso *above us only sky* suggerisce una maggiore facilità nel riconoscimento di parole collocate all'inizio o alla fine dei fraseggi, come osservato anche nel caso di *Nothing to kill*. Infine, dopo *But I'm* non è stato rilevato alcun elemento, probabilmente a causa della maggiore presenza e intensità della parte strumentale. Anche la gestione delle lettere maiuscole e minuscole e della punteggiatura appare disordinata e poco accurata. L'analisi di una porzione del file JSON prodotto dal servizio evidenzia, inoltre, che gli elementi individuati sono caratterizzati da un basso grado di confidenza.

I If you love sky imagine We And didn't. Nothing to kill. I know religion. you And we you A dreamer. But I'm

Trascrizione 4.19: Trascrizione di una traccia audio contenente la canzone *Imagine*

Per completare l'analisi, sono stati esaminati i risultati ottenuti con la seconda canzone, *Heal the World*. La lingua individuata è l'inglese, nello specifico Inglese, Stati Uniti (en-US), con un grado di confidenza pari a 0,66.

Come ipotizzato, il modello ha correttamente identificato la porzione parlata della traccia riportata nella Trascrizione 4.20, associando ad essa un grado di confidenza piuttosto elevato. In questa parte della registrazione non sono stati riscontrati errori significativi, nonostante la presenza della musica di sottofondo. Anche il segmento corrispondente all'inizio della parte

> Think about um the generations and to say we want to make it a better place for our children and our children's children so that they, they, they, they know it's a better world for them. And think if they can make it a better place. There's a place. I know. is It was to cry. Was If far.

Trascrizione 4.20: Trascrizione di una traccia audio contenente la canzone *Heal The World*

cantata, ovvero *That's a place*, è stato rilevato correttamente con un grado di confidenza di circa il 97%; la parte successiva, invece, non è stata trascritta. I vocaboli identificati nella parte cantata sono prevalentemente collocati nella parte finale del fraseggio o caratterizzati da una pronuncia più prossima al parlato, come *to cry*, confermando quanto osservato in precedenza. Nel complesso, i risultati evidenziano che il servizio è in grado di identificare correttamente porzioni di parlato anche in presenza di un sottofondo musicale, mentre incontra notevoli difficoltà nel riconoscimento del contenuto cantato.

Risultati con specificazione manuale della lingua Per il caso relativo alla canzone *Imagine*, la lingua è stata specificata manualmente come Inglese, Regno Unito (en-GB). Dalla trascrizione è possibile osservare che il risultato ottenuto è sostanzialmente identico a quello della prova precedente. Pertanto, l'indicazione manuale della lingua non ha prodotto alcun miglioramento significativo della trascrizione.

4.3.8 Analisi di un audio ad elevata complessità

In questa sezione vengono presentati alcuni esempi relativi a tracce audio caratterizzate dalla combinazione di diversi fattori che possono rendere più complessa l'identificazione dei vocaboli da parte del modello. Gli esempi considerati si avvicinano maggiormente a situazioni d'uso reali, poiché riuniscono in un'unica traccia elementi già analizzati nelle sezioni precedenti, quali la presenza di più parlanti, la varietà degli accenti, i rumori di sottofondo, l'intonazione, la velocità del parlato e l'utilizzo di termini appartenenti a specifici domini.

Audio multilingue con più parlanti e accenti differenti

In questa sottosezione viene analizzata una traccia audio caratterizzata dalla presenza di più parlanti, da un'alternanza tra diverse lingue e da accenti differenti. L'audio, accessibile tramite il QR code riportato in Figura 4.22, è tratto da uno speciale in cui alcuni campioni del motociclismo si confrontano, rivelando alcuni retroscena relativi alle gare e alle proprie modalità di allenamento. Considerate tali caratteristiche, è possibile ipotizzare che il servizio possa incontrare difficoltà nel rilevare correttamente i vocaboli, in particolare quelli pronunciati durante la sovrapposizione degli interventi dei parlanti o in corrispondenza di repentini passaggi da una lingua all'altra. Le lingue presenti nella traccia sono l'inglese, caratterizzato da accenti differenti, lo spagnolo e l'italiano, anch'esso pronunciato con accenti diversi. Inoltre, la durata della traccia è maggiore rispetto a quella degli esempi precedenti, che si aggiravano intorno ai 2 minuti, al fine di verificare se la durata possa incidere sulla qualità della trascrizione.

Risultati con specificazione manuale della lingua Nelle opzioni di richiesta è stato selezionato il riconoscimento automatico di più lingue, indicando come suggerimenti l'inglese (Inglese, Stati Uniti (en-US)), l'italiano (Italiano, IT (it-IT)) e lo spagnolo (Spagnolo, ES (es-ES)).



Figura 4.22: QR code per l'accesso alla traccia audio caratterizzata dalla combinazione di diversi fattori (fonte: `https://www.motogp.com/`)

Inoltre, è stata selezionata l'opzione per la suddivisione dei parlanti, impostando un massimo di 5 persone. Come rilevabile dalla Trascrizione 4.21, il servizio è stato in grado di riconoscere quasi tutti i vocaboli appartenenti alle 3 lingue presenti nell'audio, pur non essendo esente da errori.

E sennò anche lì eri buono? Sì. E poi però secondo me dov'è stato quando Marquez ti ha tolto il 13. 13. E lì mi avevo fatto anche prima male con la clavicola al Saxenry. E era caduto lui lui a O. L. caduto in Olanda, poi ha fatto in testa con tanti punti e la gara dopo si è caduto te. prove. Come hai fatto? La prima curva. Prima curva. E poi dopo si è caduto te, anche un'altra volta, giusto nelle prove. Poi quello quello del. Eh e e poi quello era sempre 13 quindi. Aragon.credibile. Marca l'ha vinto conto. Lui ha fatto 320 eh e io ho fatto 300. Okay. Quello era il più vicino, anche 2012. È vero è vero che a Misano te ha fatto secondo a Misano. A Misano Barbera, no. Sì, ma perché mi hanno messo l'ultimo. Eh sì, perché le le termocopererte era rimasta attaccata. Tu hai fatto secondo fatto secondo che ormai cadi ti ricordi? curva così ho detto. Lui lui cade e io approfitta. cabe los exceso. Mama mía. Ah, Cé visto cuando se me cerrado. metal. Lo viste? No? Sí, mucho. Ah, que que que salió en el replay. No, yo estaba en el box ya el replay salió y yo te he visto. Me caí en la primera vuelta. Tú te pasó eso en la tercera o cuarta, Freddy, I'm curious. Because I rode once your '84 bike, the NSR, yeah, you rode it, yes, [...] And I also wouldn't run an altitude like at Salzburg grain. The V was too close together and the to generate too much heat with the exhaust pipes over the top that when I, when I wrote, yes, they, [...] So you were like suffering more than enjoying in the last years? No, not that, I just. I like to enjoy riding, so I wanna enjoy the sensations of controlling the sky get injured. No, never, I never had a problem for this, no, no, I just didn't enjoy the bikes anymore. The bikes. I was like, [...] Però delle volte vai con la moto. Sì, un po', un po', ma non tanto eh. Io non non mi allenavo tanto in nato. Perché come ero sempre che mi facevo male o avevo paura di farmi male. Certo. Non è che giravo io tanto in moto un po', un po' di cross, all'interno così e poi un po' di mo, ma quando mi sentivo che mi mancava. Qualcosina, sai che non ero proprio posto, allora andavo. E sennò stavo a casa, facevo palestra che era quello che io bisognavo, e bici, bici? Per allenamento ho fatto tanto per imparare a guidare sul bagnato perché sai io non, io il bagnato non s- non ero forte, al 2009. um Ho detto okay. Basta? Eh no. Sì, 2009 perché ti ricordi al Sing che ero primo? Sì, ma sono caduto. Per quello era 2009? 8. 8. Sì che sei caduto curva. Sì. Che eri davanti a noi, a me e a lui e si è si è casato la 1. Lì ho detto okay, sempre perdo il campionato perché c'è una gara che piove, e io sono così e devo andare di più di quello che mi sento, e poi.

Trascrizione 4.21: Trascrizione di una traccia audio caratterizzata dalla combinazione di diversi fattori

In questo segmento emergono numerosi errori di trascrizione, dovuti principalmente alla sovrapposizione delle voci, alla pronuncia e all'alternanza tra italiano, inglese e spagnolo. Diverse parole e intere frasi non sono state rilevate, soprattutto nei punti in cui gli interlocutori parlano contemporaneamente. Alcuni nomi propri di circuiti, come *Sachsenring* e *Salzburgring*, sono stati trascritti erroneamente, rispettivamente come *Saxenry* e *Salzburg*

grain. Sono, inoltre, presenti diverse sostituzioni di parole, come *Olanda* con *O.*, *incredibile* con *credibile*, *moto* con *nato* e *Motard* con *mo*. Anche alcuni segmenti in inglese sono stati riconosciuti solo parzialmente o erroneamente, mentre la breve parte in spagnolo è stata trascritta correttamente. La punteggiatura risulta talvolta coerente con le pause del parlato, ma in altri casi appare poco accurata. In linea con quanto già riscontrato negli esempi precedenti, anche sul piano della diarizzazione il servizio individua tre parlanti a fronte dei cinque effettivamente presenti nella traccia e non associa sempre correttamente i segmenti di discorso ai rispettivi interlocutori, che si alternano in maniera più o meno distribuita nel corso della conversazione. Nel complesso, la sovrapposizione delle voci rappresenta il principale fattore di errore, mentre la pronuncia e il multilinguismo contribuiscono a ulteriori riconoscimenti errati.

Risultati con rilevamento automatico delle lingue Le lingue individuate automaticamente sono: Italiano, IT (it-IT), Inglese, Stati Uniti (en-US), Spagnolo, ES (es-ES). Non si riscontrano differenze significative rispetto alla trascrizione precedente. Gli errori si concentrano, approssimativamente, nelle medesime parti della conversazione.

Inoltre, in entrambe le prove, la maggiore durata della traccia audio non sembra aver influito sulla qualità complessiva dell'output.

Trascrizione di un contenuto audiovisivo

Per questo esempio è stata utilizzata una traccia contenente diverse scene tratte dal film *The Dark Knight* di Christopher Nolan (2008) (Figura 4.23). L'obiettivo è valutare la capacità del sistema di individuare correttamente i vocaboli presenti nella traccia nonostante la presenza di diversi elementi di disturbo, quali effetti sonori di sottofondo, colonna sonora ed enunciati sottovoce o urlati. Si tratta, di fatto, di condizioni maggiormente riconducibili a situazioni d'uso reali e caratterizzate da elementi che possono ostacolare la corretta identificazione del parlato. Si presume, di conseguenza, che l'accuratezza del riconoscimento possa diminuire nei tratti caratterizzati dalla compresenza di più parlanti o da una maggiore intensità della colonna sonora e dei rumori di sottofondo.

Figura 4.23: QR code per l'accesso alla traccia audio contenente diverse scene tratte dal film *The Dark Knight*

Nella richiesta di trascrizione è stata, inoltre, selezionata l'opzione per la generazione di file destinati ai sottotitoli, al fine di valutarne la qualità e verificarne l'effettiva idoneità a un possibile impiego nella sottotitolazione di film, video e altri contenuti multimediali. Nelle impostazioni di trascrizione è stata selezionata la lingua inglese (Inglese, US) e la suddivisione dei parlanti, impostando un massimo di 15 persone. Le evidenze emerse nella Trascrizione 4.22 mostrano un tasso di errore contenuto e dimostrano la capacità del sistema di gestire correttamente gli elementi potenzialmente in grado di interferire con il riconoscimento.

Le imprecisioni riscontrate sono riconducibili principalmente alla presenza di musica di sottofondo, alla pronuncia e agli accenti. Ad esempio, alcune delle parole riconosciute erroneamente dal sistema sono: *Attention*, trascritta come *hot*, *Aim! Fire!* come *all right*, *the*

What do you got on the roof? We're tight, but frankly, there's a lot of windows up here. Commissioner Loeb dedicated his life [...] and letters. Who was that? What happened? He took our guns. Uniforms. We must [...] safety. Stand by, Honor, guard. Hot. Ready? All right. What makes you think I wanna hear you talk? I want the Joker. [...] spot from this height. Paul wouldn't kill me. I'm counting on it. Who? Where is he? I don't know where he is. He found us. He must have friends, friends. [...] up your mind. Not so crazy as you look. I told you I'm a man of my word. [...] in 20 minutes. I want auxiliary units to cover all possible getaway routes. Doorman, what you do with all your money? [...] I want to know how much they're paying you to say who Batman really is. This is our chance. I want Lau alive. The Joker, either way. Hey Jim, get a load of this. [...] Harvey Dent didn't want [...] him now. Sergio Roman of your word. I am. [...] And it was so Boring. I've had a change of heart. I don't want Mr. Reese spoiling everything, [...] The priority is Gotham General. Wheel everybody [...] Where are we going, sir? To get Reese.

Trascrizione 4.22: Trascrizione di una traccia audio contenente scene tratte dal film *The Dark Knight*

fall e *wouldn't kill me* come *Paul*, mentre *Doorman* è stata rilevata al posto di *Joker man*. Alcuni segmenti, tra cui *Port arms*, *Can't we go someplace quieter? We can't hear each other talk.* e *What?*, non sono stati trascritti, presumibilmente a causa del volume molto elevato della musica di sottofondo. Un risultato particolarmente positivo riguarda, invece, il riconoscimento corretto dei nomi propri, inclusi quelli direttamente associati al film, come *Gotham General*. Per quanto riguarda l'identificazione dei parlanti, la prima frase della traccia evidenzia già una criticità, poichè il sistema attribuisce l'intero segmento a un unico parlante, nonostante esso corrisponda a uno scambio tra due persone. La stessa problematica si ripresenta nel resto della traccia, che viene complessivamente suddivisa tra soli 8 parlanti. Di conseguenza, anche il file di trascrizione presenta una segmentazione poco coerente, risultando frammentato e non sempre rappresentativo dell'effettiva alternanza tra i parlanti. A fronte di queste limitazioni, l'allineamento temporale della trascrizione presenta risultati significativamente migliori. Il sistema individua, infatti, con elevata precisione il momento in cui le singole frasi vengono pronunciate all'interno della traccia audio. Tale caratteristica è stata considerata particolarmente rilevante in questo caso di studio, poiché ha consentito di valutare non soltanto l'accuratezza della trascrizione, ma anche, più in generale, la qualità dei sottotitoli prodotti. In Figura 4.24 è riportato il QR code che consente di scaricare il file contenente i sottotitoli generati.

Figura 4.24: QR code per l'accesso ai sottotitoli generati dall'analisi della traccia audio contenente scene riprese dal film *The Dark Knight*

Contestualemente, è stata valutata la funzionalità di *Filtraggio del vocabolario*. A tal fine, è stato definito un apposito vocabolario filtro, riportato in Tabella 4.3, contenente una serie di termini potenzialmente tossici da individuare e mascherare nella trascrizione. Tra questi rientrano, ad esempio, *dynamite*, *kill* e altri termini analoghi.

Come specificato nella documentazione di Amazon, il sistema considera le parole inserite

Termini
guns
kill
killed
dynamite
gunpowder
gasoline
maniac
burning
criminal
burns
dead

Tabella 4.3: Vocabolario filtro

nel vocabolario filtro in modo letterale. Di conseguenza, se nel vocabolario viene inserita la forma *imprecare*, ma nel file multimediale compare, ad esempio, la forma *impreca* o *imprecando*, queste ultime non vengono filtrate. Pertanto, è necessario inserire nel vocabolario tutte le varianti morfologiche delle parole che si desiderano filtrare. Per la trascrizione analizzata è stato selezionata la modalità di filtraggio *maschera vocabolario*. Come è possibile osservare nella Trascrizione 4.23, tutte le parole precedentemente inserite nel vocabolario filtro sono state correttamente mascherate.

> [...] What happened? He took our ***. Uniforms. We must [...] Paul wouldn't *** me. [...] Let him come find you. How are you gonna let a couple more people get *** while you make up your mind. [...] You see, I'm a guy of simple taste. I enjoy, uh, ***. And *** and ***. What's it. [...] Harvey Dent didn't want to give in to this ***. Do you [...]. I am. I'm only *** my half. All you care about is money. This town deserves a better class of ***. And I'm [...] Everything ***. Who is this? [...] If Coleman Reese isn't *** in 60 minutes, [...].

Trascrizione 4.23: Trascrizione di una traccia audio contenente scene tratte dal film *The Dark Knight* abilitando la funzionalità di *Filtraggio del vocabolario*

Quindi, la funzionalità ha dimostrato di poter rappresentare uno strumento utile per la gestione e il filtraggio di questa tipologia di contenuti. A conclusione dei casi di studio presentati, è opportuno sottolineare che le ultime prove sono state concepite con l'obiettivo di mettere maggiormente alla prova il sistema e di far emergere eventuali difficoltà e potenziali limitazioni proprie dei sistemi di riconoscimento automatico del parlato. A differenza delle prime prove, nelle quali sono stati analizzati principalmente singoli aspetti del processo di trascrizione, queste ultime hanno permesso di valutare la risposta del servizio alla presenza simultanea di più fattori e di analizzare l'impatto della loro combinazione sulle prestazioni complessive del sistema. Un'ulteriore osservazione emersa complessivamente dalla sperimentazione riguarda il grado di confidenza associato alle rilevazioni. Come già constatato nei casi di studio relativi al riconoscimento del testo nelle immagini mediante Amazon Rekognition, un elevato livello di confidenza non costituisce necessariamente un'indicazione di correttezza dell'identificazione effettuata dal servizio.

CAPITOLO 5

Un case study relativo ai chatbot

In questo capitolo viene illustrato l'ultimo caso di studio proposto nella presente trattazione, relativo alla realizzazione di un chatbot mediante il servizio Amazon Lex di AWS. Il capitolo si apre con un'introduzione ai chatbot, fornendo una panoramica delle principali tipologie e del relativo funzionamento. Successivamente, vengono analizzati i potenziali rischi e le principali sfide nell'adozione di tale tecnologia. Viene, quindi, presentato il servizio utilizzato per il caso di studio, illustrandone le modalità di integrazione con gli altri servizi messi a disposizione dalla suite AWS. In seguito, viene presentato il caso di studio, con particolare riferimento alla progettazione di un chatbot finalizzato a fornire assistenza nella scelta dei corsi universitari. Vengono illustrate l'architettura adottata e le modalità di integrazione degli altri servizi AWS impiegati nella realizzazione di un'interfaccia utente attraverso la quale è possibile interagire con il bot. In particolare, viene descritta l'implementazione delle funzioni Lambda. A conclusione del capitolo, viene presentata una sezione dedicata ai test e alla valutazione del prodotto finale, con particolare riferimento alle funzionalità integrate messe a disposizione da Amazon Lex.

5.1 Introduzione ai chatbot

Con il termine *chatbot*, o più propriamente *chatterbot*, si indica un'applicazione software progettata per simulare una conversazione con un essere umano attraverso l'utilizzo della voce o del testo. L'obiettivo principale di un chatbot è fornire risposte automatiche in grado di supportare gli utenti nello svolgimento di diverse attività, quali l'assistenza ai clienti e il reperimento di informazioni. Spesso i chatbot si avvalgono di tecniche di *elaborazione del linguaggio naturale* (NLP), di sistemi di *IA generativa* e di *Large Language Model* (LLM) per comprendere l'intento dell'utente e generare risposte quanto più naturali e coerenti possibile. In questo contesto, è possibile affermare che i chatbot rappresentano un'importante evoluzione nelle modalità di erogazione dei servizi digitali.

5.1.1 Tipologie di chatbot

Attualmente, il termine ''chatbot'' viene spesso associato a sistemi conversazionali basati su tecniche di Intelligenza Artificiale. Tuttavia, i chatbot esistono da diversi decenni, antecedentemente allo sviluppo delle moderne tecnologie di IA. I chatbot possono essere classificati sulla base del livello di complessità tecnologica e delle funzionalità offerte, considerando, inoltre, che le relative caratteristiche sono soggette a una continua evoluzione. Tra le principali tipologie di chatbot è possibile distinguere le seguenti:

- *Chatbot tradizionali*. In questa categoria rientrano i chatbot caratterizzati da una struttura relativamente semplice, sviluppati antecedentemente all'introduzione delle moderne tecniche di IA. Essi presentano capacità limitate e risultano particolarmente adatti alla gestione di interazioni prevedibili e strutturate. Per tale motivo, vengono generalmente impiegati dalle aziende sui siti web per fornire risposte alle domande frequenti (FAQ) e guidare l'utente nello svolgimento di attività ripetitive. In base alla modalità con cui vengono fornite le risposte all'utente, i chatbot tradizionali possono essere ulteriormente suddivisi in:

  - *Chatbot basati su regole*: guidano l'utente attraverso una serie di passaggi e domande predefiniti. In questo caso, l'utente non può formulare liberamente una richiesta, ma deve selezionarla tra le opzioni disponibili. Alcuni di questi sistemi utilizzano alberi decisionali per indirizzare l'utente lungo percorsi predefiniti, fornendo risposte limitate a determinati scenari.
  - *Chatbot basati su parole chiave*: estraggono specifiche parole chiave all'interno delle richieste dell'utente e forniscono le risposte associate alle corrispondenti occorrenze.

- *Chatbot basati sull'IA*. A differenza dei chatbot tradizionali, questa tipologia sfrutta le moderne tecnologie di Intelligenza Artificiale per gestire una gamma più ampia di richieste e fornire risposte maggiormente articolate. All'interno di questa categoria è possibile individuare diverse tipologie:

  - *Chatbot basati su modelli generativi*: utilizzano gli LLM per generare risposte, consentendo un'interazione più naturale e fluida rispetto ai chatbot tradizionali. Tra gli esempi più noti rientrano *ChatGPT*, *Claude* e *Gemini*.
  - *Chatbot con RAG e funzioni integrate*: combinano tecniche avanzate, quali la RAG, con l'accesso a database, documentazione o altre fonti informative esterne. Inoltre, essi possono integrare funzioni specifiche del contesto aziendale, ampliando le capacità del sistema rispetto alla semplice generazione di risposte.

  In particolare, i chatbot basati sull'IA possono essere classificati anche in funzione del dominio applicativo di riferimento. Essi possono essere progettati per sostenere conversazioni su argomenti appartenenti a domini differenti e caratterizzati da un'elevata varietà tematica, oppure possono essere destinati allo svolgimento di attività specifiche, operando all'interno di un dominio più ristretto (Figura 5.1). Questa seconda tipologia consente di mantenere un maggiore controllo sul flusso della conversazione e sulle operazioni che il chatbot è in grado di eseguire, a fronte di una minore libertà concessa all'utente nell'interazione. Tale approccio è stato adottato anche per il chatbot realizzato nell'ambito del caso di studio discusso nel presente capitolo.

È, inoltre, opportuno distinguere tra chatbot basati sull'IA, *assistenti AI* e *agenti AI*. Sebbene tali termini vengano talvolta utilizzati in modo intercambiabile, essi identificano sistemi caratterizzati da funzionalità differenti, pur presentando confini non sempre nettamente definiti e potendo, in alcuni casi, sovrapporsi. Gli assistenti AI, oltre a rispondere alle domande degli utenti, possono eseguire diverse azioni, quali impostare sveglie e promemoria, interagire con altre applicazioni software e gestire calendari. *Siri*, *Amazon Alexa* e *Google Assistant* costituiscono alcuni esempi di assistenti AI. Gli agenti AI, invece, sono sistemi in grado di pianificare, prendere decisioni sulla base delle informazioni disponibili e gestire flussi di lavoro articolati in più fasi, riducendo la necessità di un intervento umano diretto.

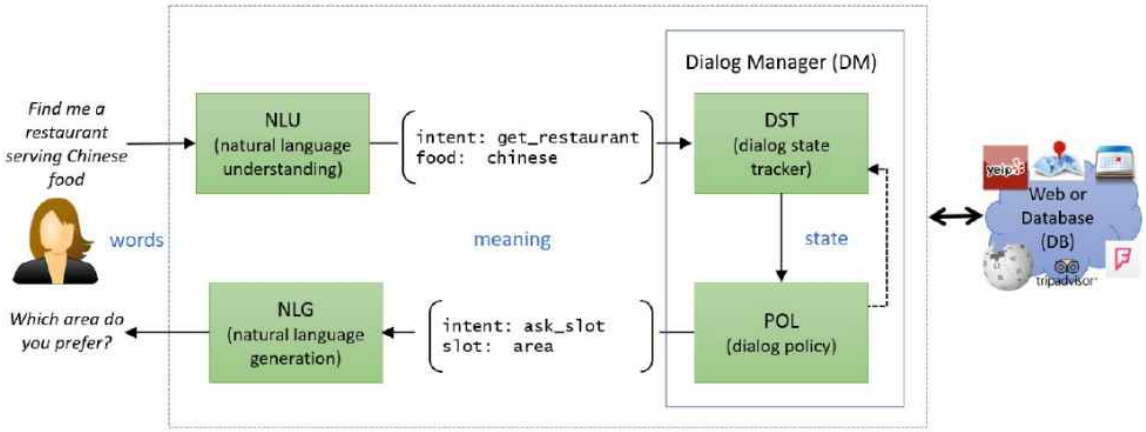


Figura 5.1: Schema illustrativo del processo di funzionamento di un chatbot multi-turn task-oriented (fonte: https://www.researchgate.net/)

5.1.2 Funzionamento dei chatbot

Sebbene i chatbot possano differire significativamente in termini di architettura, tecnologie impiegate e livello di complessità, il relativo flusso di funzionamento può essere ricondotto a una sequenza di fasi fondamentali, comune alle diverse tipologie di sistema. Nello specifico, è possibile individuare i seguenti passaggi:

- *acquisizione*: il sistema acquisisce la richiesta dell'utente attraverso un input testuale o vocale;
- *elaborazione*: l'input ricevuto viene analizzato e interpretato dal sistema al fine di determinarne il significato e l'intento sottostante;
- *output*: sulla base dell'elaborazione effettuata, il sistema genera una risposta da restituire all'utente oppure procede all'esecuzione di un'azione coerente con la richiesta ricevuta.

Dal punto di vista architetturale, il funzionamento di un chatbot può essere descritto attraverso la suddivisione del sistema in tre componenti principali:

- *Livello utente*: rappresenta il livello di interazione tra l'utente e il sistema ed è costituito dall'interfaccia attraverso la quale vengono acquisite le richieste e vengono restituiti i risultati dell'elaborazione.
- *Livello di elaborazione*: costituisce il nucleo logico del sistema e comprende i componenti responsabili dell'interpretazione delle richieste, della gestione del dialogo e dell'esecuzione delle azioni. Tale livello può essere ulteriormente suddiviso in tre moduli principali. Il primo è costituito dai moduli deputati alla comprensione del linguaggio, i quali analizzano la richiesta dell'utente e ne determinano il contesto e l'intento. Tali moduli possono, inoltre, svolgere attività di riscrittura della query, tenendo conto, ove necessario, della cronologia della conversazione, al fine di ottenere una rappresentazione più completa della richiesta e facilitare il reperimento delle informazioni pertinenti da documenti o knowledge base. Il secondo modulo è costituito dai componenti preposti alla gestione del dialogo, il cui compito consiste nel mantenere e aggiornare lo stato della conversazione, determinando le modalità con cui il sistema deve procedere nell'interazione. Infine, il terzo modulo comprende i componenti responsabili della gestione e dell'esecuzione delle azioni previste dal sistema.

- *Livello dati*: comprende le risorse informative utilizzate dal sistema durante il processo di elaborazione. In tale livello rientrano, ad esempio, documenti indicizzati, knowledge base, database e altre strutture dati dalle quali possono essere recuperate le informazioni necessarie alla formulazione delle risposte. In funzione delle caratteristiche e degli obiettivi del sistema, il livello dati può, inoltre, comprendere informazioni relative al profilo e alla cronologia degli utenti, consentendo di fornire risposte maggiormente contestualizzate e personalizzate.

5.1.3 Sfide e rischi dei chatbot

I chatbot rappresentano una tecnologia in grado di supportare gli utenti nello svolgimento di numerose attività quotidiane. Nonostante il progressivo sviluppo di tali sistemi e il raggiungimento di livelli di complessità sempre maggiori, risulta necessario prestare particolare attenzione alla loro evoluzione e alle modalità con cui vengono progettati e gestiti. L'adozione di questa tecnologia non richiede, infatti, esclusivamente un'adeguata valutazione degli aspetti tecnologici, ma anche una corretta gestione dei processi di governance, manutenzione e pianificazione. In tale contesto, emergono diverse criticità e potenziali rischi associati all'utilizzo dei chatbot, tra i quali è possibile individuare i seguenti:

- *Distorsioni*: i chatbot possono manifestare forme di distorsione o pregiudizio derivanti, tra gli altri fattori, dai dati utilizzati durante il processo di addestramento o dalle informazioni alle quali il sistema ha accesso. Dunque, risulta necessario prevedere adeguati meccanismi di monitoraggio e valutazione del comportamento del chatbot, al fine di individuare e, ove possibile, mitigare eventuali risposte caratterizzate da tali distorsioni.

- *Gestione di richieste complesse*: i chatbot possono presentare limitazioni nella gestione di richieste particolarmente articolate, soprattutto nei casi in cui sia necessario ricorrere a valutazioni che richiedono il giudizio umano. Pertanto, il sistema deve essere progettato in modo da riconoscere i propri limiti operativi e prevedere, quando necessario, il trasferimento della conversazione a un operatore umano.

- *Risposte errate o fuorvianti*: i chatbot possono generare risposte inaccurate, incomplete o potenzialmente fuorvianti. Nel caso dei sistemi basati sull'IA generativa, tale fenomeno può manifestarsi attraverso le cosiddette *allucinazioni*. La probabilità di generare tali risposte può aumentare, ad esempio, quando il sistema dispone di fonti informative non sufficientemente affidabili oppure quando vengono formulate richieste al di fuori del dominio di conoscenza per il quale è stato progettato.

- *Privacy e sicurezza*: i chatbot possono elaborare quantità significative di dati, inclusi dati sensibili relativi agli utenti o alle organizzazioni. È, perciò, necessario adottare adeguate misure di protezione dei dati e di sicurezza, nonché implementare meccanismi di controllo degli accessi alle risorse e ai sistemi aziendali, garantendo, al contempo, il rispetto delle normative in materia di protezione dei dati personali.

- *Trasparenza nei confronti degli utenti*: è opportuno che gli utenti siano adeguatamente informati circa la natura dell'interazione con il sistema, soprattutto quando essa avviene con un chatbot anziché con un operatore umano. Dovrebbero, inoltre, essere fornite informazioni chiare in merito alle modalità di trattamento dei dati e alle risorse informative alle quali il sistema può avere accesso.

5.1.4 Casi d'uso dei chatbot

I chatbot possono essere impiegati in numerosi settori e contesti applicativi differenti. Tra i principali ambiti applicativi è possibile individuare i seguenti:

- *Servizio clienti*: rappresenta uno dei principali ambiti di impiego dei chatbot. Tali sistemi consentono di fornire assistenza agli utenti senza richiedere necessariamente l'intervento di un operatore umano. In particolare, possono essere utilizzati per reperire informazioni relative a prodotti e servizi, monitorare lo stato degli ordini e fornire supporto nella risoluzione di problematiche comuni.
- *Supporto aziendale*: i chatbot possono essere integrati all'interno degli ambienti di lavoro per supportare i dipendenti nello svolgimento di diverse attività organizzative. Tra queste rientrano, ad esempio, la consultazione delle politiche aziendali, la gestione delle richieste IT più frequenti e l'accesso centralizzato alle risorse e alle informazioni aziendali.
- *Recupero delle informazioni*: questi sistemi possono facilitare il reperimento rapido di informazioni all'interno di siti web, documenti e knowledge base.
- *Automazione dei flussi di lavoro*: l'integrazione dei chatbot all'interno dei processi organizzativi consente di automatizzare determinate attività di routine. Tra queste rientrano, ad esempio, l'inoltro delle richieste al team competente, l'aggiornamento dei registri aziendali e la raccolta delle informazioni necessarie alla gestione di nuovi eventi o richieste.

5.2 Amazon Lex

Prima di procedere con l'illustrazione del caso di studio oggetto del presente capitolo, è opportuno fornire una panoramica del servizio utilizzato per la realizzazione del chatbot, ossia *Amazon Lex*. Amazon Lex è un servizio della suite AWS che consente di progettare e sviluppare interfacce conversazionali, quali chatbot e agenti conversazionali, integrabili all'interno di applicazioni, siti web e piattaforme di comunicazione. Il servizio mette a disposizione funzionalità basate su NLU e ASR, consentendo al sistema di interpretare le richieste degli utenti e di gestire le interazioni attraverso modalità sia testuali sia vocali. Amazon Lex consente, inoltre, di integrare LLM, al fine di migliorare, in determinati scenari, la classificazione degli intenti e la risoluzione degli slot. L'integrazione di tali funzionalità permette di aumentare la flessibilità del sistema e di supportare la gestione di conversazioni formulate in linguaggio naturale. Il servizio può essere impiegato in diversi contesti applicativi e trova particolare impiego nella realizzazione di bot destinati all'assistenza clienti.

5.2.1 Integrazione con altri servizi AWS

Amazon Lex mette a disposizione le funzionalità necessarie per la realizzazione di chatbot caratterizzati da differenti livelli di complessità. Tuttavia, è possibile integrare il servizio con altre componenti dell'ecosistema AWS al fine di estenderne le capacità e ottenere un maggiore livello di personalizzazione, in relazione agli obiettivi e ai requisiti del sistema. Tra i principali servizi e funzionalità integrabili con Amazon Lex è possibile individuare i seguenti:

- *Amazon Bedrock*: consente di integrare funzionalità di IA generativa e di avvalersi di modelli linguistici e agenti per la realizzazione di esperienze conversazionali più avanzate. Amazon Bedrock mette, inoltre, a disposizione le *Knowledge Base*, che consentono

di utilizzare informazioni provenienti da fonti informative esterne e modelli quali *Anthropic Claude* per supportare specifiche funzionalità di elaborazione degli enunciati.

- *Assisted NLU*: permette di migliorare la classificazione degli intenti e la risoluzione degli slot mediante l'impiego di LLM.
- *AWS Lambda*: consente di estendere il comportamento del chatbot attraverso l'esecuzione di funzioni personalizzate.
- *Amazon Comprehend*: consente di effettuare sentiment analysis associato alle richieste dell'utente e, di conseguenza, supportare la gestione del flusso conversazionale.
- *Amazon Polly*: consente di convertire il testo delle risposte generate dal chatbot in parlato mediante tecniche di sintesi vocale, rendendo possibile la restituzione delle risposte attraverso un'interazione vocale.
- *Amazon Transcribe*: consente all'utente di interagire con il bot mediante input vocale in alternativa alla consueta modalità testuale.

5.2.2 Funzionamento di Amazon Lex

L'interfaccia di Amazon Lex per la creazione e la configurazione di un nuovo bot presenta una struttura orientata a semplificare le principali operazioni di sviluppo e gestione. Dalla pagina iniziale è possibile creare un nuovo bot, oppure gestire quelli precedentemente sviluppati e procedere, ove necessario, alle operazioni di importazione ed esportazione. La pagina dedicata alla configurazione del bot consente di scegliere tra diverse modalità per la creazione di un nuovo chatbot. In particolare, è possibile ricorrere alla modalità tradizionale oppure alla modalità basata sull'IA generativa, mediante l'integrazione con Amazon Bedrock. Nella modalità tradizionale, è possibile selezionare tre modalità di creazione del bot:

- *Creazione di un bot vuoto*: permette di sviluppare un chatbot partendo da una configurazione priva di lingue, intenti e slot preconfigurati, lasciando allo sviluppatore la definizione completa della struttura del sistema.
- *Inizio da un modello di esempio*: permette di creare un bot a partire da un modello contenente lingue, intenti e slot preconfigurati, consentendo allo sviluppatore di apportare eventuali modifiche e adattarlo alle specifiche esigenze.
- *Inizio dalle trascrizioni*: consente di generare automaticamente gli intenti a partire dalle trascrizioni di conversazioni precedentemente caricate. Tale modalità è attualmente disponibile esclusivamente per la lingua inglese (Stati Uniti).

Nella sezione *Configurazione bot* è possibile inserire il nome con cui identificare il bot e fornire una breve descrizione delle relative finalità. È, inoltre, possibile impostare il timeout della sessione, ovvero l'intervallo di tempo durante il quale la sessione del bot rimane attiva in assenza di nuovi input da parte dell'utente. Il valore massimo configurabile è pari a 24 ore. Durante tale intervallo, Amazon Lex mantiene i dati raccolti mediante gli slot e gli attributi associati alla sessione, consentendo all'utente di riprendere l'interazione e fornire eventuali informazioni non precedentemente specificate. La configurazione del bot consente di selezionare una o più lingue supportate, considerando che alcune funzionalità sono disponibili esclusivamente per la lingua inglese. Analogamente, la selezione della regione AWS riveste particolare importanza, in quanto la disponibilità di determinate funzionalità può variare in funzione dell'area geografica selezionata. Un ulteriore parametro configurabile è rappresentato dalla soglia di confidenza relativa alla classificazione dell'intento, espressa

mediante un valore compreso tra 0,00 e 1,00. La possibilità di configurare tale parametro permette di influenzare il comportamento del sistema in presenza di classificazioni caratterizzate da differenti livelli di confidenza e può risultare utile anche durante la fase di test per analizzare gli effetti derivanti da variazioni nella formulazione degli enunciati associati a un determinato intento. A seguito della creazione del bot, viene generato automaticamente un nuovo intento denominato *NewIntent*, che può essere successivamente modificato. Un *intento* (o intent) rappresenta un obiettivo che l'utente intende raggiungere attraverso l'interazione con il chatbot, ad esempio la prenotazione di un volo o di una camera d'albergo oppure la richiesta di informazioni relative a un determinato prodotto. Ogni bot deve disporre di almeno un intento. Per impostazione predefinita, è presente un intento integrato denominato *FallbackIntent*, utilizzato da parte del sistema quando non è in grado di associare l'input dell'utente a uno degli altri intenti definiti. Per ciascuna lingua associata al bot è possibile creare nuovi intenti, che possono essere distinti nelle seguenti tipologie:

- *intenti vuoti*: intenti privi di enunciati di esempio predefiniti;
- *intenti incorporati*: intenti appartenenti a una libreria standard integrata; per tali intenti sono disponibili due operazioni, quali l'aggiunta o la rimozione degli enunciati di esempio e la configurazione degli slot necessari.

Un intento è costituito da una serie di componenti che concorrono al completamento della richiesta dell'utente, tra cui rientrano i seguenti:

- *sample utterances*: frasi rappresentative che descrivono le possibili formulazioni utilizzate dall'utente per richiamare uno specifico intento;
- *risposta iniziale*: consente di definire il messaggio restituito all'utente all'avvio dell'elaborazione dell'intento;
- *slot*: rappresentano le informazioni che il bot deve acquisire dall'utente per poter completare l'intento;
- *conferma*: consente di richiedere all'utente una conferma prima di procedere al completamento dell'intento;
- *fulfillment*: quando abilitato, consente di richiamare una funzione AWS Lambda per soddisfare l'intento e fornire all'utente un riscontro relativo allo stato di completamento dell'operazione;
- *risposta di chiusura*: consente di definire il messaggio restituito al termine della gestione dell'intento.

Inoltre, per ciascun componente dell'intento è possibile abilitare la funzionalità di *Ramificazione condizionale* attraverso la quale è possibile definire il passo successivo della conversazione sulla base delle condizioni definite. Prima di procedere alla configurazione degli intenti, è opportuno definire i *tipi di slot* necessari alla gestione delle informazioni richieste. Gli slot rappresentano, infatti, i parametri variabili che devono essere forniti da parte dell'utente affinché il bot possa completare un determinato intento. Sono disponibili due principali tipologie di slot:

- *slot di tipo personalizzato*: consentono di definire e, quindi, acquisire valori personalizzati specifici per un determinato intento;
- *slot di tipo integrato*: consentono di acquisire valori appartenenti a tipologie standard già definite dal servizio, quali indirizzi e-mail, date, città e numeri.

Nel caso in cui vengano associati uno o più slot a un intento, è possibile configurare specifiche risposte e azioni da eseguire in funzione dell'esito dell'acquisizione. In particolare, l'*Acquisizione slot: risposta al successo* consente di definire il comportamento del sistema a seguito della riuscita acquisizione dello slot, mentre l'*Acquisizione slot: risposta ai fallimenti* permette di specificare il passaggio successivo nel caso in cui il valore richiesto non venga correttamente acquisito. Infine, dalla pagina dedicata alla gestione delle lingue associate a uno specifico bot, selezionando una specifica lingua, è possibile accedere alle seguenti funzionalità:

- *Comprensione del linguaggio naturale (NLU) assistita*: consente di migliorare la precisione nell'identificazione degli intenti e nella risoluzione degli slot attraverso l'impiego di LLM.
- *Risoluzione assistita degli slot*: permette di aumentare la precisione nell'acquisizione dei valori degli slot mediante l'impiego di LLM. Tale funzionalità può essere utilizzata esclusivamente con gli slot di tipo integrato.

Le due opzioni non possono essere abilitate simultaneamente. Tale limitazione rappresenta un aspetto rilevante ai fini della valutazione del servizio e sarà analizzata più approfonditamente nella Sezione 5.3.9.

5.3 Case study relativo alla creazione di un chatbot

In questa sezione viene presentato il caso di studio relativo all'utilizzo del servizio Amazon Lex per la realizzazione di un chatbot finalizzato a fornire assistenza nella scelta dei corsi opzionali offerti dall'Università Politecnica delle Marche per il corso di studi in Ingegneria Informatica e dell'Automazione. Lo sviluppo del chatbot mediante Amazon Lex ha lo scopo di valutare le funzionalità messe a disposizione dal servizio e di individuare gli aspetti a cui prestare particolare attenzione nella progettazione di un chatbot. Il chatbot si propone come uno strumento di supporto agli studenti nella definizione del proprio piano di studi, sia nell'ambito del percorso di laurea triennale che magistrale. In una prima fase, la sua implementazione è stata circoscritta al corso di studi in Ingegneria Informatica e dell'Automazione; tuttavia, il bot è stato concepito con la possibilità di consentirne una successiva estensione agli altri percorsi di studio offerti dall'università. Inoltre, esso potrebbe essere integrato nelle piattaforme web già in uso presso l'università per renderlo facilmente consultabile dagli studenti, i quali possono incontrare difficoltà nel reperire informazioni relative all'offerta formativa e nell'individuare le scelte più appropriate per la composizione del proprio piano di studi.

5.3.1 Criteri di analisi

L'obiettivo principale del caso di studio non consiste semplicemente nella creazione di un chatbot mediante Amazon Lex e nella successiva valutazione del suo comportamento, ma anche nel comprendere le modalità attraverso cui integrare tale servizio all'interno di un'applicazione web, per valutarne un possibile impiego in un contesto applicativo reale. Per tale motivo, oltre a illustrare il processo di realizzazione del flusso conversazionale mediante Amazon Lex, viene descritta l'architettura complessiva adottata per la realizzazione dell'interfaccia web. In questo modo, è stato possibile evidenziare sia gli aspetti positivi che le eventuali problematiche riscontrate durante la fase di implementazione. La valutazione ha considerato alcuni parametri relativi al comportamento del bot, tra cui il riconoscimento degli intenti, la risoluzione degli slot, la frequenza con cui il bot richiama l'intento *Fallback*, la

completezza delle informazioni restituite e la coerenza dello stato conversazionale. L'analisi ha tenuto conto anche delle funzionalità basate sull'IA messe a disposizione dal servizio, che sono state abilitate e impiegate ove possibile. La realizzazione del bot è stata effettuata mediante la modalità tradizionale, partendo dalla creazione di un bot vuoto, per disporre di una maggiore flessibilità nella definizione della struttura e del flusso conversazionale. Per motivazioni di natura tecnica che non hanno consentito l'accesso ai modelli V3 di Anthropic mediante Amazon Bedrock, non è stato possibile realizzare una seconda versione del bot mediante la modalità basata su IA generativa. La possibilità di sviluppare il bot secondo tale modalità e di effettuare un confronto tra le due implementazioni viene, pertanto, rimandata a una futura trattazione.

5.3.2 Architettura adottata

In questa sezione, prima di analizzare nel dettaglio il flusso conversazionale pensato per il chatbot, viene descritta l'architettura complessiva adottata per la realizzazione del bot e dell'annessa interfaccia utente, come mostrato in Figura 5.2. Nelle sezioni successive verranno approfondite le modalità con cui i diversi componenti sono stati impiegati; per il momento, ne viene offerta una panoramica generale, volta a introdurre il comportamento del bot e a descrivere le principali scelte progettuali adottate.

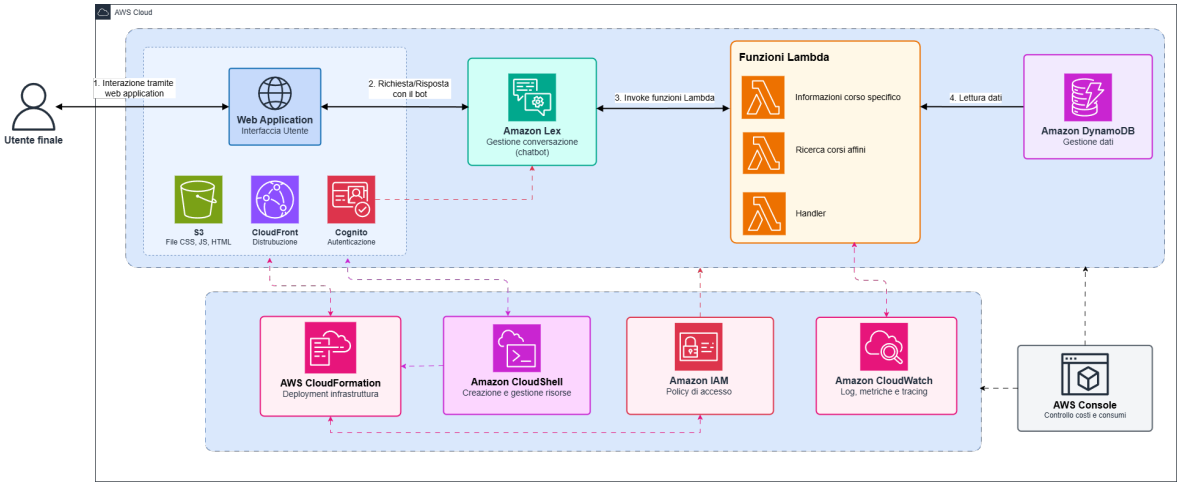


Figura 5.2: Architettura impiegata per la realizzazione del bot

L'obiettivo originario era sfruttare le funzionalità offerte da ulteriori servizi AWS integrabili con il bot, per incrementare le prestazioni del chatbot. A tal proposito, sono state inizialmente vagliate diverse possibili soluzioni architetturali che, tuttavia, non è stato possibile implementare a causa dei costi associati all'utilizzo di tali servizi e dei permessi necessari alla loro integrazione. Tra le soluzioni considerate rientrano le seguenti:

- Per ottenere dati costantemente aggiornati, è stata considerata la possibilità di creare una knowledge base mediante Amazon Bedrock, basata sul sito web contenente le informazioni relative ai corsi disponibili. In tal modo, a fronte di eventuali modifiche alle informazioni presenti nel sito, il chatbot avrebbe potuto accedere direttamente ai contenuti aggiornati, consentendo all'utente di ottenere informazioni più recenti e risposte maggiormente dettagliate.
- In alternativa, è stata considerata la possibilità di utilizzare documenti opportunamente strutturati e formattati, per facilitarne il parsing, contenenti le informazioni relative ai

corsi e archiviati all'interno di un bucket Amazon S3. Su tali documenti sarebbe stata, successivamente, costruita una knowledge base mediante Amazon Bedrock.

In questi scenari, la knowledge base collegata al bot Amazon Lex permetterebbe di demandare alcune richieste degli utenti a uno dei modelli di LLM selezionabili dalla console di Amazon Bedrock, quali Claude 3.5 Haiku o Amazon Nova Lite. Tuttavia, a causa delle limitazioni relative ai permessi necessari per l'accesso a tali modelli, è stato necessario optare per un'architettura più semplice, ma comunque funzionale rispetto agli obiettivi del caso di studio. Tale architettura include:

- funzioni AWS Lambda per la gestione delle richieste dell'utente che non possono essere soddisfatte mediante il solo flusso conversazionale definito in Amazon Lex;
- un database realizzato mediante Amazon DynamoDB, contenente le informazioni relative ai corsi offerti, interrogato dalle funzioni Lambda per il recupero dei dati necessari durante la conversazione.

Introduzione ai servizi utilizzati

In generale, la console di gestione centralizzata di AWS, già descritta nella Sezione 2.4.1, ha consentito di accedere ai diversi servizi utilizzati e di monitorare i costi e i consumi. I servizi utilizzati per la realizzazione del chatbot e della relativa interfaccia utente, oltre ad Amazon Lex, sono i seguenti:

- *Amazon Lambda*: utilizzato per la creazione di funzioni Lambda utili alla gestione di particolari aspetti della conversazione con l'utente non direttamente realizzabili tramite i soli strumenti offerti da Amazon Lex[1].
- *Amazon CloudWatch*: utilizzato per gestire il deployment delle funzioni Lambda create, nonché per analizzare accuratamente il flusso di esecuzione delle richieste e verificare la corretta integrazione delle funzioni Lambda.
- *Amazon IAM*: utilizzato per la gestione delle autorizzazioni e dei permessi necessari ai diversi componenti dell'architettura.
- *Amazon DynamoDB*: utilizzato per la creazione di un database NoSQL contenente i dati relativi ai corsi opzionali offerti dall'Università Politecnica delle Marche, con l'obiettivo di poterli recuperare durante la conversazione e fornire informazioni più dettagliate sui corsi.
- *Amazon CloudShell*: utilizzato per l'esecuzione di comandi da terminale e per la gestione delle operazioni necessarie alla creazione del database DynamoDB e della web application, sfruttando alcuni sample già messi a disposizione da Amazon Web Services e successivamente adattati allo specifico caso d'uso.
- *Amazon CloudFormation*: utilizzato per consentire il deployment automatizzato dell'infrastruttura, attraverso la definizione di un template in cui sono specificati i servizi impiegati e le relative configurazioni.

[1] Inoltre, sono state utilizzate funzioni Lambda già implementate e consultabili nella documentazione ufficiale di Amazon, impiegate per la gestione della logica dell'interfaccia del bot.

5.3.3 Configurazione del chatbot

In questa sezione vengono descritte le scelte effettuate durante la configurazione del bot, denominato `Study Plan Assistant`. Di seguito viene riportata la descrizione associata al bot, finalizzata a definirne lo scopo e a fornire un'indicazione generale delle funzionalità previste: *The bot's purpose is to assist students in creating their study plan for the Computer Engineering and Automation program. It should guide them in choosing the elective courses offered by the Polytechnic University of Marche, both for the bachelor's and master's degrees.* Il timeout della sessione è stato impostato a 15 minuti, per garantire un intervallo di tempo sufficientemente ampio per l'interazione con l'utente, stimando una durata media di una conversazione. La lingua scelta per l'implementazione è l'inglese, in quanto consente di sfruttare alcune funzionalità aggiuntive che risultano non disponibili, ad esempio, per la lingua italiana. Successivamente, si potrà valutare l'integrazione di altre lingue, con particolare riferimento all'italiano, per rendere il servizio più accessibile agli studenti che non sono di madrelingua inglese. Anche la scelta della regione di distribuzione risulta rilevante, poiché alcune opzioni sono disponibili esclusivamente in determinate aree geografiche. Pertanto, è stata mantenuta la regione predefinita *Stati Uniti, Virginia settentrionale (us-east-1)*. Inoltre, è stata valutata la possibilità di integrare, in seguito, l'interazione mediante input vocale. Per il livello di confidenza è stato impostato un valore pari a 0.65, con l'obiettivo di evitare che richieste correttamente interpretabili vengano scartate a causa di una soglia eccessivamente restrittiva. Nelle sezioni successive verranno analizzati anche i risultati ottenuti modificando tale parametro. Inoltre, è stata inizialmente abilitata l'opzione di risoluzione assistita degli slot. Ulteriori specifiche sulle modalità di svolgimento dei test relativi a tale funzionalità e sui risultati ottenuti sono descritti nella Sezione 5.3.9.

5.3.4 Flusso di conversazione

Prima di procedere alla realizzazione del chatbot, è stato definito il flusso della conversazione che esso avrebbe dovuto seguire, mostrato in Figura 5.3.

La conversazione ha inizio chiedendo all'utente se sia interessato a ricevere informazioni relative a uno specifico corso oppure se desideri assistenza nella costruzione di uno study plan. Successivamente, viene richiesto di specificare se l'interesse riguarda il percorso di laurea magistrale o quello di laurea triennale e, in base alla scelta effettuata, viene richiesto l'anno di riferimento. A seconda delle risposte fornite dall'utente, viene richiamato uno dei seguenti intenti:

- *Specific Course*, finalizzato a fornire informazioni relative a uno specifico corso. Al termine della consultazione, viene domandato all'utente se sia interessato a procedere con la costruzione di uno study plan oppure se necessiti di ulteriori informazioni.
- *Study Plan*, finalizzato all'organizzazione del processo di costruzione dello study plan. In base all'anno selezionato, il bot procede con la relativa gestione e, successivamente, verifica se l'utente sia interessato a ricevere informazioni relative ad anni accademici precedenti o successivi. In caso contrario, la conversazione può essere conclusa.

In particolare, qualora l'utente scelga di procedere con la costruzione di uno study plan per uno specifico anno accademico, il bot attiva l'intento dedicato alla sua gestione. Innanzitutto, vengono fornite le informazioni riguardanti i corsi disponibili per l'anno selezionato; successivamente, viene richiesto all'utente se sia interessato ad approfondire uno specifico corso. In caso affermativo, viene richiesto quale corso desideri approfondire; altrimenti, viene verificato se l'utente sia interessato a ricevere una proposta di study plan basata sui corsi già selezionati. Qualora l'utente manifesti interesse per uno study plan personalizzato, il flusso

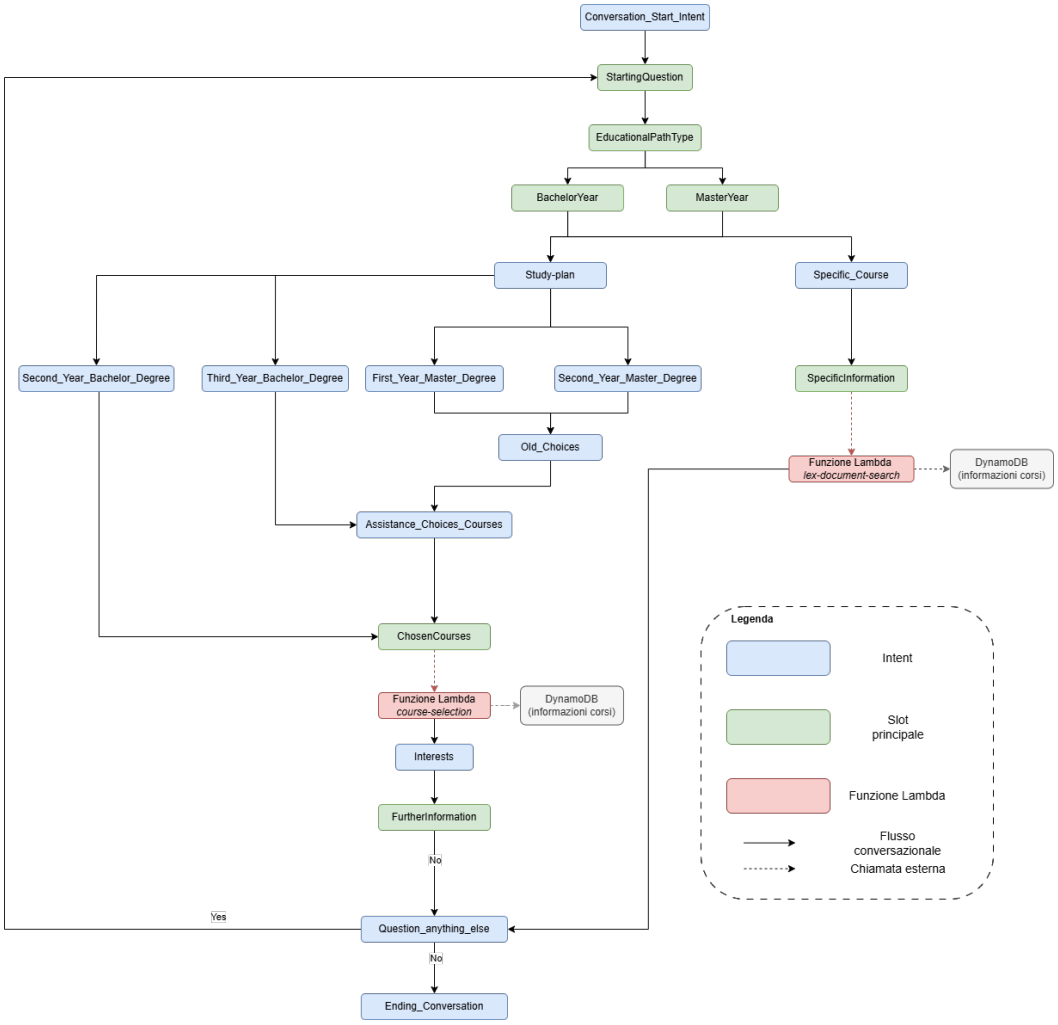


Figura 5.3: Flusso logico principale del chatbot

conversazionale viene ulteriormente articolato mediante una serie di domande finalizzate a raccogliere informazioni sulle sue preferenze. In particolare, nel caso in cui venga espresso interesse per anni accademici precedenti, viene richiesto se sia necessaria assistenza anche relativamente a tali anni. In caso di risposta negativa, la conversazione prosegue verificando se l'utente desideri ricevere ulteriori informazioni. Qualora la risposta sia affermativa, il flusso riparte dalla domanda iniziale; altrimenti, la conversazione viene terminata. Al fine di fornire informazioni maggiormente dettagliate, il bot richiede all'utente di indicare se il proprio interesse sia maggiormente orientato verso l'informatica, verso l'automazione oppure verso entrambe le aree. Quest'ultima possibilità comporta una maggiore ampiezza dell'insieme di corsi considerati e, di conseguenza, può condurre a una proposta meno esaudiente. Per quest'ultimo scenario, le specifiche già raccolte vengono ulteriormente integrate mediante una domanda relativa agli argomenti di maggiore interesse dell'utente, correlati ai corsi disponibili. Sulla base delle preferenze espresse, il bot elabora una possibile configurazione dello study plan. Infine, lo scambio viene concluso, domandando all'utente se abbia un eventuale interesse a ricevere informazioni relative agli anni accademici successivi. Nella Tabella 5.1 sono raggruppati tutti gli intenti realizzati per il chatbot, che verranno ulteriormente approfonditi nella Sezione 5.3.5.

Gli intenti sono stati sviluppati con l'obiettivo di gestire nel modo più efficace possibile le interazioni con l'utente, cercando di soddisfare le principali richieste che possono emergere

Categoria	Intent	Descrizione
Gestione della conversazione	*Conversation_Start_Intent*	Gestisce l'inizio della conversazione, visualizzando un messaggio di benvenuto e chiedendo all'utente se necessita di assistenza per informazioni su un corso specifico o per la creazione di un piano di studi.
	Question_Anything_Else	Verifica se l'utente necessita di ulteriore assistenza.
	Ending_Conversation	Gestisce la conclusione della conversazione, ponendo fine al flusso di interazione con l'utente.
	Help_Intent	Intent integrato dedicato alla gestione delle richieste di aiuto formulate dall'utente.
	FallbackIntent	Intent integrato richiamato quando il bot non riesce a comprendere o interpretare correttamente la richiesta dell'utente.
Informazioni sui corsi	*Specific_Course*	Fornisce informazioni relative a uno specifico corso di studio.
	Second_Year_Bachelor_Degree	Gestisce le richieste di informazioni relative ai corsi del secondo anno della laurea triennale.
	Third_Year_Bachelor_Degree	Fornisce informazioni relative ai corsi del terzo anno della laurea triennale.
	First_Year_Master_Degree	Fornisce informazioni relative ai corsi del primo anno della laurea magistrale.
	Second_Year_Master_Degree	Gestisce le richieste di informazioni relative ai corsi del secondo anno della laurea magistrale.
Costruzione dello study plan	*Study_Plan*	Instrada il flusso della conversazione verso la costruzione di un possibile piano di studi (*study plan*).
	Old_Choices_First_Master	Fornisce assistenza nella costruzione dello *study plan* relativo al primo anno della laurea magistrale, tenendo conto delle scelte effettuate in precedenza.
	Old_Choices_Second_Master	Fornisce assistenza nella costruzione dello *study plan* relativo al secondo anno della laurea magistrale, tenendo conto delle scelte effettuate in precedenza.
	Assistance_Choices_Courses	Verifica se l'utente necessita di ulteriore assistenza nella scelta dei corsi relativi a uno specifico anno di studio.
Personalizzazione	*Interests*	Raccoglie gli interessi dell'utente al fine di fornire suggerimenti e indicazioni maggiormente pertinenti.

Tabella 5.1: Principali intent del bot organizzati per funzione nel flusso conversazionale

durante la conversazione. Nella Tabella 5.2, prima di osservare il comportamento dei singoli intenti, vengono riportati tutti gli slot di tipo personalizzato definiti.

Sono stati utilizzati anche slot di tipo integrato, come *AMAZON.Confirmation* e *AMAZON.AlphaNumeric*.

5.3.5 Gestione degli intent

Di seguito vengono illustrati gli intenti sviluppati per il chatbot, descrivendone le modalità di gestione delle diverse interazioni e il contributo al flusso conversazionale complessivo. Per gli intent dedicati a uno specifico anno accademico, nella sezione *Risposta*

Categoria	Slot	Descrizione
Gestione del flusso	*Starting_Question*	Consente di selezionare il tipo di assistenza richiesta, distinguendo tra informazioni relative a un corso specifico e costruzione dello study plan.
	Educational_Path_Type	Permette di definire il percorso di studi di interesse, distinguendo tra laurea triennale e laurea magistrale.
	Degree_Programme_Year	Specifica l'anno di corso di interesse, ovvero primo, secondo o terzo anno.
Selezione dei corsi	*Second_Year_Bachelor_Degree_Courses*	Contiene l'elenco dei corsi offerti per il secondo anno della laurea triennale.
	Third_Year_Bachelor_Degree_Courses	Contiene l'elenco dei corsi offerti per il terzo anno della laurea triennale.
	First_Year_Master_Degree_Courses	Contiene l'elenco dei corsi offerti per il primo anno della laurea magistrale.
	Second_Year_Master_Degree_Courses	Contiene l'elenco dei corsi offerti per il secondo anno della laurea magistrale.
Interessi e ambiti di studio	*Branch*	Contiene l'elenco dei principali settori di interesse, distinguendo tra informatica e automazione.
	SpecificBranch	Contiene l'elenco dei possibili settori di interesse specifici relativi al percorso magistrale, quali Cybersecurity e Intelligenza Artificiale.
	TopicsBachelor	Contiene l'elenco dei possibili argomenti di interesse relativi al percorso di laurea triennale.
	TopicsMaster	Contiene l'elenco dei possibili argomenti di interesse relativi al percorso di laurea magistrale.

Tabella 5.2: Principali slot del bot organizzati per funzione

iniziale sono state inserite le informazioni relative ai corsi previsti per l'anno considerato. Inoltre, una volta soddisfatto l'intent, il flusso conversazionale prosegue attraverso l'intent *Question_Anything_Else*.

Conversation_Start_Intent

Il primo intento è stato progettato con l'obiettivo di avviare la conversazione con l'utente. Dopo aver mostrato un messaggio di benvenuto, il bot richiede all'utente se desideri ricevere informazioni relative a uno specifico corso oppure assistenza nella costruzione del proprio piano di studi. Tale informazione viene acquisita mediante lo slot personalizzato *Starting_Question*, opportunamente configurato, aggiungendo nella sezione *Affermazioni di esempio* alcune frasi che l'utente potrebbe utilizzare per fornire i valori dello slot, ad esempio *I want to receive help for a specific course* e *I want to receive help for my study plan*. Successivamente, il bot richiede di specificare il percorso di studio di interesse, distinguendo tra laurea triennale e laurea magistrale, mediante lo slot *EducationalPathType*. A seguito della corretta acquisizione dello slot, gestita mediante la voce *Acquisizione slot: risposta al successo*, il flusso conversazionale prosegue attraverso una *Ramificazione condizionale*, che determina il passo successivo in funzione del valore acquisito, come mostrato in Figura 5.4. Tale modalità di gestione è stata adottata frequentemente anche negli intenti successivi.

Per la gestione delle ramificazioni condizionali (*Default flow*) e delle *Risposte al fallimento* è stata generalmente adottata una logica comune all'interno del chatbot. In particolare,

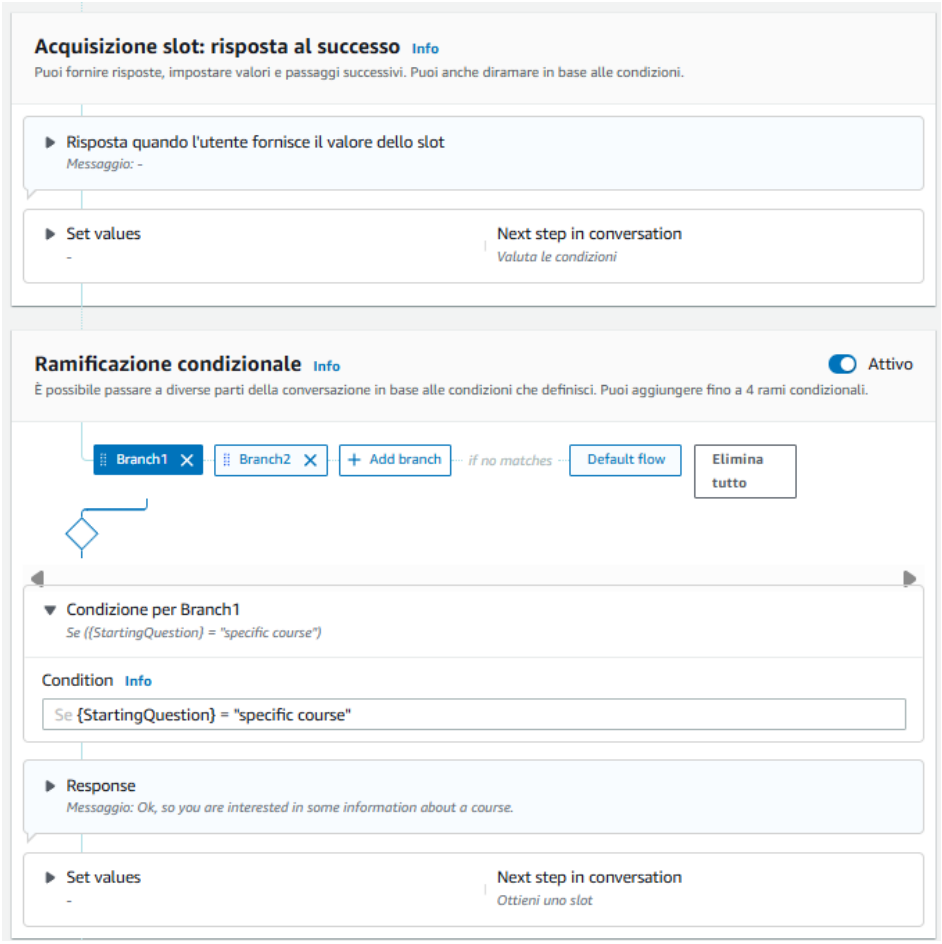


Figura 5.4: Sezione *Acquisizione slot: risposta al successo* con abilitazione della *Ramificazione condizionale*

qualora nessuna delle condizioni previste venga soddisfatta, il flusso viene indirizzato verso il *FallbackIntent*. Infine, in base al percorso selezionato, viene richiesto l'anno di interesse. A tale scopo, sono stati definiti due slot distinti, *BachelorYear* e *MasterYear*, rispettivamente per il percorso triennale e magistrale. La separazione dei due slot consente di controllare in maniera più precisa il branching.

Question_anything_else

Questo intento è stato progettato per gestire le situazioni in cui l'utente ha raggiunto un punto della conversazione nel quale non sono presenti ulteriori informazioni da acquisire. In tale circostanza, il bot richiede all'utente se desideri ricevere ulteriore assistenza o concludere la conversazione. Per la gestione della risposta è stato utilizzato lo slot integrato *AMAZON.Confirmation*[2].

Ending_Conversation

L'intento, richiamato dall'intento *Question_Anything_Else*, ha lo scopo di gestire la conclusione della comunicazione. Per tale motivo, è stata abilitata la *Risposta di chiusura*, mediante la quale il bot restituisce un messaggio finale all'utente e termina il dialogo.

[2]Si tratta di uno slot di tipo integrato per la gestione di risposte affermative o negative, quali *yes* e *no*.

Specific_Course

Se l'utente, nell'intento *Conversation_Start_Intent* dichiara di voler ricevere informazioni relative a un corso specifico, viene indirizzato all'internt *Specific_Course*, ideato per fornire informazioni riassuntive relative al corso selezionato. Per tale motivo, sono stati definiti quattro slot personalizzati, distinti sulla base del percorso accademico e dell'anno di erogazione del corso, ovvero *SecondYearBachelorDegree*, *ThirdYearBachelorDegree*, *FirstYearMasterDegree* e *SecondYearMasterDegree*. Gli slot includono esclusivamente i corsi a scelta. Inoltre, gli slot sono stati configurati come obbligatori per l'intento, in modo da garantire l'acquisizione delle informazioni necessarie alla corretta gestione delle ramificazioni presenti.

A seguito dell'acquisizione di uno degli slot relativi ai corsi, il flusso prosegue verso lo slot *SpecificInformation*, di tipo *AMAZON.AlphaNumeric*[3], mediante il quale viene domandato quali informazioni relative al corso siano di interesse. In assenza di una corrispondenza valida, il flusso viene indirizzato verso il *FallbackIntent*. Considerato che, in caso di risposta positiva, gli slot conducono all'intent *SpecificInformation*, il bot avrebbe potuto essere costruito prevedendo un unico slot di tipo *Courses*, contenente l'insieme dei corsi disponibili. La scelta di utilizzare slot distinti consente di mantenere un maggiore controllo sulla logica conversazionale, facilitando, inoltre, eventuali estensioni future nel caso di un aumento del numero di corsi o di una maggiore suddivisione tra laurea triennale e magistrale. Tale criterio di progettazione è stato mantenuto anche negli intenti successivi in cui risulta necessario distinguere i due percorsi accademici. A ciò si aggiungono la limitazione delle possibili ramificazioni associate ai singoli slot e l'esigenza di preservare il significato proprio di *Intent*, che sarebbe stato compromesso da un ulteriore accorpamento della logica.

Una volta acquisito lo slot, il flusso viene indirizzato a *Evadi intento*, che consente di proseguire con la sezione *Fulfillment*. La gestione della richiesta prosegue mediante l'invocazione, da parte di Amazon Lex, della funzione Lambda `handler`, la cui implementazione è descritta nella Sezione 5.3.6. In caso di esecuzione corretta, il bot comunica l'adempimento dell'operazione e procede alla *Risposta di chiusura*. Successivamente, una *Ramificazione condizionale*, basata sui valori acquisiti negli slot precedenti, determina il successivo percorso conversazionale. Per consentire tale gestione, i valori degli slot rilevanti vengono memorizzati tra gli attributi di sessione mediante la sezione *Set values*, come mostrato in Figura 5.5. In caso di errore durante il fulfillment o di timeout della sessione, il flusso viene indirizzato al *FallbackIntent*.

Study_plan

Questo intent viene richiamato nel caso in cui l'utente abbia espresso interesse per suggerimenti relativi alla composizione del proprio study plan. Il suo scopo è gestire la fase generale del dialogo finalizzata a fornire all'utente indicazioni in merito a possibili combinazioni dei corsi a scelta. Analogamente al caso precedente, la sezione relativa alla *Risposta iniziale* è stata lasciata vuota. A tale scopo, sono stati inseriti alcuni slot finalizzati a verificare se l'utente sia interessato a conoscere possibili combinazioni di corsi sulla base degli insegnamenti già selezionati negli anni precedenti, quali *StudyPlanBachelorDegree*, *StudyPlanMasterDegree*, *PossibleBachelorStudyPlan* e *PossibleMasterStudyPlan*. Tramite le opportune ramificazioni condizionali associate a ciascuno slot, vengono gestite le diverse logiche necessarie a restituire le informazioni richieste e a proseguire il dialogo adeguatamente.

[3]Il tipo di slot *AMAZON.AlphaNumeric* permette all'utente di fornire un valore contenente testo, numeri e caratteri speciali.

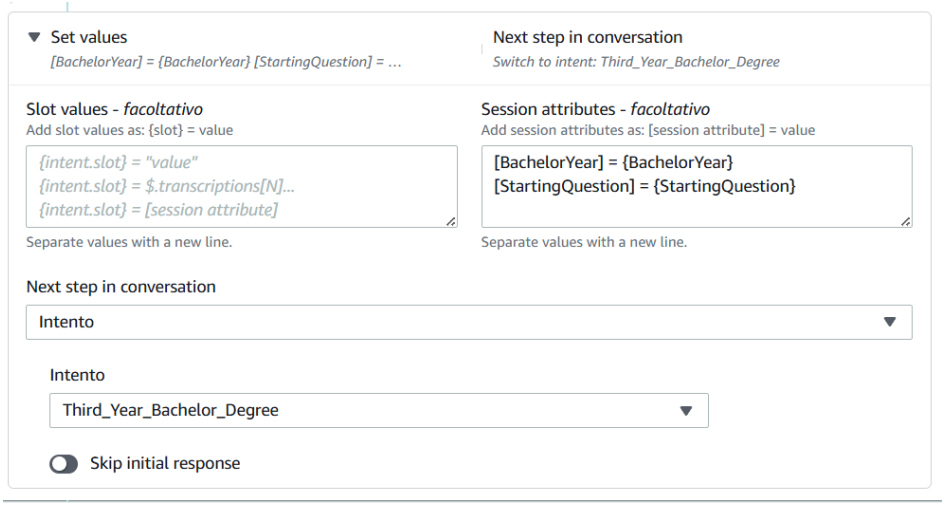


Figura 5.5: Schermata per la configurazione dei *Session attributes*

Second_Year_Bachelor_Degree

Questo intent ha lo scopo di gestire le informazioni relative alle raccomandazioni per la composizione dello study plan per il secondo anno del corso di laurea triennale. Il flusso prosegue attraverso l'intent *Study_Plan*, in particolare mediante lo slot *StudyPlanBachelorDegree*, che verifica l'interesse dell'utente verso uno specifico corso. In base alla risposta, il dialogo viene indirizzato verso l'intent *Specific_Course* oppure verso lo slot *PossibleBachelorStudyPlan*, che verifica se l'utente intenda comporre lo study plan sulla base dei corsi già scelti o che intende scegliere per il secondo anno. Da tale slot, il flusso prosegue verso *ChosenCourses* o *Interests*, a seconda delle preferenze espresse dall'utente, come mostrato in Figura 5.6.

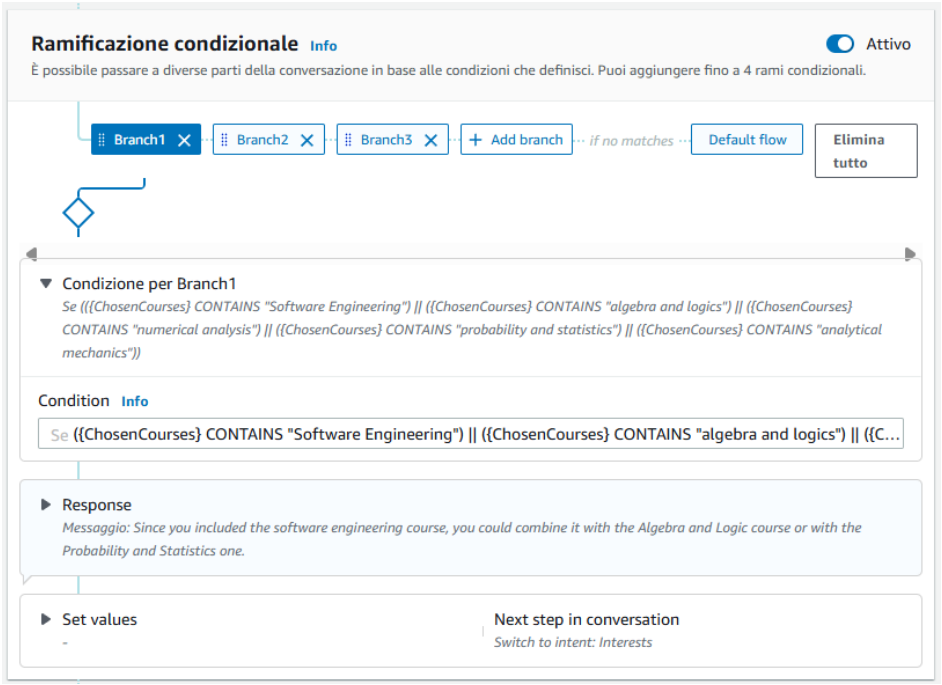


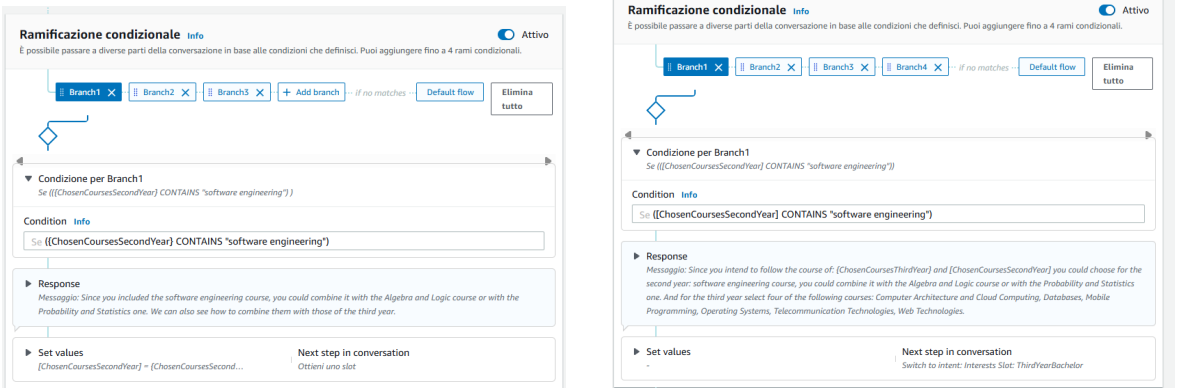
Figura 5.6: Ramificazione condizionale *ChosenCourses*

Nel caso in cui l'utente manifesti interesse sia per l'automazione sia per l'informatica, esso viene indirizzato verso lo slot *TopicsInterest*, che consente di specificare gli argomenti

di interesse relativi ai corsi del secondo anno. Per tale slot è stata abilitata l'opzione *Slot multivalore*, in modo da permettere la selezione di più valori. Sulla base della risposta dell'utente vengono, quindi, restituiti i suggerimenti pertinenti mediante opportune ramificazioni. Infine, la conversazione prosegue attraverso lo slot *FurtherInformation*, che verifica l'eventuale interesse dell'utente a ricevere informazioni anche sul terzo anno della laurea triennale. In caso affermativo, il flusso viene indirizzato all'intent *Third_Year_Bachelor_Degree*, impostando i *Session attribute* `[BachelorYear] = "third year"` e `[EducationalPathType] = "bachelor degree"`, necessari per il corretto instradamento degli intent successivi.

Third_Year_Bachelor_Degree

L'intent ha lo scopo di supportare l'utente nella composizione dello study plan per il terzo anno del corso di laurea triennale, tenendo conto anche dei corsi già selezionati, o che l'utente intende selezionare, per il secondo anno. Il flusso logico riprende in larga parte quello dell'intent relativo al secondo anno. In aggiunta, mediante lo slot *AssistanceSecondYear*, l'utente può essere indirizzato all'intent *Second_Year_Bachelor_Degree* qualora necessiti di informazioni sui corsi del secondo anno. Le successive ramificazioni gestiscono la raccolta dei corsi già scelti o da scegliere per il secondo anno e di quelli di interesse per il terzo anno, tramite lo slot *ChosenCoursesThirdYear*, fornendo informazioni pertinenti sulla base delle risposte dell'utente, come mostrato in Figura 5.7.



(a) Verso lo slot *ChosenCoursesSecondYear* **(b)** Verso lo slot *ChosenCoursesThirdYear*

Figura 5.7: Ramificazione condizionale relativa all'intent *Third Year Bachelor Degree*

Infine, lo slot *FurtherInformation* consente di proseguire la conversazione verso *First_Year_Master_Degree*, nel caso in cui l'utente sia interessato a ricevere informazioni sul primo anno del corso di laurea magistrale. Nel primo caso, sono stati impostati i *Session attributes* `[EducationalPathType] = "master degree"` e `[MasterYear] = "first year"`, necessari per il corretto instradamento degli intent successivi.

First_Year_Master_Degree

Questo intent ha lo scopo di gestire le informazioni legate al primo anno del percorso magistrale. L'intento segue la logica applicata negli intenti precedentemente descritti. In aggiunta, lo slot *PossibleCoursesBachelor* domanda all'utente se sia interessato a ricevere suggerimenti relativi a possibili combinazioni di corsi, anche sulla base di quelli già scelti o che intende selezionare per il percorso di laurea triennale. In caso affermativo, il flusso viene indirizzato all'intent *Old_Choices_First_Master*, nello specifico allo slot *PossibleSecondYearCourses*; in alternativa, prosegue verso lo slot *ChosenCoursesFirstYear* dell'intent corrente, attraverso il

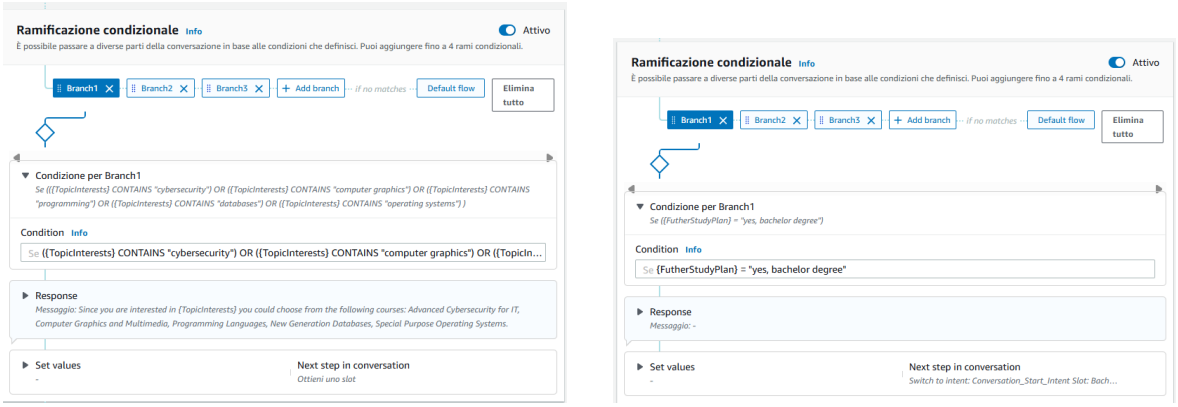
quale vengono raccolte le preferenze dell'utente relative ai corsi del primo anno della magistrale. Sulla base dei corsi indicati vengono proposte possibili soluzioni e, successivamente, il flusso viene instradato verso lo slot *FirstYearMaster* dell'intent *Interests*. Eventualmente, lo slot *TopicsInterests* consente all'utente di specificare gli argomenti di proprio interesse. Per tale scopo, sono stati definiti appositi enunciati, ad esempio *I'm interested in {TopicsInterests}*, così da consentire il riconoscimento dei valori direttamente a partire dalle espressioni dell'utente. In conclusione, lo slot *FurtherInformation* verifica l'eventuale interesse dell'utente a proseguire la conversazione per ricevere dettagli relativi al secondo anno della laurea magistrale. Qualora la risposta sia positiva, il flusso viene indirizzato all'intent *Second_Year_Master_Degree*, impostando il *Session attribute* `[MasterYear] = "second year"`.

Old_Choices_First_Master

Questo intent è stato sviluppato come supporto all'intent *First_Year_Master_Degree*, con lo scopo di raccogliere ulteriori informazioni necessarie per formulare raccomandazioni relative alla composizione dello study plan del primo anno della laurea magistrale. La sezione *Risposta iniziale* è stata lasciata vuota. Il flusso inizia dallo slot *PossibleSecondYearCourses*, che verifica se l'utente abbia già selezionato i corsi del secondo anno della triennale. Nel caso di risposta affermativa, la conversazione prosegue verso lo slot *PossibleThirdYearCourses*, mentre, in caso opposto, viene attivato lo slot *AssistanceSecondYear*, che consente, qualora necessario, di indirizzare l'utente verso l'intent *Second_Year_Bachelor_Degree* per ottenere le informazioni relative a tale anno. Analogamente, la gestione del terzo anno può prevedere il ricorso all'intent *Third_Year_Bachelor_Degree* nel caso in cui l'utente necessiti di supporto nella scelta dei relativi corsi. Una volta acquisiti i corsi del secondo e del terzo anno effettivamente selezionati dall'utente mediante gli slot *ChosenCoursesSecondYear* e *ChosenCoursesThirdYear*, il flusso viene ricondotto all'intent *First_Year_Master_Degree*, nello specifico allo slot *ChosenCoursesFirstYear*, attraverso il quale prosegue la raccolta delle preferenze relative al primo anno della laurea magistrale. In questo modo, le informazioni relative agli insegnamenti selezionati durante la triennale possono essere utilizzate nelle successive fasi di elaborazione delle raccomandazioni.

Second_Year_Master_Degree

Questo intent è stato sviluppato per gestire la composizione dello study plan relativo al secondo anno del corso di laurea magistrale. Il flusso riprende la logica dell'intent dedicato al primo anno del percorso magistrale, con la gestione dei corsi di interesse dell'utente tramite lo slot *ChosenCoursesSecondYearMaster*. Successivamente, lo slot *PossibleFirstYear* verifica se l'utente desidera comporre lo study plan anche sulla base dei corsi già scelti per il primo anno della magistrale. Nel caso in cui l'utente sia interessato, il bot richiama l'intent *Old_Choices_Second_Master*, che gestisce tali informazioni secondo una logica analoga a quella illustrata per *Old_Choices_First_Master*. Sulla base delle scelte effettuate, vengono forniti consigli, cercando di raffinare la combinazione dei vari corsi; qualora necessario, il flusso prosegue verso la gestione degli interessi tramite lo slot *TopicInterests*. Le opportune ramificazioni consentono di fornire suggerimenti coerenti con gli argomenti indicati, come mostrato in Figura 5.8a. Infine, lo slot *FutherStudyPlan* verifica se l'utente desidera proseguire la conversazione con informazioni relative al percorso di laurea triennale o al primo anno della laurea magistrale. In base alla risposta, il flusso viene indirizzato verso il percorso corrispondente, come mostrato in Figura 5.8b.



(a) Verso lo slot *Topicsinterests* **(b)** Verso lo slot *FutherStudyPlan*

Figura 5.8: Ramificazione condizionale relativa all'intent *Second Year Master Degree*

Old_Choices_Second_Master

L'obiettivo di questo intent è integrare il flusso di conversazione dell'intent *Second_Year_Master_Degree*, semplificandone la gestione. In particolare, l'intent consente di recuperare le informazioni relative ai corsi già selezionati dall'utente negli anni precedenti, così da poterle utilizzare nella composizione dello study plan del secondo anno della laurea magistrale. Il flusso parte dallo slot *PossibleCoursesFirstYear*, che domanda all'utente se abbia già selezionato i corsi del primo anno della magistrale. In tal caso, il flusso prosegue verso lo slot *ChosenCourseFirstYear*; altrimenti, viene attivato lo slot *AssistanceMaster* dell'intent *Assistance_Choices_Courses*, che consente di fornire il supporto necessario alla scelta dei corsi. Successivamente, lo slot *PossibleBachelorCourses* verifica se sia necessario considerare anche i corsi selezionati durante il percorso di laurea triennale. In questa eventualità, il flusso prosegue attraverso gli slot *PossibleCoursesSecondYear* e *PossibleCoursesThirdYear*, che verificano, rispettivamente, la selezione dei corsi del secondo e del terzo anno della triennale. In base alle risposte, vengono acquisiti i corsi tramite gli slot *ChosenCoursesSecondYearBachelor* e *ChosenCoursesThirdYear*, oppure viene fornito il supporto necessario mediante i relativi slot dell'intent *Assistance_Choices_Courses*. Una volta raccolte tutte le informazioni necessarie, dallo slot *ChosenCoursesThirdYear*, il passo successivo è *Evadi intento*, che permette di richiamare la funzione Lambda *courses-selection*, i cui dettagli sono riportati nella Sezione 5.3.6. Quindi, è stato abilitato il *Fulfillment* dell'intent per permettere l'esecuzione della funzione e, al termine, la conversazione prosegue verso la *Risposta di chiusura*, anch'essa opportunamente configurata. In conclusione, il flusso viene indirizzato verso l'intent *Interests*, come mostrato in Figura 5.9.

Assistance_Choices_Courses

Questo intent è stato sviluppato per gestire le richieste di assistenza relative alla scelta dei corsi degli anni precedenti al secondo anno della laurea magistrale. In particolare, esso integra il flusso degli intent *Second_Year_Bachelor_Degree* e *Old_Choices_Second_Master*, semplificandone la struttura e la raccolta delle informazioni necessarie alla composizione dello study plan. Il flusso ha inizio dallo slot *AssistanceMaster* per verificare la necessità dell'utente di ricevere assistenza nella scelta dei corsi del primo anno della magistrale. In tal caso, il flusso viene indirizzato verso l'intent *First_Year_Master_Degree*; altrimenti, prosegue verso lo slot *PossibleBachelorCourses* dell'intent *Old_Choices_Second_Master*, sulla base delle informazioni già disponibili. Analogamente, lo slot *AssistanceSecondYearBachelor* verifica l'eventuale necessità di supporto per i corsi del secondo anno della triennale. Lo slot *AssistanceThirdYear* gestisce in

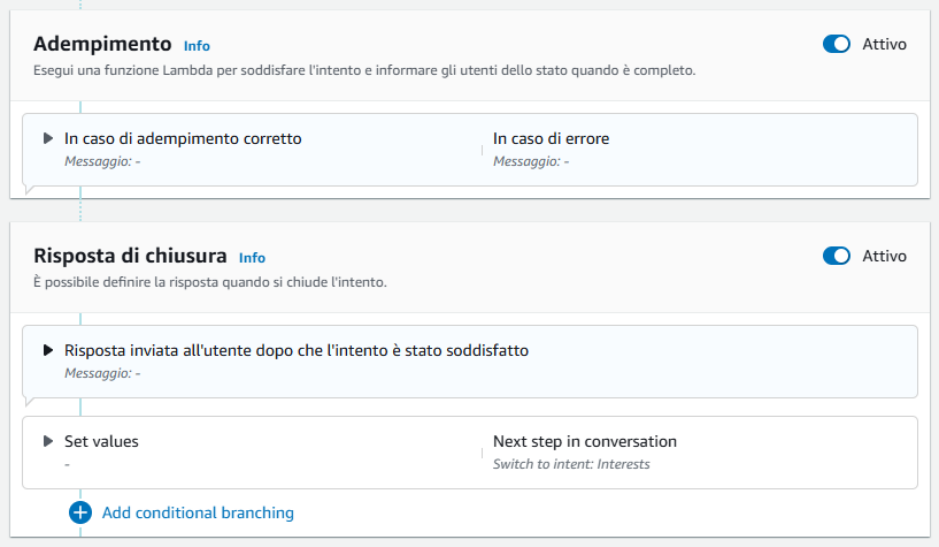


Figura 5.9: Abilitazione dei passaggi di conversazione *Adempimento* e *Riposta di chiusura*

maniera analoga la richiesta di assistenza relativa al terzo anno della triennale, indirizzando il flusso verso *Third_Year_Bachelor_Degree* oppure verso *Evadi intento*. In quest'ultimo caso, è stato abilitato il *Fulfillment* dell'intent per consentire il richiamo della funzione Lambda *courses-selection*. In seguito, la *Risposta di chiusura*, anch'essa opportunamente configurata, permette di proseguire la conversazione verso l'intent *Interests*.

Interests

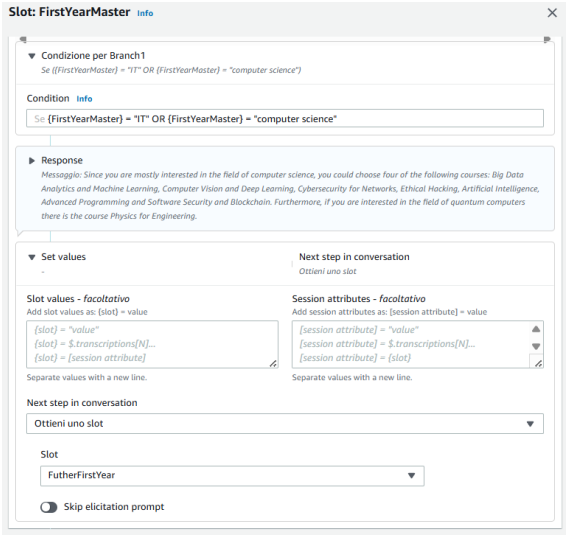
Lo scopo di questo intent è fornire ulteriori dettagli sui possibili corsi opzionali, anche sulla base degli interessi espressi dall'utente. Nella sezione *Risposta iniziale* è stato inserito un messaggio introduttivo alle domande successive. Infatti, l'intent è costituito da una serie di slot che consentono di individuare l'area di maggiore interesse dell'utente, distinguendo principalmente tra informatica e automazione. In base all'intent da cui viene richiamato, il flusso viene indirizzato verso le opportune ramificazioni consentendo anche di fornire risposte meno specifiche qualora l'utente non esprima una preferenza netta, come mostrato in Figura 5.10. Inoltre, sono stati definiti ulteriori slot per approfondire gli interessi dell'utente in merito al percorso magistrale, domandandogli se sia maggiormente incline agli argomenti relativi alla cybersecurity o all'IA.

Help_Intent

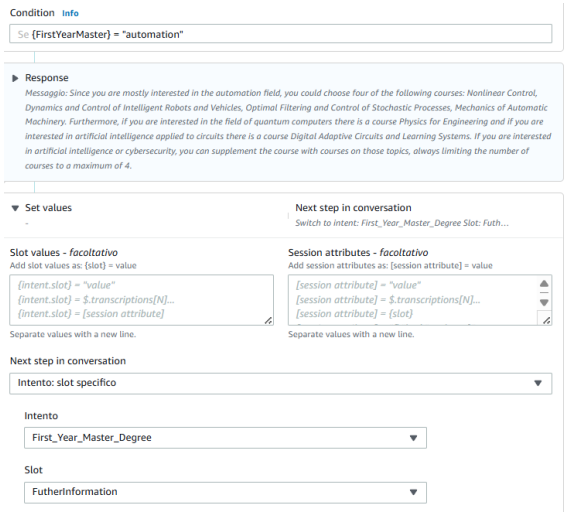
Questo intent è stato introdotto per gestire le richieste di assistenza formulate dall'utente. Per tale motivo, è stato utilizzato l'intent integrato *AMAZON.HelpIntent*, già predisposto per la gestione di questo tipo di richieste.

FallbackIntent

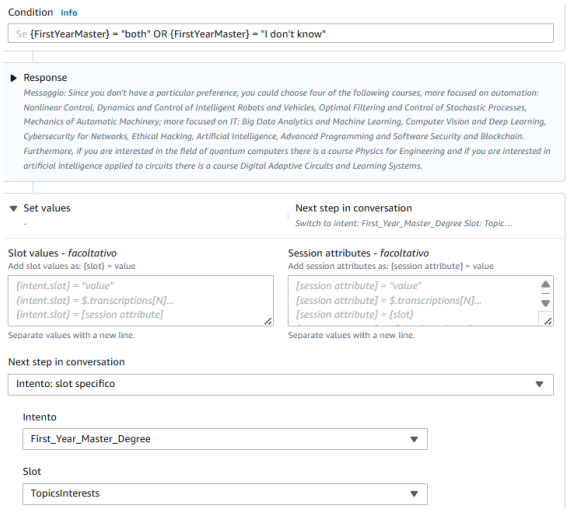
L'intent è già presente nella configurazione iniziale del bot e viene richiamato quando viene ricevuto un input non previsto dal sistema. Per adattarlo al contesto specifico del bot, è stata modificata la relativa *Risposta iniziale*, inserendo il messaggio *We're sorry, but we're unable to handle your request.* Una volta richiamato, l'intent restituisce il messaggio configurato e, successivamente, richiama l'intent *Question_Anything_Else*.



(a) Informatica



(b) Automazione



(c) Entrambi o l'utente è incerto

Figura 5.10: Ramificazione condizionale relativa a *FirstYearMaster*

5.3.6 Funzioni Lambda

Le funzioni Lambda implementate per la gestione delle informazioni più elaborate relative al chatbot sono illustrate di seguito, riportando esclusivamente le parti ritenute maggiormente significative. Per consentire l'accesso alle funzioni Lambda sviluppate e ad alcuni dei file JSON utilizzati per i test, è stato predisposto un repository GitHub accessibile, dalla quale è possibile reperire tutto il materiale necessario alla realizzazione del progetto: `https://github.com/LucreziaMulieri/Assistant_Study_Plan`. Per definire l'azione successiva all'esecuzione della funzione viene utilizzato il parametro `dialogAction`, strutturato in modo da contenere le informazioni relative alla fase successiva proposta da Amazon Lex. Per entrambe le funzioni, `lex-document-search` e `courses-selection`, il campo `type` è stato impostato sul valore `Close`[4], ritenuto più appropriato per la gestione della conversazione. Infatti, una volta completata l'esecuzione della funzione, l'obiettivo è

[4]Il valore `close` indica il completamento del processo di *Fulfillment* dell'intent e specifica che non è prevista un'ulteriore interazione con l'utente nell'ambito dello stesso flusso.

valutare le condizioni necessarie per determinare le successive ramificazioni condizionali, senza attendere un'ulteriore risposta da parte dell'utente. Inoltre, per entrambe le funzioni, nella sezione *Configurazione* della funzione Lambda, alla voce *Autorizzazioni*, sono state configurate le autorizzazioni necessarie per l'accesso alle risorse. In particolare, è stata creata una policy inline associata alla funzione, che ne consente l'utilizzo del servizio Amazon DynamoDB, oltre alle autorizzazioni necessarie per l'integrazione della funzione con il bot. Infine, è stata aggiunta la variabile d'ambiente `COURSES_TABLE`, il cui valore è impostato a `courses`, corrispondente al nome della tabella DynamoDB utilizzata dalla funzione.

Funzione handler

La prima funzione sviluppata è `handler`, il cui compito è richiamare dinamicamente le altre due funzioni nei punti appropriati della conversazione. Nello specifico, si è resa necessaria la possibilità di collegare al chatbot due funzioni Lambda distinte. Poiché Amazon Lex consente di associare una sola funzione Lambda al bot, è stata implementata tale funzione di routing, riportata nel Listato 5.1.

```
    .....

def lambda_handler(event, context):
    print(event)

    intent = (
        event.get("sessionState", {})
            .get("intent", {})
            .get("name", "")
    )

    if intent in ("Old_Choices_Second_Master", "Assistance_Choices_Courses"):
        return _invoke(SELECTED_COURSES, event)
    if intent == "Specific_Course":
        return _invoke(DOCUMENT_SEARCH, event)

    return {
        "sessionState": {"dialogAction": {"type": "ElicitIntent"}},
        "messages": [{
            "contentType": "PlainText",
            "content": (
                "Hello! Welcome to the help desk for computer and automation engineering courses, how can I
                    help you? "
                "Do you need information on a specific course or assistance in composing a study plan?"
            ),
        }],
    }
```

Listato 5.1: Funzione Lambda handler

Tale logica è stata implementata all'interno di `lambda_handler` mediante una struttura condizionale `if`, che consente di invocare la funzione dedicata alla gestione dello study plan in corrispondenza degli slot *Old_Choices_Second_Master* e *Assistance_Choices_Courses*. Per l'intent *Specific_Course*, invece, viene richiamata la funzione dedicata alla gestione delle informazioni relative al corso.

Funzione Lambda lex-document-search

La funzione riportata nel Listato 5.2 viene utilizzata per fornire all'utente informazioni aggiuntive relative a un determinato corso, specificato dall'utente. Ad esempio, la funzione consente di fornire dettagli relativi al programma didattico, ai crediti formativi e alla lingua di erogazione di un corso.

```
    ....

#client
logger.info("Module loading started")
dynamodb = boto3.resource("dynamodb")
logger.info("Boto3 client created")

#environment variables
COURSES_TABLE = os.getenv("COURSES_TABLE", "courses")

#dynamoDB table connection
table = dynamodb.Table(COURSES_TABLE)

.....

      # If the user asked for something specific
   if specific:
      specific_lower = normalize(specific)
      parts = []

      if any(w in specific_lower for w in ["exam", "test", "assessment", "evaluation"]):
         parts.append(f"Exam format: {item.get('exam_format', 'not available')}")
      if any(w in specific_lower for w in ["syllabus", "program"]):
         parts.append(f"Syllabus: {item.get('syllabus', 'not available')}")
      if any(w in specific_lower for w in ["topic", "topics", "arguments", "content", "subjects"]):
         topics = ", ".join(item.get("topics", [])) or "not specified"
         parts.append(f"Topics: {topics}")
      if any(w in specific_lower for w in ["professor", "teacher", "instructor", "lecturer"]):
         parts.append(f"Professor: {item.get('professor', 'not available')}")
      if any(w in specific_lower for w in ["prerequisite", "requirements", "required"]):
         parts.append(f"Prerequisites: {item.get('prerequisites', 'not available')}")
      if any(w in specific_lower for w in ["language", "taught in"]):
         parts.append(f"Language: {item.get('language', 'not available')}")
      if any(w in specific_lower for w in ["credit", "cfu", "ects"]):
         parts.append(f"Credits: {item.get('credits', 'n/a')}")

      if parts:
         return f"{display}:\n" + "\n".join(parts), True

   # Full response if no specific info requested
   topics = ", ".join(item.get("topics", [])) or "not specified"

   return (
      (
         f"{display} ({item.get('credits', '?')} credits)\n"
         f"Area: {item.get('area', 'n/a')}\n"
         f"Professor: {item.get('professor', 'n/a')}\n"
         f"Language: {item.get('language', 'n/a')}\n\n"
         f"Summary: {item.get('summary', 'n/a')}\n\n"
         f"Syllabus: {item.get('syllabus', 'n/a')}\n"
         f"Prerequisites: {item.get('prerequisites', 'n/a')}"
      ),
      True,
   )

    ....
```

Listato 5.2: Funzione Lambda per la gestione delle informazioni relative ad un corso specifico

Funzione Lambda courses-selection

La funzione dedicata all'individuazione di possibili abbinamenti tra i corsi è riportata nel Listato 5.3. In particolare, essa restituisce una lista di corsi affini alle preferenze espresse dall'utente, orientando la selezione sulla base del settore di interesse. Ad esempio, qualora le scelte effettuate evidenzino una maggiore propensione verso l'ambito dell'automazione, la funzione privilegia corsi appartenenti a tale settore rispetto a quelli afferenti all'ambito informatico.

Le variabili d'ambiente `BOT_ID` e `SLOT_TYPE_ID_*` sono utilizzate, rispettivamente, per indicare l'id del bot che richiama la funzione e per recuperare i valori contenuti nei diversi slot, che rappresentano i corsi opzionali previsti per ciascun anno.

5.3.7 Dynamo DB

Tramite Amazon DynamoDB è stato realizzato il database interrogato dalla funzione Lambda `lex-document-search`. Esso contiene le informazioni relative a tutti i corsi previsti per ciascun anno di entrambi i percorsi di studio, come mostrato in Figura 5.11. Per

```
......

# All course names stored in lowercase for consistent comparison and suggestion.
AREA_SUGGESTIONS = {
   "cybersecurity": {
      "master_year_1": ["cybersecurity for networks", "ethical hacking", "software security and blockchain
          "],
      "master_year_2": ["advanced cybersecurity for it"],
      "bachelor_year_3": [],
      "bachelor_year_2": [],
   },
   "ai_ml": {
      "master_year_1": ["big data analytics and machine learning",
                    "computer vision and deep learning", "artificial intelligence",
                    "digital adaptive circuits and learning systems"],
      "master_year_2": ["data science"],
      "bachelor_year_3": [],
      "bachelor_year_2": [],
   },
  .......
}

AREA_LABELS = {
   "cybersecurity": "Cybersecurity",
   "ai_ml":        "AI / Machine Learning",
   "automation":   "Automation and Control",
   "computer science": "Computer Science",
}
......

#build the response based on the slots
def _build_lines(slots, session, include_bachelor_y3: bool, include_master_y1: bool, include_bachelor_y2:
     bool) -> list:
   def get_session_courses(key: str) -> list:
      value = session.get(key, "")
      if not value:
         return []
      return [v.strip() for v in value.split(",") if v.strip()]

   m2 = get_session_courses("ChosenCoursesSecondYearMaster")
   m1 = get_session_courses("ChosenCourseFirstYear") if include_master_y1 else []
   b3 = get_session_courses("ChosenCoursesThirdYear") if include_bachelor_y3 else []
   b2 = get_session_courses("ChosenCoursesSecondYearBachelor") if include_bachelor_y2 else []

   all_chosen = m2 + m1 + b3 + b2
   areas     = detect_areas(all_chosen)
   chosen_lower = {normalize(c) for c in all_chosen}

   lines = []

   # Master year 2
   candidates = _get_candidates(areas, "master_year_2", chosen_lower)
   lines.append(
      f"Since you're interested in {format_list(m2)} for your second year of your master's degree, "
      f"you could choose from the following courses: {format_list(candidates)}."
   )
   .....

   return lines

.....
```

Listato 5.3: Funzione Lambda per fornire suggerimenti relativi ai corsi

ogni corso vengono memorizzati i seguenti attributi: identificativo del corso; nome del corso; area di interesse, ad esempio cybersecurity; numero di crediti formativi (CFU); modalità di svolgimento dell'esame; lingua di erogazione del corso; eventuali prerequisiti; docente responsabile dell'insegnamento; riassunto degli obiettivi formativi del corso; programma, con l'indicazione degli argomenti trattati durante il corso; lista degli argomenti principali; anno di riferimento, ad esempio il primo anno del corso di laurea magistrale.

La scelta di Amazon DynamoDB è motivata principalmente dalla sua agevole integrazione con gli altri servizi AWS e dalla natura del caso d'uso considerato. Infatti, il volume di dati da gestire è piuttosto contenuto, almeno allo stato attuale del bot, e l'accesso al database è previsto esclusivamente in lettura.

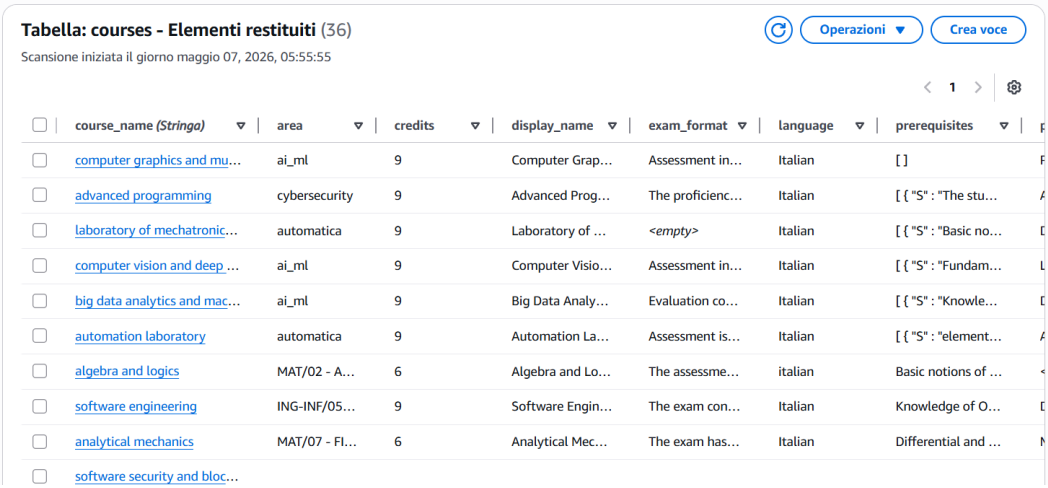


Tabella: courses - Elementi restituiti (36)

Scansione iniziata il giorno maggio 07, 2026, 05:55:55

Operazioni | Crea voce

course_name *(Stringa)*	area	credits	display_name	exam_format	language	prerequisites
computer graphics and mu...	ai_ml	9	Computer Grap...	Assessment in...	Italian	[ ]
advanced programming	cybersecurity	9	Advanced Prog...	The proficienc...	Italian	[{ "S" : "The stu...
laboratory of mechatronic...	automatica	9	Laboratory of ...	*<empty>*	Italian	[{ "S" : "Basic no...
computer vision and deep ...	ai_ml	9	Computer Visio...	Assessment in...	Italian	[{ "S" : "Fundam...
big data analytics and mac...	ai_ml	9	Big Data Analy...	Evaluation co...	Italian	[{ "S" : "Knowle...
automation laboratory	automatica	9	Automation La...	Assessment is...	Italian	[{ "S" : "element...
algebra and logics	MAT/02 - A...	6	Algebra and Lo...	The assessme...	italian	Basic notions of ...
software engineering	ING-INF/05...	9	Software Engin...	The exam con...	Italian	Knowledge of O...
analytical mechanics	MAT/07 - FI...	6	Analytical Mec...	The exam has...	Italian	Differential and ...
software security and bloc...						

Figura 5.11: Database NoSQL realizzato mediante Amazon DynamoDB

5.3.8 Integrazione del chatbot in un'applicazione web

Per rendere il chatbot fruibile dall'utente finale e consentirne l'eventuale integrazione all'interno di una web application esistente, è stata utilizzata un'interfaccia web già disponibile, successivamente personalizzata e adattata al bot Amazon Lex precedentemente realizzato. Inoltre, il deployment dell'architettura adottata fa riferimento alla soluzione proposta da AWS (fonte: `https://aws.amazon.com/it/blogs/machine-learning/deploy-a-web-ui-for-your-chatbot/`), mostrata in Figura 5.12.

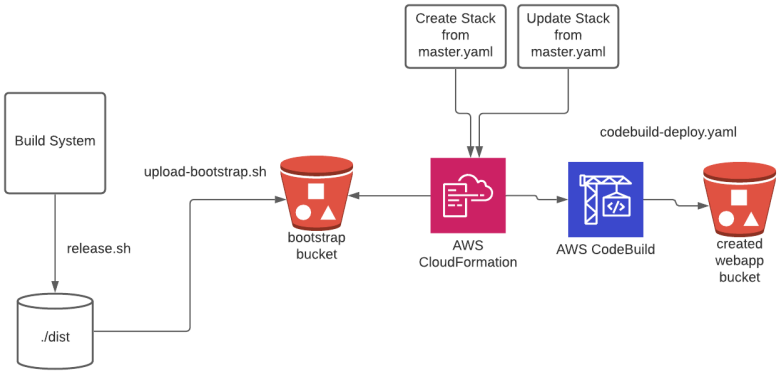


Figura 5.12: Schema dell'architettura adottata per l'interfaccia utente (fonte: `https://github.com/aws-samples/aws-lex-web-ui`)

Prima di procedere con la configurazione dell'interfaccia è necessario verificare alcuni prerequisiti relativi alle risorse AWS utilizzate. In particolare, è necessario disporre delle autorizzazioni necessarie per accedere ai servizi CloudFormation, Amazon S3, Amazon Cognito e Amazon Lex. Inoltre, è necessario disporre di un bot Amazon Lex V2 correttamente associato a un alias che punti a una versione del bot diversa dallo stato *DRAFT*. L'alias consente, infatti, di associare un identificativo stabile a una specifica versione del bot, rendendo possibile l'aggiornamento della configurazione selezionata da parte delle applicazioni client. A seguito di eventuali modifiche al bot, è necessario creare una nuova versione e associarla all'alias opportuno. Nel caso considerato, oltre all'alias predefinito, è stato creato l'alias *TestVersionBot*, associato alla versione del bot utilizzata dall'interfaccia web. Nella configurazione dell'alias è, inoltre, necessario specificare, per la lingua utilizzata dal bot, la funzione Lambda da

invocare durante le fasi di inizializzazione, convalida e fulfillment. Nel caso considerato, è stata associata la funzione Lambda `handler`, precedentemente descritta.

Per la configurazione e il deployment dell'interfaccia è stato utilizzato AWS CloudShell, dal quale sono stati eseguiti i comandi necessari. Una volta verificati i prerequisiti, il primo passaggio consiste nella creazione di un file JSON contenente i parametri necessari alla configurazione della Web UI. Il comando utilizzato è riportato nel Listato 5.4.

```
   cat > ~/params.json << 'ENDOFFILE'
[
 {"ParameterKey": "LexV2BotId",        "ParameterValue": "BOT_ID"},
 {"ParameterKey": "LexV2BotAliasId",   "ParameterValue": "ALIAS_ID"},
 {"ParameterKey": "LexV2BotLocaleId",  "ParameterValue": "en_US"},
 {"ParameterKey": "WebAppConfToolbarTitle", "ParameterValue": "Study Plan Assistant"},
 {"ParameterKey": "WebAppConfBotInitialText", "ParameterValue": "Hello! Welcome to the help desk for
      computer and automation engineering courses."},
 {"ParameterKey": "WebAppConfBotInitialUtterance", "ParameterValue": "hi"},
 {"ParameterKey": "TitleLogoImgUrl",   "ParameterValue": ""},
 {"ParameterKey": "BotAvatarImgUrl",   "ParameterValue": ""},
.....
]
ENDOFFILE
```

Listato 5.4: Comando per la creazione del file JSON contenente i parametri di configurazione dell'interfaccia del bot

Il parametro `LexV2BotId` viene ''valorizzato'' con l'id del bot.

Analogamente, `LexV2BotAliasId` contiene l'id dell'alias precedentemente creato. Il parametro `WebAppConfToolbarTitle` consente, invece, di specificare il nome visualizzato nell'interfaccia web.

Il parametro `WebAppConfBotInitialText` permette di definire il messaggio iniziale visualizzato dal chatbot all'apertura della conversazione. Per consentire al bot di avviare automaticamente il dialogo, senza richiedere un'interazione esplicita da parte dell'utente, è possibile utilizzare la funzionalità *PutSession* della console Amazon Lex. In particolare, il parametro `WebAppConfBotInitialUtterance` è stato impostato al valore *hi*, precedentemente definito tra le *sample utterances* dell'intent *Conversation_Start_Intent*. In tal modo, l'interfaccia invia automaticamente un'utterance iniziale a Lex, simulando la prima interazione dell'utente e consentendo l'avvio della conversazione. Quindi, il primo slot *StartingQuestion* dell'intent viene sollecitato senza che l'utente debba interrogare il bot. Gli ulteriori parametri presenti nel file consentono di personalizzare l'aspetto dell'interfaccia, ad esempio il colore dello sfondo e il font dei caratteri dei messaggi. Tramite il servizio *CloudFormation* è stato realizzato il deployment dell'interfaccia, utilizzando il template fornito dalla soluzione AWS. Tale template descrive le risorse necessarie alla realizzazione della Web UI e ne gestisce automaticamente la creazione, rispettando le dipendenze esistenti tra le diverse risorse. In particolare, il deployment genera le seguenti risorse:

- un *bucket S3*[5], utilizzato per ospitare i file della Web UI, inclusi HTML, JavaScript e CSS;
- una *distribuzione Amazon CloudFront*, che consente di rendere accessibile l'interfaccia tramite un URL;
- un *Amazon Cognito Identity Pool*, utilizzato per gestire l'autenticazione degli utenti verso Amazon Lex;
- un *progetto AWS CodeBuild*, utilizzato per la compilazione e il deployment dei file della web application;
- i *ruoli IAM* necessari per consentire la corretta comunicazione tra le risorse coinvolte.

[5]Per ulteriori informazioni su Amazon S3 si veda la Sezione 4.2.2.

Al completamento del deployment, Amazon CloudFront rende disponibile un URL attraverso il quale è possibile accedere all'interfaccia web del chatbot, mostrata in Figura 5.13

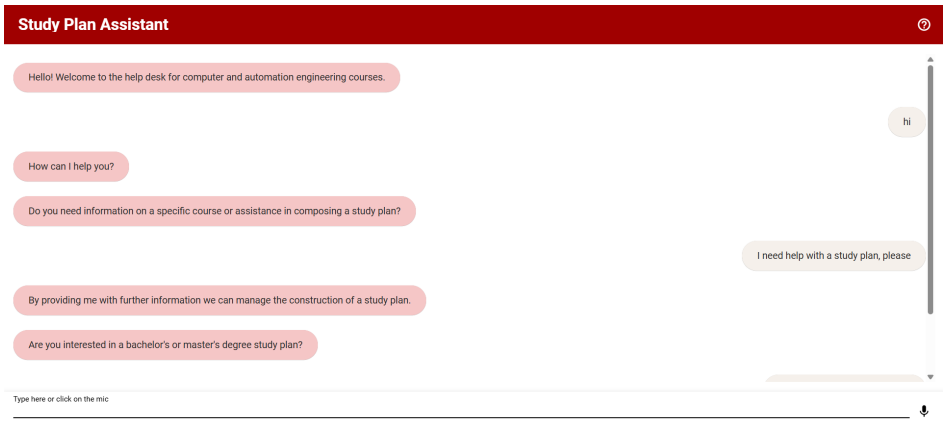


Figura 5.13: Interfaccia web del chatbot

In questo modo, l'interfaccia realizzata consente all'utente di interagire con il chatbot attraverso un ambiente web semplice e intuitivo. In aggiunta, la soluzione può essere ulteriormente personalizzata e integrata all'interno di un'applicazione web esistente, adattandola alle esigenze del sito ospitante.

5.3.9 Verifica e valutazione del chatbot

In questa sezione vengono illustrati i test effettuati per comprovare il corretto funzionamento del chatbot e delle funzioni Lambda sviluppate. Per la valutazione del comportamento del chatbot è stata sfruttata anche la funzionalità *Analisi* disponibile nella console Amazon Lex, che consente di monitorare una serie di parametri tra cui il tasso di riconoscimento degli enunciati, le prestazioni degli intent e gli intent maggiormente utilizzati o classificati come *failed*. Prima di verificarne il comportamento all'interno del flusso conversazionale, le funzioni Lambda sono state sottoposte a test individuali mediante specifici eventi di input.

Test del chatbot

Durante lo sviluppo del chatbot, sono stati effettuati diversi *Build* mediante la console Amazon Lex. La sezione *Test* della console ha, inoltre, consentito di interagire direttamente con il chatbot, verificando il corretto svolgimento del flusso logico della conversazione. Un supporto fondamentale alla fase di test è stato fornito da Amazon CloudWatch. Nella sezione *Gestione dei log*, esso consente di accedere ai gruppi di log associati alle funzioni Lambda e di analizzarne l'esecuzione in corrispondenza dei diversi eventi. Di conseguenza, i log hanno consentito di individuare più facilmente eventuali errori, nello specifico quelli relativi alla corretta comunicazione tra il chatbot e le funzioni Lambda nonché quelli relativi al corretto comportamento della funzione `handler` nell'invocazione di `lex-document-search` e `course-selection`. L'obiettivo era verificare che le due funzioni venissero richiamate nel punto giusto della conversazione, senza generare eventi anomali o chiamate improvvise. Inoltre, è stato possibile monitorare i valori trasferiti dal chatbot alle funzioni Lambda e viceversa. Sono state, quindi, effettuate diverse prove per valutare la capacità del chatbot di adempiere alle richieste dell'utente. In una prima fase, è stata abilitata esclusivamente la funzionalità di risoluzione assistita degli slot, per verificarne l'efficacia nell'acquisizione delle specifiche richieste. Su un campione di circa 100 conversazioni, comprendente anche le interazioni effettuate durante lo sviluppo del chatbot, la dashboard di Amazon Lex ha

restituito i risultati riportati nelle Tabelle 5.3, 5.4 e 5.5. In diverse occasioni è stato osservato un ritorno improvviso all'intent di avvio della conversazione. Gli intent che prevedono l'invocazione delle funzioni Lambda hanno evidenziato il maggiore tasso di fallimento, mostrando alcune difficoltà iniziali nel coordinamento tra essi e le funzioni Lambda coinvolte, come mostrato in Figura 5.14. Nel complesso, i risultati evidenziano una certa efficacia della risoluzione assistita degli slot nell'acquisizione dei valori relativi agli slot integrati.

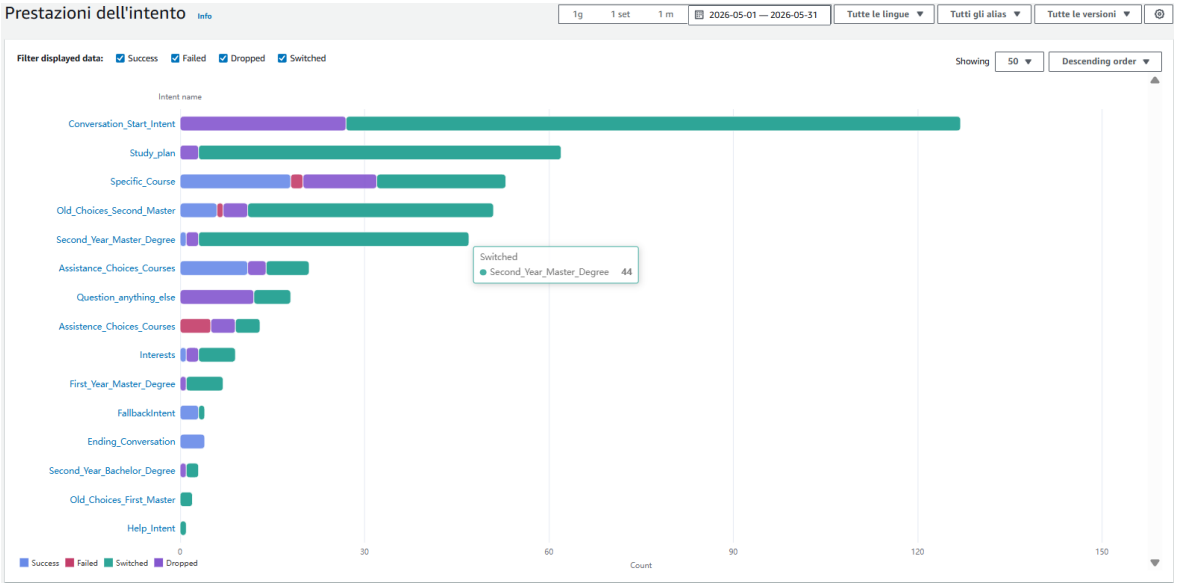


Figura 5.14: Grafico delle prestazioni dell'intento

Successivamente, sono state effettuate ulteriori prove utilizzando la funzionalità di comprensione del linguaggio assistita, che ha evidenziato un miglioramento nel riconoscimento degli intent e, di conseguenza, nella qualità del flusso conversazionale, come mostrato nelle Tabelle 5.3, 5.4 e 5.5. Nonostante il miglioramento osservato, il chatbot ha mostrato una certa dipendenza dagli enunciati definiti nella configurazione degli intent. In alcuni casi, infatti, un enunciato non previsto ha determinato il ritorno all'intent di avvio della conversazione o al *FallbackIntent*; in altri, è stata nuovamente richiesta l'immissione di uno slot.

Sono state, infine, effettuate ulteriori prove impostando la soglia di confidenza a 0,85, al fine di osservare eventuali variazioni nel comportamento del chatbot. Come evidenziato dai valori riportati nelle Tabelle 5.3, 5.4 e 5.5, rispetto alle configurazioni precedenti, in questa impostazione si osserva una maggiore percentuale di conversazioni completate con esito *success*, accompagnata da una riduzione dei casi classificati come *dropped*[6]. A livello degli intent, si rileva, inoltre, un aumento della percentuale di *switched*[7] e una riduzione dei casi *dropped*, mentre non sono stati registrati casi *failed*. Tali variazioni suggeriscono un comportamento differente del chatbot al variare della soglia di confidenza, senza, tuttavia, consentire di attribuire univocamente a tale parametro un miglioramento della qualità della conversazione, anche considerando la diversa dimensione dei campioni analizzati. L'aumento della percentuale di *switched*, osservato anche nella configurazione con comprensione del linguaggio assistita, non costituisce di per sé un indicatore di miglioramento, ma può essere interpretato come un comportamento maggiormente dinamico nel riconoscimento degli intent. Nell'ultimo caso, inoltre, non è stata rilevata alcuna ricaduta nel *FallbackIntent*.

[6]Il termine *dropped* indica una conversazione nella quale l'utente non fornisce una risposta prima che questa possa essere classificata come *success* o *failed*.

[7]Con *switched* si indica il caso in cui il chatbot riconosce un intent differente da quello attualmente attivo e vi trasferisce la gestione della conversazione, prima che l'intent possa essere classificato come *success* o *failed*.

Categoria	Parametro	Risoluzione assistita degli slot	Comprensione del linguaggio assistita	Soglia di confidenza 0,85
Campione di test	Numero di conversazioni	≈ 100	≈ 50	≈ 15
Ripartizione delle prestazioni delle conversazioni	*Success*	33%	59%	80%
	Dropped	60%	41%	20%
	Failed	8%	–	–

Tabella 5.3: Prestazioni delle conversazioni nelle diverse configurazioni di test

Categoria	Parametro	Risoluzione assistita degli slot	Comprensione del linguaggio assistita	Soglia di confidenza 0,85
Campione di test	Numero di intent analizzati	≈ 400	≈ 300	75
Ripartizione delle prestazioni degli intent	*Switched*	40%	84,29%	86,67%
	Success	10,43%	9,37%	10,67%
	Dropped	16,82%	6,34%	2,67%
	Failed	1,90%	–	–
Intent di fallback	*FallbackIntent*	$\approx 0,95\%$	$\approx 1,80\%$	0%
Avvio della conversazione	Intent di avvio	$\approx 30\%$	$\approx 18,43\%$	–

Tabella 5.4: Prestazioni degli intent nelle diverse configurazioni di test

Categoria	Parametro	Risoluzione assistita degli slot	Comprensione del linguaggio assistita	Soglia di confidenza 0,85
Campione di test	Numero di utterance analizzate	> 1000	> 500	≈ 100
Riconoscimento degli enunciati	*Riconosciuti*	99%	99%	100%
	Non riconosciuti	1%	1%	0%

Tabella 5.5: Riconoscimento degli enunciati nelle diverse configurazioni di test

In aggiunta, è stato valutato il comportamento del timeout di sessione, impostato a 15 minuti. Tale valore si è dimostrato adeguato a mantenere le informazioni raccolte durante una sessione anche in presenza di conversazioni articolate. Un valore inferiore avrebbe potuto determinare la scadenza prematura della sessione, mentre un valore eccessivamente elevato avrebbe potuto aumentare il rischio di sovrapposizione tra gli attributi associati a sessioni successive.

Test della funzione lex-document-search

La funzione `lex-document-search` è stata testata mediante la creazione di specifici test event[8]. Un esempio di evento utilizzato durante i test è riportato nel Listato 5.5.

[8]I test event sono oggetti JSON utilizzati per simulare le richieste inviate dal servizio AWS associato alla funzione, in questo caso Amazon Lex.

```
{
  "sessionState": {
    "intent": {
      "name": "Specific_Course",
      "slots": {
        "SecondYearMasterDegree": {
          "value": {
            "interpretedValue": "data science",
            "resolvedValues": ["data science"]
          }
        },
        "FirstYearMasterDegree": null,
        "SecondYearBachelorDegree": null,
        "ThirdYearBachelorDegree": null,
        "SpecificInformationMaster": {
          "value": {
            "interpretedValue": "exam",
            "resolvedValues": ["exam"]
          }
        },
        "SpecificInformationBachelor": null
      }
    },
    "sessionAttributes": {
      "EducationalPathType": "master degree",
      "MasterYear": "second year",
      "BachelorYear": "",
      "StartingQuestion": "study plan"
    }
  }
}
```

Listato 5.5: Test della funzione `lex-document-search`

In particolare, l'evento simula una richiesta di informazioni aggiuntive relativa al corso di Data Science, nello specifico sulle modalità di svolgimento dell'esame. Dopo aver verificato il corretto comportamento della funzione, sono state effettuate delle controprove direttamente attraverso il chatbot, includendo anche richieste contenenti simultaneamente diverse tipologie di informazioni relative a uno stesso corso, quali, ad esempio, il docente, il syllabus e i crediti formativi. Anche in questi casi, sono stati ottenuti gli output attesi.

Test della funzione course-selection

In modo analogo è stata testata la funzione `course-selection`, mediante la creazione di un test event volto a simulare differenti combinazioni di preferenze espresse dall'utente. Sono stati inseriti intenzionalmente più corsi per anno, allo scopo di verificare se la funzione sia effettivamente in grado di gestire correttamente più scelte da parte dell'utente. A seguito del superamento dei test su diversi scenari d'uso, che hanno generato gli output previsti, il bot finale è stato testato non solo tramite la console Amazon Lex, bensì simulando una sessione di comunicazione attraverso l'interfaccia web. Questa procedura ha permesso di analizzare l'esperienza d'uso dal punto di vista dell'utente finale. Nel complesso, l'interazione con il chatbot è risultata scorrevole e non ha evidenziato particolari limitazioni o comportamenti inattesi. La soluzione, tuttavia, non è stata progettata per consentire all'utente di formulare richieste completamente libere, ma per guidarlo attraverso il flusso conversazionale definito. Tale scelta progettuale è riflessa anche nella fase iniziale della conversazione, nella quale è il chatbot ad avviare l'interazione, chiedendo all'utente come possa assisterlo.

CAPITOLO 6

Discussione

In questo capitolo viene proposta una discussione in merito ai case study affrontati nel corso della trattazione. In particolare, vengono analizzati i risultati ottenuti e già descritti nei rispettivi capitoli, effettuando un confronto più dettagliato tra i risultati ottenuti nelle singole prove, con particolare riferimento ai case study dedicati, rispettivamente, al riconoscimento delle immagini e al riconoscimento vocale automatico. Infine, vengono riportate le osservazioni emerse dall'ultimo caso di studio, relativo alla realizzazione di un chatbot dedicato all'assistenza nella selezione dei corsi a scelta dell'Università Politecnica delle Marche.

6.1 Discussione in merito ai case study

Di seguito viene presentata una discussione relativa ai casi di studio illustrati nella presente trattazione. Per quelli inerenti al riconoscimento delle immagini e al riconoscimento vocale automatico viene fornito un confronto tra le diverse prove eseguite, mediante l'utilizzo di tabelle nelle quali sono riassunte le osservazioni emerse dall'analisi, riportate nelle rispettive Sezioni 6.1.1 e 6.1.2. In particolare, una delle tabelle riporta le caratteristiche relative, rispettivamente, alle immagini e alle tracce audio utilizzate nelle prove, mentre l'altra sintetizza i risultati ottenuti dall'analisi e le relative osservazioni. Il caso di studio relativo alla realizzazione di un chatbot mediante il servizio Amazon Lex si differenzia, invece, dai precedenti. Infatti, esso ha previsto non solo la valutazione del singolo servizio, ma anche la sua integrazione con altri servizi dell'ecosistema AWS.

6.1.1 Discussione in merito al riconoscimento di immagini

Dall'analisi dei casi di studio relativi al riconoscimento delle immagini, incentrati sulla rilevazione del testo mediante il servizio Amazon Rekognition, sono emersi diversi aspetti significativi in merito alle capacità e ai limiti dello strumento. Nei primi casi di studio sono state selezionate immagini che, in base alle ipotesi formulate, non presentavano particolari elementi in grado di ostacolare il riconoscimento del testo (Tabelle 6.1 e 6.2). In tali condizioni, il servizio ha generalmente confermato le aspettative, mostrando un buon livello di accuratezza nell'individuazione degli elementi testuali, come mostrato nella Tabella 6.3. Un'unica immagine ha costituito un'eccezione, poiché una parola non è stata rilevata integralmente. Tale risultato potrebbe essere riconducibile alla dimensione del testo o alle caratteristiche del font utilizzato, che potrebbero averne reso più complessa l'identificazione.

Un aspetto particolarmente significativo emerso durante l'analisi riguarda proprio la relazione tra le dimensioni degli elementi testuali e l'accuratezza del riconoscimento. Contrariamente a quanto inizialmente ipotizzato, gli elementi di dimensioni maggiori non sono risultati necessariamente più semplici da individuare rispetto a quelli di dimension inferiori. In alcuni casi, infatti, il riconoscimento degli elementi di dimensioni maggiori è risultato meno accurato. Una possibile spiegazione è che tali elementi vengano interpretati dal sistema come parte dello sfondo o come componenti grafiche dell'immagine, rendendone più difficile l'identificazione come testo.

Dai casi di studio è, inoltre, emerso che Amazon Rekognition presenta alcune limitazioni legate al modo in cui il testo viene individuato all'interno dell'immagine. In particolare, il comportamento osservato suggerisce che il sistema si concentri principalmente sull'identificazione dei caratteri e sulla loro aggregazione in parole e righe sulla base della disposizione spaziale, senza che emerga, nei casi analizzati, una significativa interpretazione semantica del contenuto. Ciò può contribuire a spiegare alcune delle imprecisioni osservate, soprattutto quando la disposizione degli elementi testuali si discosta da quella convenzionale o quando il testo è inserito in un contesto grafico particolarmente articolato.

Il secondo gruppo di casi di studio ha riguardato immagini caratterizzate da testi generalmente leggibili, ma contraddistinti da alcune particolarità grafiche (Tabelle 6.1 e 6.2). Anche in questo caso il servizio ha mostrato complessivamente un buon livello di accuratezza, sebbene siano emerse alcune limitazioni, osservabili nella Tabella 6.3. Tra queste si possono individuare difficoltà nel riconoscimento di determinate disposizioni dei caratteri, in particolare quando uno stesso elemento grafico è condiviso da più parole, nonché problemi nell'elaborazione di testi caratterizzati da una qualità visiva inferiore. Al contrario, le dimensioni ridotte del testo e l'utilizzo di caratteri graficamente più particolari non hanno rappresentato, nei casi analizzati, un ostacolo significativo al riconoscimento.

Un'ulteriore parte dell'analisi è stata dedicata alle immagini contenenti segnaletica stradale, con il duplice obiettivo di valutare il comportamento del sistema in condizioni più vicine a contesti applicativi reali e di verificarne la capacità di riconoscere alfabeti differenti da quello latino. Le immagini considerate presentavano, in alcuni casi, condizioni potenzialmente sfavorevoli, quali riflessi, bagliori, scarsa luminosità e prospettive non ottimali. Nonostante ciò, tali caratteristiche non hanno generalmente impedito il corretto rilevamento del testo. Anche gli esempi contenenti alfabeti arabo e cirillico sono stati riconosciuti correttamente, mentre il sistema non è stato in grado di rilevare il testo appartenente a sistemi di scrittura logografici, come il cinese e il giapponese.

Per quanto riguarda le immagini contenenti insegne, invece, la prospettiva e le caratteristiche delle scritte luminose hanno costituito un ostacolo più significativo, come inizialmente ipotizzato. Un ulteriore fattore di difficoltà è risultato essere l'elevato numero di elementi testuali presenti contemporaneamente nell'immagine, soprattutto quando associati a una notevole densità di componenti grafiche. Il caso di studio relativo alla locandina del film *Inception* ha evidenziato in modo particolare questa problematica. La presenza ravvicinata di numerosi elementi ha prodotto errori sia nell'individuazione delle parole che nella loro corretta assegnazione alle rispettive righe.

Anche la presenza di caratteristiche grafiche non convenzionali ha evidenziato alcune limitazioni del servizio. Testi inseriti all'interno di pitture murali, caratteri disposti secondo un ordine inverso rispetto al normale senso di lettura o scritte composte da elementi grafici particolari, come la scritta *Happy Birthday* realizzata mediante palloncini, hanno posto maggiori difficoltà al sistema. Al contrario, il testo disposto secondo orientamenti differenti, come quello ruotato o verticale, è stato generalmente rilevato, sebbene in alcuni casi la disposizione delle parole all'interno della struttura testuale non sia stata ricostruita correttamente.

Un ulteriore elemento emerso dall'analisi riguarda il rapporto tra il livello di confidenza

restituito dal servizio e la correttezza effettiva del riconoscimento. Infatti, in alcuni casi, Amazon Rekognition ha associato livelli di confidenza elevati a elementi testuali riconosciuti in modo errato. Ciò evidenzia come il valore di confidenza non possa essere considerato, singolarmente, una garanzia di correttezza del risultato, ma debba essere interpretato congiuntamente al contenuto effettivamente rilevato e alla sua corrispondenza con gli elementi presenti nell'immagine.

In generale, i casi di studio hanno evidenziato come l'accuratezza di Amazon Rekognition dipenda non soltanto dalla presenza del testo, ma anche dalle sue caratteristiche grafiche e dalla relazione con gli altri elementi presenti nell'immagine. Distorsioni, dissolvenze, effetti grafici, disposizioni non convenzionali e un'elevata densità di contenuti possono compromettere sia il rilevamento dei caratteri sia la corretta ricostruzione della struttura testuale. Proprio per questo motivo, i casi di studio finali sono stati selezionati con l'obiettivo di sottoporre il servizio a condizioni progressivamente più complesse e di mettere in evidenza le limitazioni a cui sono ancora frequentemente soggetti i sistemi di riconoscimento del testo nelle immagini.

In conclusione, Amazon Rekognition si è dimostrato uno strumento complessivamente valido per il rilevamento del testo e potenzialmente utile per essere integrato all'interno di applicazioni che richiedono funzionalità di questo tipo. Tuttavia, i risultati ottenuti evidenziano che il servizio non è esente da errori, soprattutto in presenza di condizioni visive o grafiche particolari. Pertanto, il suo impiego richiede una valutazione attenta dei risultati restituiti, l'adozione di strategie e procedure aggiuntive per verificarne l'affidabilità e un utilizzo accurato dei parametri disponibili per gestire l'analisi da parte del servizio stesso.

Categoria	**Caso**	**Livello di complessità**	**Dimensione del testo**	**Numero di vocaboli**	**Contrasto testo/sfondo**
Testo chiaramente leggibile	CS1	Molto basso	Media	Basso	Molto elevato
	CS2	Molto basso	Media	Basso	Molto elevato
	CS3	Basso	Media e grande	Basso	Medio
Testo leggibile	CS4	Medio	Media e piccola	Alto	Medio
	CS5	Medio	Grande	Molto basso	Molto elevato
	CS6	Medio	Media	Alto	Medio
	CS7	Medio	Piccola e media	Medio	Medio
Cartelli stradali	CS8	Alto	Molto piccola, piccola e media	Medio	Basso
	CS9	Alto	Piccola e media	Basso	Elevato
	CS10	Alto	Media	Molto basso	Medio
	CS11	Alto	Piccola	Molto basso	Basso
Insegne	CS12	Molto alto	Piccola e media	Basso	Basso
	CS13	Alto	Piccola e media	Basso	Basso
Locandine di film	CS14	Molto alto	Molta piccola, piccola e media	Molto alto	Molto basso
Graffiti	CS15	Molto alto	Grande	Molto basso	Medio
	CS16	Molto alto	Grande	Molto basso	Medio
Aspetti particolari	CS17	Alto	Grande	Molto basso	Medio
	CS18	Alto	Media e piccola	Molto basso	Molto elevato
	CS19	Molto alto	Grande	Molto basso	Molto basso
	CS20	Molto alto	Grande, media e piccola	Alto	Elevato

Tabella 6.1: Caratteristiche delle immagini e livelli di complessità dei casi di test (prima parte)

Categoria	Caso	Elementi grafici	Elementi di punteggiatura	Alfabeto	Condizioni di illuminazione
Testo chiaramente leggibile	CS1	Nesssuno	Punti, virgolette, parantesi tonde e trattino	Latino	Molto favorevoli
	CS2	Nessuno	Punto e virgolette alte	Latino	Molto favorevoli
	CS3	Font	Nessuno	Latino	Favorevoli
Testo leggibile	CS4	Apici, cifre romane e cifre arabe	Virgole e punti	Latino	Medie
	CS5	Disposizione verticale dei caratteri e font	Nessuno	Latino	Molto favorevoli
	CS6	Cifre romane	Virgole e punti	Latino	Medie
	CS7	Font	Virgole e punti	Latino	Medie
Cartelli stradali	CS8	&	Nessuno	Latino	Medie
	CS9	Nessuno	Nessuno	Latino e arabo	Favorevoli
	CS10	Nessuno	Nessuno	Latino e cirillico	Sfavorevoli
	CS11	Nessuno	Nessuno	Latino e scrittura logografica	Medie
Insegne	CS12	Nessuno	Nessuno	Latino	Medie
	CS13	Font e insegna luminosa	Nessuno	Latino	Sfavorevoli
Locandine di film	CS14	Loghi	Virgole e punti	Latino	Medie
Graffiti	CS15	Pittura murale	Nessuno	Latino	Favorevoli
	CS16	Caratteri invertiti	Nessuno	Latino	Favorevoli
Aspetti particolari	CS17	Palloncini	Nessuno	Latino	Medie
	CS18	Scritta ruotata di 90° e 180°	Nessuno	Latino	Molto favorevoli
	CS19	Distorsione e dissolvenza	Nessuno	Latino	Molto favorevoli
	CS20	Scritte ruotate di 90° e 180° e scritta contenente ulteriori elementi	Nessuno	Latino	Molto favorevoli

Tabella 6.2: Caratteristiche delle immagini e livelli di complessità dei casi di test (seconda parte)

6.1.2 Discussione in merito al riconoscimento vocale automatico

Come per i casi di studio relativi all'analisi del riconoscimento del testo presente all'interno delle immagini mediante Amazon Rekognition, in questa sezione viene riportata una panoramica delle osservazioni emerse dai casi di studio riguardanti il riconoscimento vocale automatico mediante Amazon Transcribe. I primi casi di studio sono stati realizzati principalmente con l'obiettivo di comprendere le capacità del servizio in condizioni di facile identificazione del parlato (Tabelle 6.4 e 6.5). In tali circostanze, Amazon Transcribe ha soddisfatto le aspettative, dimostrandosi in grado di riconoscere correttamente i vocaboli presenti all'interno delle tracce audio. Nel complesso, il servizio si è, quindi, mostrato piuttosto preciso nella trascrizione del parlato, soprattutto in presenza di un singolo interlocutore, di condizioni acustiche favorevoli e di una pronuncia sufficientemente chiara, come mostrato nella Tabella 6.6.

Successivamente, la complessità delle tracce è stata progressivamente aumentata, introducendo condizioni maggiormente rappresentative di situazioni reali, con l'obiettivo di

Categoria	Caso	Rilevamento testo	Accuratezza della trascrizione
Testo chiaramente leggibile	CS1	Completo	Molto elevata
	CS2	Completo	Molto elevata
	CS3	Parziale	Media
Testo leggibile	CS4	Parziale	Media
	CS5	Parziale	Bassa
	CS6	Parziale	Media
	CS7	Completa	Molto elevata
Cartelli stradali	CS8	Completa	Elevata
	CS9	Parziale	Media
	CS10	Completa	Bassa
	CS11	Parziale	Molto bassa
Insegne	CS12	Parziale	Bassa
	CS13	Parziale	Media
Locandine di film	CS14	Parziale	Media
Graffiti	CS15	Assente	–
	CS16	Completa	Bassa
Aspetti particolari	CS17	Parziale	Bassa
	CS18	Completa	Elevata
	CS19	Assente	–
	CS20	Parziale	Media

Tabella 6.3: Risultati di rilevamento e trascrizione del testo per i casi di test

osservare il comportamento del servizio e l'eventuale presenza di problematiche generalmente associate a questa tipologia di sistemi (Tabelle 6.4 e 6.5). In questo caso, sono emerse alcune delle limitazioni già discusse nella Sezione 4.1.2.

Una delle principali criticità osservate riguarda il riconoscimento e la suddivisione dei parlanti. Infatti, la modalità di diarizzazione si è dimostrata sufficientemente efficace nei casi in cui gli interlocutori risultavano chiaramente distinti, mentre ha evidenziato maggiori difficoltà in presenza di sovrapposizioni del parlato. In tali circostanze, alcune porzioni pronunciate da un interlocutore sono state associate a un altro, producendo una suddivisione errata del contenuto e rendendo la trascrizione ottenuta frammentata.

Nei casi caratterizzati da lievi sovrapposizioni o dalla presenza simultanea di più parlanti, il servizio ha, inoltre, mostrato la tendenza ad associare il contenuto a un unico interlocutore, non riuscendo a distinguere correttamente le diverse voci.

Nei casi in cui all'interno della stessa traccia sono contenute più lingue, il riconoscimento automatico delle lingue ha mostrato alcune difficoltà, in particolare, in corrispondenza dei cambi di lingua. Inoltre, la rilevazione della lingua non è risultata sempre corretta. Ad esempio, una traccia in lingua inglese con pronuncia indiana è stata identificata come inglese con pronuncia britannica. Analogamente, nel caso di parlato inglese caratterizzato da un forte accento russo, il servizio ha rilevato la presenza della lingua russa anziché dell'inglese. Questi risultati evidenziano come la presenza di accenti particolarmente marcati possa rappresentare un ulteriore elemento di difficoltà per il riconoscimento automatico della lingua.

Anche l'utilizzo di un vocabolario personalizzato, introdotto con l'obiettivo di migliorare il riconoscimento di termini appartenenti a linguaggi settoriali, non ha prodotto i risultati attesi. Nei casi di studio in cui tale funzionalità è stata utilizzata, il vocabolario personalizzato non ha apportato benefici significativi oppure, in alcune porzioni della trascrizione, ha persino costituito un ulteriore elemento di difficoltà per la corretta identificazione dei vocaboli.

Per quanto riguarda altre caratteristiche della traccia audio, il servizio ha mostrato una

buona capacità di gestione del rumore di sottofondo. Anche in presenza di registrazioni caratterizzate da una qualità audio ridotta e dalla presenza di rumori ambientali, la maggior parte delle parole pronunciate è stata riconosciuta correttamente. La gestione del rumore non ha, tuttavia, impedito il verificarsi di errori, soprattutto per quanto riguarda la corretta suddivisione dei parlanti.

Un comportamento differente è stato osservato nel caso di una traccia contenente una canzone. Nelle porzioni caratterizzate dalla presenza di parlato accompagnato da musica di sottofondo, il contenuto è stato generalmente riconosciuto, mentre nelle parti propriamente cantate sono stati rilevati pochi vocaboli, alcuni dei quali trascritti in maniera errata.

Anche la variazione della velocità della traccia audio non ha prodotto differenze rilevanti nei risultati ottenuti. Sia nel caso della traccia rallentata sia in quello della traccia velocizzata, la trascrizione è risultata identica a quella ottenuta dalla registrazione riprodotta alla velocità originale. Analogamente, non è stata osservata una significativa influenza della durata dell'audio sulla correttezza complessiva della trascrizione.

La formattazione del testo prodotto è risultata, nel complesso, coerente. In particolare, la punteggiatura è stata generalmente introdotta in modo appropriato, consentendo di restituire il contenuto del parlato in una forma sufficientemente leggibile. Inoltre, è stato osservato l'utilizzo delle lettere maiuscole per l'identificazione di determinati nomi propri. Tuttavia, in alcuni casi, la punteggiatura è risultata meno coerente, soprattutto nelle situazioni caratterizzate da una maggiore complessità del parlato.

Un ulteriore aspetto osservato riguarda la presenza, all'interno delle trascrizioni, di esitazioni, incertezze e ripetizioni presenti nel parlato originale. Tali elementi non vengono sistematicamente eliminati dal risultato prodotto dal servizio. Questo comportamento, insieme alla presenza di alcuni errori nei casi più complessi, suggerisce che il risultato della trascrizione rappresenti in larga misura il contenuto così come viene riconosciuto dal sistema, senza un'importante fase di rielaborazione successiva finalizzata alla correzione o alla normalizzazione del testo.

Le maggiori difficoltà sono state osservate, soprattutto, nei casi in cui erano presenti contemporaneamente più fattori di complessità, quali più parlanti, sovrapposizioni del parlato, passaggi da una lingua all'altra, accenti marcati e rumore di sottofondo, come evidenziato nella Tabella 6.6. Si tratta di condizioni che si avvicinano maggiormente a contesti reali di applicazione e nelle quali il servizio ha mostrato maggiori limitazioni. In tali situazioni, risulta, pertanto, opportuno prevedere una revisione manuale della trascrizione prodotta, soprattutto quando è richiesta un'elevata accuratezza del risultato finale.

Alcune funzionalità di Amazon Transcribe non sono state analizzate nel corso della presente trattazione. Il settore medico, ad esempio, non è stato approfondito, poiché Amazon mette a disposizione un servizio specifico, Amazon Transcribe Medical, le cui caratteristiche potrebbero essere oggetto di un'analisi separata. Similmente, la possibilità di creare un modello personalizzato non è stata sottoposta a valutazione, poiché l'obiettivo principale della sperimentazione era analizzare le capacità del servizio nelle sue condizioni standard, variando soltanto alcune delle impostazioni di trascrizione disponibili.

Nel complesso, Amazon Transcribe si è dimostrato un servizio sufficientemente accurato nella trascrizione del parlato e, pertanto, potenzialmente utile per essere integrato all'interno di sistemi che richiedono funzionalità di riconoscimento vocale automatico. Tuttavia, i casi di studio hanno evidenziato alcune limitazioni, in particolare nella diarizzazione e nella gestione di condizioni caratterizzate dalla sovrapposizione di più voci, da accenti marcati o dalla presenza simultanea di più fattori di complessità.

Per applicazioni nelle quali la corretta associazione tra parlante e contenuto trascritto assume particolare importanza, come nel caso della generazione automatica di sottotitoli per contenuti multimediali, il risultato prodotto dal servizio dovrebbe, quindi, essere sottoposto a

un'attenta revisione. Inoltre, l'effettiva qualità della trascrizione dipende dalle condizioni della traccia audio e dalle impostazioni utilizzate, rendendo necessario valutare con attenzione i diversi parametri offerti dal servizio in funzione del contesto applicativo.

Categoria	**Caso**	**Livello di complessità**	**Rumore di fondo**	**Sottofondo musicale**	**Numero parlanti**	**Lingua**
Trascrizione in condizioni standard	CS1	Molto basso	Assente	Assente	1	Inglese
	CS2	Basso	Assente	Assente	2	Inglese
	CS3	Molto basso	Molto basso	Assente	1	Italiano
Gestione contenuti multilingua	CS4	Medio	Assente	Assente	1	Alternanza italiano, inglese
	CS5	Medio	Assente	Assente	1	Multilingue (olandese, inglese, francese, italiano, spagnolo, portoghese, tedesco)
Analisi di condizioni acustiche particolari	CS6	Alto	Assente	Assente	5	Inglese
	CS7	Alto	Elevato	Assente	≈ 7	Inglese
	CS8	Medio	Presenza di effetti sonori	Basso	1	Inglese
	CS9	Medio	Presenza di effetti sonori	Basso	1	Inglese
Influenza accenti sulla trascrizione	CS10	Medio	Basso	Assente	≈ 5	Inglese
	CS11	Medio	Basso	Assente	2	Inglese
	CS12	Alto	Assente	Assente	1	Inglese
Gestione linguaggio specialistico	CS13	Alto	Assente	Assente	1	Inglese
	CS14	Medio	Assente	Assente	1	Inglese
Trascrizione contenuti musicali	CS15	Alto	Assente	Alto	1	Inglese
	CS16	Alto	Assente	Basso e alto	2	Inglese
Audio ad elevata complessità	CS17	Molto alto	Basso	Assente	5	Multilingue (italiano, spagnolo, inglese)
	CS18	Molto alto	Medio	Basso e alto	≈ 15	Inglese

Tabella 6.4: Caratteristiche del parlato e livelli di complessità dei casi di test (prima parte)

6.1.3 Discussione in merito alla realizzazione di un chatbot

Dalla realizzazione di tale progetto sono scaturite diverse osservazioni, anche in relazione all'integrazione delle funzionalità di IA fornite da AWS, finalizzate alla creazione di un flusso conversazionale maggiormente flessibile e in grado di soddisfare le specifiche richieste degli utenti. A causa dell'impossibilità di sfruttare pienamente le funzionalità integrabili con Amazon Lex mediante Amazon Bedrock, non è stato possibile valutare in modo esaustivo le opzioni basate sull'IA messe a disposizione dalla suite AWS. Pertanto, è stata adottata

Categoria	Caso	Velocità	Accenti e pronuncia	Linguaggio specialistico	Numero di parole	Volume del parlato
Trascrizione in condizioni standard	CS1	Normale	Assenza di particolare pronuncia	Assente	Basso	Medio
	CS2	Normale	Assenza di particolare pronuncia	Assente	Medio	Medio
	CS3	Normale	Assenza di particolare pronuncia	Assente	Medio	Medio
Gestione contenuti multilingua	CS4	Normale	Pronuncia inglese su lingua italiana	Assente	Medio	Medio
	CS5	Normale	Assenza di particolare pronuncia	Assente	Medio	Medio
Analisi di condizioni acustiche particolari	CS6	Normale	Assenza di particolare pronuncia	Assente	Alto	Alto, medio e basso
	CS7	Normale	Assenza di particolare pronuncia	Assente	Medio	Medio e basso
	CS8	Lento	Assenza di particolare pronuncia	Assente	Medio	Medio
	CS9	Veloce	Assenza di particolare pronuncia	Assente	Medio	Medio
Influenza accenti sulla trascrizione	CS10	Normale	Accento indiano	Assente	Medio	Medio
	CS11	Normale	Accento scozzese	Assente	Medio	Medio
	CS12	Normale	Accento russo	Assente	Medio	Medio
Gestione linguaggio specialistico	CS13	Normale	Assenza di particolare pronuncia	Linguaggio tecnico	Molto alto	Medio
	CS14	Normale	Assenza di particolare pronuncia	Linguaggio finanziario	Alto	Medio
Trascrizione contenuti musicali	CS15	Normale	Pronuncia propria del canto	Assente	Medio	Medio
	CS16	Normale	Pronuncia propria del canto	Assente	Medio	Medio
Audio ad elevata complessità	CS17	Normale	Accento spagnolo su lingua italiana e inglese	Assente	Molto alto	Medio e basso
	CS18	Normale	Assenza di particolare pronuncia	Assente	Molto alto	Alto, medio e basso

Tabella 6.5: Caratteristiche del parlato e livelli di complessità dei casi di test (seconda parte)

Categoria	Caso	Rilevamento del parlato	Accuratezza della trascrizione	Identificazione della lingua	Suddivisione dei parlanti
Trascrizione in condizioni standard	CS1	Completo	Molto alta	-	-
	CS2	Completo	Molto alta	-	Molto accurata
	CS3	Completo	Molto alta	-	-
Gestione contenuti multilingua	CS4	Parziale	Bassa	Accurata	-
	CS5	Completo	Media	Media	-
Analisi di condizioni acustiche particolari	CS6	Parziale	Media	Molto accurata	Molto bassa
	CS7	Completo	Media	Molto accurata	Molto bassa
	CS8	Completo	Molto alta	-	-
	CS8	Completo	Molto alta	-	-
Influenza accenti sulla trascrizione	CS9	Completo	Alta	Media	Bassa
	CS10	Completo	Media	Molto accurata	Bassa
	CS11	Parziale	Molto bassa	Molto bassa	-
Gestione linguaggio specialistico	CS12	Completo	Molto bassa	-	-
	CS13	Completo	Media	-	-
Trascrizione contenuti musicali	CS14	Parziale	Molto bassa	Accurata	-
	CS15	Parziale	Bassa	Molto accurata	-
Audio ad elevata complessità	CS16	Parziale	Media	Accurata	Molto bassa
	CS17	Completo	Media	-	Molto bassa

Tabella 6.6: Risultati di rilevamento e trascrizione del parlato per i casi di test

un'architettura più semplice, che ha, comunque, consentito di raggiungere l'obiettivo proposto, mostrando come il servizio possa essere utilizzato da aziende e organizzazioni per la realizzazione di bot conversazionali adattabili a diversi casi d'uso.

In particolare, l'abilitazione della funzionalità di comprensione del linguaggio naturale (NLU) assistita ha consentito di migliorare la comprensione degli intenti dell'utente rispetto alla mancata abilitazione di tale funzionalità. Tuttavia, il bot ha, spesso, dimostrato di essere fortemente vincolato agli enunciati dichiarati durante la configurazione degli intent, limitando la flessibilità nella gestione delle richieste formulate dall'utente.

Anche la funzionalità di risoluzione assistita degli slot si è rivelata efficace solo parzialmente, poichè il suo utilizzo risulta principalmente circoscritto agli slot preconfigurati e offre una flessibilità limitata nella gestione di quelli definiti dallo sviluppatore. Invece, l'incremento del valore della soglia di confidenza ha determinato un miglioramento delle metriche utilizzate per l'analisi del chatbot. Tale risultato suggerisce che l'aumento di questo parametro possa comportare un miglioramento nell'interpretazione degli enunciati dell'utente e, di conseguenza, la corretta associazione degli stessi ai relativi intent.

Nel complesso, il servizio ha dimostrato di poter aumentare il livello di personalizzazione nella risoluzione delle richieste dell'utente attraverso l'integrazione con altri servizi offerti da AWS. Nel caso preso in esame, tale integrazione è stata realizzata mediante funzioni Lambda, impiegate per raccogliere informazioni specifiche legate al caso d'uso, successivamente estese attraverso la realizzazione di un database mediante Amazon DynamoDB. Tale struttura

si rivela, tuttavia, molto debole nel caso in cui sia necessario effettuare un aggiornamento costante delle fonti informative, come nel caso di studio realizzato. Pertanto, l'adozione di un'architettura differente, ad esempio mediante la realizzazione di una knowledge base con Amazon Bedrock collegata direttamente alle sorgenti dati di interesse, potrebbe garantire un processo di aggiornamento maggiormente automatizzato. Infatti, nel caso di un progetto destinato a un utilizzo nel lungo periodo, risulta necessario prevedere un adeguato livello di manutenzione sia del bot sia dell'architettura sulla quale esso si basa.

In generale, Amazon Lex si è rivelato, per il caso d'uso presentato, un servizio adatto alla creazione di chatbot destinati a differenti esigenze, quali l'assistenza ai dipendenti nella gestione di domande relative all'azienda, l'integrazione all'interno di siti web per la gestione delle domande frequenti degli utenti e l'assistenza ai clienti prima dell'eventuale trasferimento della conversazione a un operatore umano. Dunque, il servizio si dimostra particolarmente adatto alla costruzione di bot caratterizzati da funzionalità relativamente semplici e da flussi conversazionali orientati alla risoluzione di esigenze specifiche.

Quindi, la possibilità di lavorare su un dominio specifico senza articolare particolarmente le proprie funzionalità rappresenta una caratteristica coerente con il paradigma dei chatbot task-oriented, i quali non devono necessariamente essere considerati una limitazione, ma possono risultare particolarmente adeguati in contesti nei quali sia richiesto lo svolgimento di compiti specifici e ben definiti, come quello analizzato nel presente caso di studio.

Tra i possibili sviluppi futuri rientrano l'integrazione dell'interazione vocale, l'introduzione di ulteriori lingue, il miglioramento delle prestazioni attraverso l'utilizzo di una knowledge base e l'ampliamento del numero di corsi dell'università gestiti dal chatbot.

La realizzazione del progetto era, inoltre, volta a mettere in evidenza l'architettura alla base di un chatbot. Trattandosi di un sistema relativamente semplice, non è stato possibile approfondire in modo significativo tutti gli aspetti ai quali, generalmente, le aziende e organizzazioni dovrebbero prestare attenzione nella realizzazione di bot conversazionali, descritti nel dettaglio nella Sezione 5.1.3. Nonostante ciò, l'utilizzo di Amazon Cognito ha permesso di evidenziare l'importanza della gestione dell'autenticazione degli utenti e, più in generale, degli aspetti legati alla sicurezza dell'interazione.

Inoltre, è stata posta l'attenzione alla necessità di rendere esplicita, sin dall'inizio della conversazione, la natura automatizzata dell'interazione. Nel progetto non è stato, invece, previsto un meccanismo di trasferimento della conversazione a un operatore umano, funzionalità che potrebbe essere presa in considerazione nell'ambito di sviluppi futuri.

Infine, le metriche messe a disposizione da Amazon Lex non registrano direttamente dati personali degli utenti, ma si concentrano su parametri strettamente legati alla valutazione degli intent, degli slot e degli enunciati elaborati dal bot. Tale caratteristica consente di limitare la presenza di informazioni personali all'interno dei dati utilizzati per il monitoraggio e la valutazione delle prestazioni del sistema.

Conclusioni

Giunti al termine dell'elaborato, è possibile affermare che gli obiettivi delineati nell'Introduzione sono stati raggiunti. Nella prima parte della tesi è stato introdotto il concetto di Intelligenza Artificiale, al fine di migliorare la comprensione dei casi di studio presentati successivamente. Poichè i servizi adoperati nell'elaborato si basano sul Machine Learning e sul Deep Learning, sono state introdotte tali tecnologie. Per completare il quadro introduttivo, l'Intelligenza Artificiale è stata collocata all'interno dell'attuale panorama tecnologico e scientifico, sottolineando il rapido progresso dei modelli di apprendimento, oltre ai benefici e ai rischi associati a tale tecnologia. Tali aspetti hanno costituito il punto di riferimento per le analisi svolte successivamente. Il capitolo relativo ad AWS è stato principalmente dedicato alla presentazione dell'offerta dell'ecosistema AWS, descrivendone i principali servizi e, in particolare, quelli che adottano tecniche di IA. Una volta forniti gli strumenti necessari alla comprensione dell'analisi proposta, sono stati esaminati singolarmente i servizi selezionati, ovvero Amazon Rekognition, Amazon Transcribe e Amazon Lex.

Per Amazon Rekognition è stato proposto un caso di studio nel quale sono state analizzate le capacità del servizio di individuare il testo contenuto nelle immagini. Queste ultime sono state selezionate secondo un livello progressivo di complessità. Nelle prime prove, le caratteristiche delle immagini sono state tali da favorire l'individuazione del testo, ad esempio il testo caratterizzato dall'utilizzo di un font semplice, dall'assenza di elementi sovrapposti e da condizioni di illuminazione favorevoli. Invece, nelle ultime immagini erano presenti un numero considerevole di vocaboli, condizioni di illuminazione sfavorevoli, distorsioni del testo e dissolvenza. Questo approccio ha permesso di evidenziare come il servizio non abbia riscontrato particolari difficoltà nelle prime prove, mentre non sia stato in grado di riconoscere e trascrivere correttamente i vocaboli negli ultimi casi.

In modo analogo, è stata condotta l'analisi del servizio Amazon Transcribe. In questo modo, sono emerse alcune delle problematiche ancora riscontrabili in questi sistemi. In particolare, le ultime prove erano volte a osservare il comportamento del servizio in presenza di più condizioni sfavorevoli, quali rumore di fondo, presenza di più lingue, sovrapposizione di più parlanti e utilizzo di accenti particolari.

Per l'ultimo caso di studio è stato realizzato un chatbot tramite Amazon Lex, analizzandone il comportamento in relazione alle funzionalità di IA messe a disposizione dal servizio. Inoltre, è stata osservata l'architettura alla base dei chatbot attraverso la realizzazione di un'interfaccia utente accessibile tramite sito web.

Pertanto, il presente lavoro intende fornire un contributo all'analisi del comportamento di sistemi che integrano funzionalità di Intelligenza Artificiale, con l'obiettivo di facilitare la comprensione delle limitazioni ancora presenti in tali sistemi e delle condizioni nelle quali,

invece, essi mostrano le migliori prestazioni.

Per quanto riguarda i possibili sviluppi futuri, il lavoro potrebbe essere esteso attraverso un'analisi di ulteriori aspetti legati al riconoscimento delle immagini e allo Speech-to-Text, non trattati all'interno del presente elaborato. Infatti, AWS mette a disposizione, per entrambi i servizi, funzionalità specifiche. Ad esempio, Amazon Transcribe Medical, basato su Amazon Transcribe ma specificamente orientato al settore medico, potrebbe essere analizzato al fine di valutarne le capacità in tale contesto. In merito ad Amazon Rekognition, potrebbe essere interessante valutare anche le funzionalità relative al riconoscimento del testo all'interno dei video e osservarne il comportamento, considerando la possibile applicazione di tali tecnologie in contesti quali la robotica e i sistemi di guida autonoma. Un ulteriore approfondimento potrebbe riguardare l'analisi delle funzionalità che consentono di creare modelli personalizzati, disponibili per entrambi i servizi. Pertanto, sarebbe possibile osservare il comportamento dei sistemi sulla base dei dataset selezionati per il loro addestramento, valutando l'eventuale miglioramento delle prestazioni rispetto ai modelli predefiniti. Infine, il chatbot realizzato mediante Amazon Lex potrebbe essere ulteriormente sviluppato abilitando l'interazione vocale e introducendo il supporto a un numero maggiore di lingue. Un ulteriore intervento riguarderebbe la possibilità di modificare la fonte utilizzata per il reperimento delle informazioni relative ai corsi, sfruttando le funzionalità di Amazon Bedrock integrabili con Amazon Lex. In questo modo, sarebbe possibile confrontare i risultati ottenuti con quelli illustrati nel presente elaborato e valutare l'eventuale miglioramento delle prestazioni del chatbot attraverso l'inclusione di funzionalità di IA generativa. Inoltre, tale approccio potrebbe contribuire alla manutenzione del bot, eliminando la necessità di aggiornare manualmente il database contenente le informazioni sui corsi, che potrebbero essere recuperate direttamente dai siti web dell'università.

Bibliografia

ALEXEEV, B., BARRETO, K., LI, Y., LICHTMAN, J. D., PRICE, L., SHAH, J. I., TANG, Q. e TAO, T. (2026), «Primitive sets and von Mangoldt chains: Erdős Problem 1196 and beyond», URL https://arxiv.org/abs/2605.00301.

BISHOP, C. M. e NASRABADI, N. M. (2006), *Pattern Recognition and Machine Learning*, vol. 4, Springer.

BOSTROM, N. (2014), *Superintelligence: Paths, Dangers, Strategies*.

CHAPELLE, O., SCHÖLKOPF, B. e ZIEN, A., curatori (2006), *Semi-Supervised Learning*, The MIT Press.

DUTTA, A. (2025), «Taxonomy of Artificial Intelligence», .

GAO, J., XIONG, C., BENNETT, P. e CRASWELL, N. (2023), *Neural Approaches to Conversational Information Retrieval*, vol. 44, Springer.

GOODFELLOW, I., BENGIO, Y. e COURVILLE, A. (2016), *Deep Learning*, The MIT Press, URL http://www.deeplearningbook.org.

GOSCINSKI, A., BROBERG, J. e BUYYA, R. (2011), *Cloud computing: Principles and Paradigms*, Wiley Online Library.

HWANG, K., DONGARRA, J. J. e FOX, G. C. (2012), *Distributed and Cloud Computing: From Parallel Processing to the Internet of Things*, Morgan Kaufmann.

JEGHAM, N., ABDELATTI, M., KOH, C. Y., ELMOUBARKI, L. e HENDAWI, A. (2025), «How Hungry is AI? Benchmarking Energy, Water, and Carbon Footprint of LLM Inference», URL https://arxiv.org/abs/2505.09598.

JURAFSKY, D. e MARTIN, J. H. (2026), *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models*, 3rd ed., URL https://web.stanford.edu/~jurafsky/slp3/, online manuscript released August 19, 2026.

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

MURPHY, K. P. (2012), *Machine learning: A Probabilistic Perspective*, The MIT Press.

RUSSELL, S. e NORVIG, P. (2020), *Artificial intelligence: A Modern Approach*, Pearson, 4/E.

SAJADIEH, S., FATTORINI, L., PERRAULT, R., GIL, Y., PARLI, V., SANTARLASCI, L., PAVA, J., MASLEJ, N., ALTMAN, R., BRYNJOLFSSON, E., BRODLEY, C., CLARK, J., DIGNUM, V., KUMAR, V., LANDAY, J., LYONS, T., MANYIKA, J., NIEBLES, J. C., SHOHAM, Y., TABASSI, E., WALD, R., WALSH, T. e WELD, D. (2026), «Artificial Intelligence Index Report 2026», URL `https://arxiv.org/abs/2606.15708`.

SUTTON, R. S. e BARTO, A. G. (2015), *Reinforcement Learning: An Introduction*, Adaptive Computation and Machine Learning series, The MIT Press, second edition, in progress.

SZELISKI, R. e OTHERS (2021), *Computer vision: Algorithms and Applications*, Springer London.

Sitografia

- Amazon Web Services (AWS) – https://aws.amazon.com/
- AWS Documentation – https://docs.aws.amazon.com/
- BBC – https://www.bbc.com/
- Bloomberg – https://www.bloomberg.com/
- Gartner – https://www.gartner.com/
- Geopop — https://www.geopop.it/
- Github – https://github.com/
- Google Cloud – https://cloud.google.com/
- Guida degli insegnamenti – https://guide.univpm.it/
- HAI Stanford University Human-Centered Artificial Intelligence – https://hai.stanford.edu/
- Harvey – https://www.harvey.ai/
- IBM – https://www.ibm.com/
- Kubernetes – https://kubernetes.io/it/
- Kurzweil Technologies – https://www.kurzweiltech.com/
- Macaulay Library – https://www.macaulaylibrary.org/
- Namirial Focus – https://focus.namirial.com/
- Polizia postale – https://www.commissariatodips.it/
- Shaip by Ubiquity – https://it.shaip.com/
- Vtiger – https://www.vtiger.com/
- Wikipedia – www.wikipedia.org

Ringraziamenti

La conclusione di questo elaborato sancisce anche il termine di questo meraviglioso percorso di studi. Sembra davvero ieri che ho iniziato questa nuova avventura e, senza neanche rendermene conto, un piccolo passo alla volta, sono arrivata al suo capolinea. È dalle scuole medie che ha iniziato a farsi strada nella mia testa una vocina, dicendomi che, quando sarei diventata finalmente ''grande'', avrei conseguito una laurea per poter svolgere il lavoro dei miei sogni. Infatti, è proprio in quegli stessi anni che ho iniziato ad appassionarmi maggiormente alla scienza, alla tecnologia e anche all'informatica, e da quel momento ho capito che forse avrei dovuto dedicare la mia vita a questo. Quando ho iniziato questo percorso di studi e camminavo tra i corridoi dell'università, così grandi allo sguardo di una ragazza appena uscita dal liceo e con un po' di timore nei confronti del futuro, ho pensato che stavo iniziando a realizzare uno dei miei tanti sogni. Non è stato semplice. Ci sono stati momenti in cui mi è sembrato di non stare facendo le cose nel modo giusto. E adesso, guardando indietro, ho capito di aver concluso uno stupendo viaggio, che spero possa aprire la strada a tante altre nuove esperienze. Se sono riuscita ad arrivare fino in fondo è anche grazie a tutte le persone che mi sono sempre state vicine e sono molto grata e felice che facciano parte della mia vita.

Per tale motivo, ci tengo a ringraziare tutti coloro che mi hanno supportato e sostenuto fin dall'inizio, in particolar modo i miei genitori e tutta la mia famiglia. A mamma, che mi ha sempre incoraggiato ogni volta che dovevo sostenere un esame. A papà, con cui abbiamo fatto dei meravigliosi discorsi e con cui mi sono divertita a confrontarmi sui nuovi argomenti che affrontavo a lezione. Alla mia sorellina, che ha ascoltato tutte le mie lunghissime chiacchiere, saltando da un argomento all'altro, con molta pazienza e sempre disposta ad ascoltarmi in ogni momento. A tutte le persone che ho conosciuto durante questo percorso e che hanno contributo al raggiungimento di questo obiettivo. Ai miei amici, che mi hanno offerto il loro appoggio e mi hanno regalato dei momenti di sana spensieratezza, lasciandomi il ricordo delle risate e delle lunghe chiacchierate notturne. Un ringraziamento va al professor Domenico Ursino, che si è subito mostrato disponibile a mettere a disposizione la sua professionalità e mi ha seguito con attenzione durante il tirocinio e per tutta la stesura di questo elaborato, trasmettendomi la passione che nutre nei confronti del suo lavoro.

E, infine, vorrei ringraziare me stessa, che, nonostante alcuni momenti di sconforto e le giornate no, non si è mai abbattuta e ha sempre continuato a lavorare con grande dedizione e, soprattuto, con la curiosità e la volontà di voler scoprire sempre cose nuove. Spero di non abbandonare mai questo modo di guardare il mondo con gli occhi di una bambina, desiderosa di conoscere come funziona ciò che ci circonda. Mi auguro di poter proseguire questo bellissimo cammino, cercando di essere sempre fedele ai miei ideali. Porterò nel

cuore le notti passate a studiare e tutti questi momenti. Mi emoziona un po' scrivere questi ringrazimenti perchè conseguire una laurea, sebbene per molti possa sembrare una tappa quasi naturale della vita, rappresenta per me un obiettivo molto importante. Mi rende molto orgogliosa e spero possa rendere orgogliose anche le persone a cui voglio più bene, che hanno fatto il possibile affinchè io potessi raggiungere questo traguardo.

Grazie per questa stupenda opportunità!