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

ORGANIZZAZIONE DELL'IMPRESA

Web Scraping: Analisi della diffusione della tecnologia di AI nella filiera della moda

Autore

Niccolò Ciotti
Luca Renzi

ANNO ACCADEMICO 2024-2025

Indice

Elenco delle figure

Elenco delle tabelle

CAPITOLO 1

Obiettivo

1.1 Obiettivo

L'obiettivo del progetto è analizzare lo sviluppo e la diffusione dell'intelligenza artificiale nel contesto delle imprese italiane della moda. La sfida principale consiste nell'individuare, tramite un'attività di web scraping, le aziende che menzionano l'utilizzo o l'interesse verso l'intelligenza artificiale, analizzando la presenza di specifiche parole chiave legate a questa tecnologia. Per ottenere i risultati attesi, è necessario verificare la presenza di una serie di parole chiave correlate all'intelligenza artificiale (es. intelligenza artificiale, machine learning, deep learning, computer vision, ecc...) nella homepage del sito web aziendale. Qualora tali termini non fossero presenti nella homepage, si procede con l'analisi dei link interni alla stessa, cercando menzioni rilevanti in pagine secondarie.

L'ipotesi è che le imprese utilizzino i propri siti per comunicare l'adozione o l'interesse verso l'AI.

Infine, si analizzano i risultati ottenuti attraverso l'uso di statistiche descrittive e si includono riflessioni personali sull'adozione dell'intelligenza artificiale.

CAPITOLO 2

Tecnologie AI

2.1 Scelta delle parole chiave

Le parole chiave selezionate rappresentano alcune delle principali tecnologie e concetti associati all'intelligenza artificiale attualmente impiegati in diversi settori industriali, incluso quello della moda. La scelta è stata guidata dall'obiettivo di identificare la presenza e la diffusione di soluzioni basate su IA nei siti web aziendali del comparto "Made in Italy".

In particolare:

- **AI / IA** (e varianti come *A.I.*, *I.A.*): terminologia sintetica e di uso comune, ideale per catturare riferimenti generali all'intelligenza artificiale, sia in italiano che in contesti internazionali.
- **Intelligenza artificiale**: forma estesa in italiano, spesso utilizzata in testi istituzionali e materiali promozionali per enfatizzare l'uso strategico della tecnologia.
- **Machine Learning**: approccio statistico-funzionale di base per molti servizi personalizzati (es. previsioni di vendita, raccomandazioni prodotti), frequentemente citato in contesti applicativi.
- **Deep Learning**: sottocategoria di Machine Learning basata su reti neurali profonde, con impiego in attività di riconoscimento immagini e analisi complesse, indicativa di progetti avanzati.
- **Computer Vision**: disciplina che trasforma le immagini in dati interpretabili, fondamentale per soluzioni di prova virtuale, controllo qualità e analisi delle tendenze visive nel fashion.

- **LLM (Large Language Models)**: modelli linguistici di ultima generazione utilizzati per la generazione automatica di descrizioni prodotto, chatbot conversazionali e analisi semantica dei feedback clienti.
- **Reti neurali**: concetto chiave alla base di molte tecniche di IA e Deep Learning, spesso menzionato in sezioni tecniche o divulgative per spiegare il funzionamento degli algoritmi.

L'inclusione di questi termini consente di rilevare in modo più completo e preciso la presenza di tecnologie IA all'interno delle attività comunicative e operative delle imprese, contribuendo a valutare il grado di adozione dell'intelligenza artificiale nel settore moda.

2.2 Intelligenza artificiale

L'intelligenza artificiale è un ramo dell'informatica che si occupa dello sviluppo di sistemi in grado di eseguire compiti che, se svolti da un essere umano, richiederebbero intelligenza. Tali compiti includono la comprensione del linguaggio naturale, il ragionamento logico, la pianificazione, il riconoscimento di immagini e la capacità decisionale.

Nel contesto aziendale, l'IA viene utilizzata per ottimizzare processi, migliorare il servizio clienti tramite chatbot, automatizzare attività ripetitive e prendere decisioni basate su grandi quantità di dati. L'adozione di tecnologie IA è in crescita in numerosi settori, dalla sanità alla finanza, dalla logistica all'industria manifatturiera.

In particolare, nel settore della moda, l'intelligenza artificiale viene utilizzata per prevedere tendenze, analizzare i comportamenti dei consumatori e personalizzare le esperienze di acquisto. I brand impiegano l'IA per raccogliere ed elaborare dati provenienti da e-commerce, social media e CRM (Customer Relationship Management), al fine di identificare gusti emergenti e anticipare la domanda.

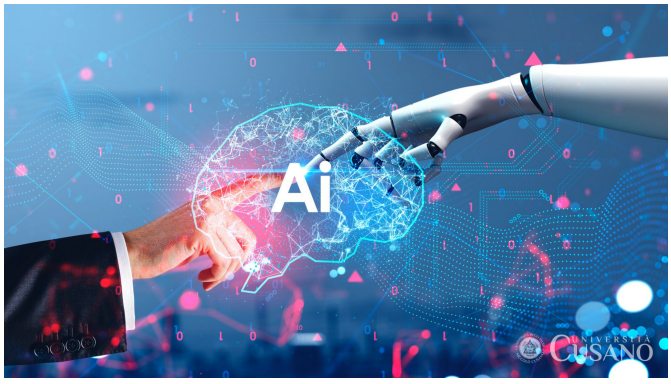

Figura 2.1: Intelligenza artificiale

2.3 Machine Learning

Il machine learning (apprendimento automatico) è una sottoarea dell'intelligenza artificiale che si concentra sullo sviluppo di algoritmi capaci di apprendere dai dati e migliorare le proprie prestazioni nel tempo, senza essere esplicitamente programmati per ogni compito.

Le applicazioni aziendali del machine learning sono molteplici, ma nello specifico nel contesto moda, è spesso usato per ottimizzare la gestione dell'inventario, prevedere le vendite e ridurre gli sprechi. Gli algoritmi possono apprendere dai dati di vendita passati per prevedere quali articoli avranno maggiore successo in specifici mercati o stagioni.

Viene anche utilizzato per identificare pattern nei comportamenti dei clienti, segmentare il pubblico e attivare campagne di marketing mirate. Alcuni marchi di lusso sfruttano il machine learning per rilevare potenziali falsificazioni di prodotti online, analizzando testi, immagini e schemi di pubblicazione nei marketplace.

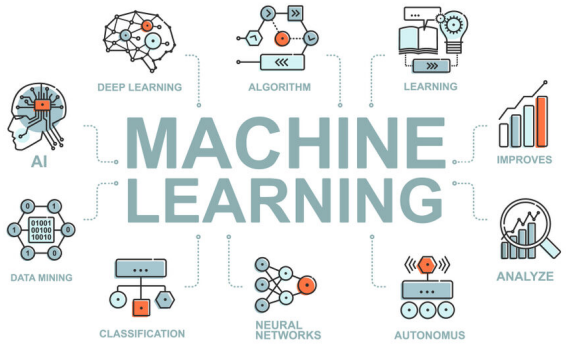

Figura 2.2: Utilizzo del machine learning

2.4 Deep Learning

Il deep learning è una branca del machine learning basata sull'uso di reti neurali artificiali profonde. Questo approccio ha ottenuto risultati straordinari in compiti complessi come il riconoscimento vocale, la traduzione automatica e l'elaborazione di immagini.

Le reti neurali profonde apprendono automaticamente rappresentazioni gerarchiche dei dati, rendendole particolarmente efficaci per affrontare problemi non lineari e ad alta dimensionalità.

Nel settore moda, il deep learning trova applicazione soprattutto nell'ambito dell'analisi visiva, come nel riconoscimento automatico di capi d'abbigliamento, classificazione delle immagini dei prodotti e prova virtuale. Alcune app mobili, ad esempio, permettono di scattare una foto a un capo e trovare prodotti simili online grazie a modelli di deep learning addestrati su grandi dataset di immagini.

Un altro utilizzo interessante riguarda la generazione automatica di nuovi design o pattern tessili, in cui le reti neurali vengono "istruite" con migliaia di stili esistenti e sono in grado di creare varianti originali, mantenendo coerenza estetica con il marchio.

2.4.1 Reti neurali

Le reti neurali sono modelli matematici ispirati al funzionamento del cervello umano. Sono composte da strati di nodi, detti neuroni artificiali, che elaborano informazioni numeriche e le trasmettono agli strati successivi.

Ogni neurone riceve degli input, li combina attraverso pesi, applica una funzione di attivazione e produce un output. Questo meccanismo consente alla rete di apprendere relazioni complesse tra i dati, migliorando le proprie prestazioni attraverso un processo di addestramento su grandi quantità di esempi.

Nelle applicazioni pratiche, l'output finale di una rete neurale è spesso una distribuzione di probabilità: ad esempio, in un problema di classificazione, la rete assegna a ciascuna classe una probabilità che indica quanto è "sicura" della propria previsione. Il risultato scelto è quello con la probabilità più alta.

Le reti neurali sono alla base di tecnologie come il riconoscimento vocale, la traduzione automatica, il rilevamento di oggetti nelle immagini e i modelli linguistici avanzati come gli LLM.

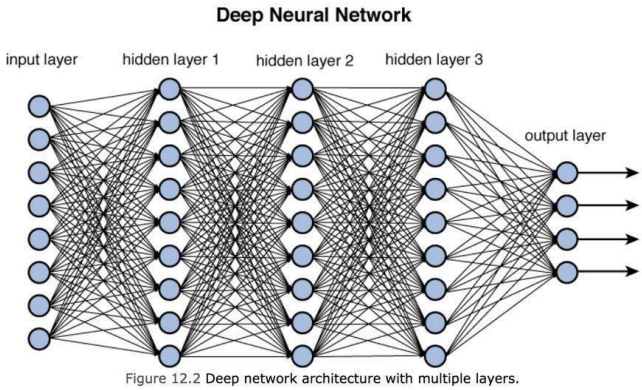

Figura 2.3: Rete neurale

2.5 Computer Vision

La computer vision è un'area dell'intelligenza artificiale che si occupa di permettere ai computer di "vedere" e interpretare immagini o video, imitando la capacità visiva umana. Ciò include il riconoscimento di oggetti, la classificazione di immagini, la rilevazione di anomalie e la segmentazione di scene.

La computer vision è una tecnologia chiave nella gestione automatizzata della produzione e del controllo qualità. Nelle fabbriche tessili o calzaturiere, le telecamere

intelligenti possono rilevare difetti nei tessuti, imperfezioni nella cucitura o variazioni nel colore dei materiali in tempo reale.

Nel retail, la computer vision viene anche utilizzata nei negozi fisici per monitorare il comportamento dei clienti, analizzando i percorsi, le interazioni con i prodotti e il tempo trascorso davanti a determinati articoli. Inoltre, sta emergendo l'uso della prova virtuale tramite realtà aumentata, che consente agli utenti di "indossare" digitalmente capi o accessori prima dell'acquisto.

Figura 2.4: Utilizzo della computer vision tramite detection

2.6 LLM

I Large Language Models rappresentano un'evoluzione avanzata nel campo dell'intelligenza artificiale e, in particolare, del natural language processing (NLP). Basati su architetture di deep learning, come le reti neurali Transformer, questi modelli sono addestrati su enormi quantità di testi per apprendere il funzionamento del linguaggio in modo statistico e contestuale.

Gli LLM sono in grado di comprendere e generare testo naturale con un alto grado di coerenza e pertinenza, risultando efficaci in attività complesse come la scrittura automatica, la traduzione, il riassunto di documenti, la risposta a domande e l'analisi semantica.

Nel settore moda, i LLM trovano applicazione in molteplici ambiti: dalla generazione automatica di descrizioni prodotto alla creazione di contenuti per e-commerce, fino al supporto nei chatbot per l'assistenza clienti. Possono anche essere utilizzati per analizzare le recensioni degli utenti, estrarre tendenze emergenti e suggerire raccomandazioni personalizzate.

Un uso innovativo riguarda la co-creazione di campagne pubblicitarie o naming di collezioni, in cui il modello fornisce ispirazione testuale sulla base dello stile del brand. Grazie alla loro flessibilità, gli LLM si integrano facilmente nei flussi creativi e operativi, potenziando l'automazione senza sacrificare la qualità comunicativa.

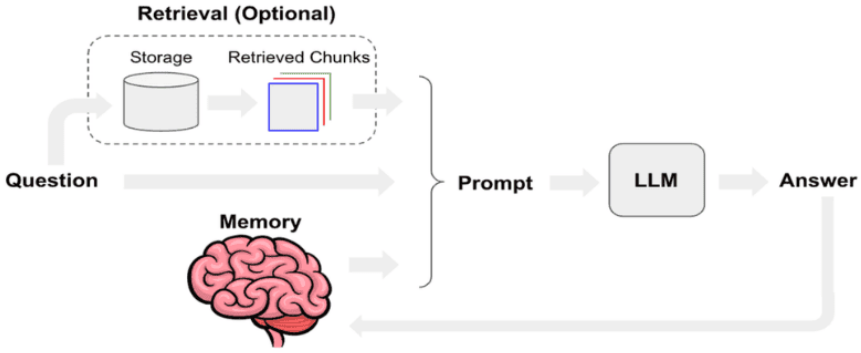

Figura 2.5: LLM

CAPITOLO 3

Il web scraping

3.1 Introduzione

Il web scraping è una tecnica informatica utilizzata per estrarre dati da siti web che prevede l'utilizzo di software o script per automatizzare il processo di raccolta di informazioni di vario tipo, come testo, immagini o tabelle. Il processo si articola generalmente in quattro fasi principali:

- Richiesta della pagina web: il programma invia una richiesta al sito come farebbe un normale browser.
- Ricezione del contenuto: il sito restituisce il codice HTML della pagina.
- Analisi del contenuto: lo scraper interpreta la struttura del codice per localizzare i dati desiderati.
- Estrazione dei dati: le informazioni vengono estratte e salvate in formati utili come file CSV, database o fogli di calcolo.

Il web scraping è ampiamente utilizzato in numerosi ambiti, ad esempio per raccogliere prezzi da siti e-commerce, aggregare notizie da portali di informazione, monitorare recensioni di prodotti, oppure analizzare contenuti pubblicati su blog e forum.

Nel nostro caso, per effettuare l'analisi si è partiti da un foglio Excel contenente l'elenco delle imprese del settore della moda del Made in Italy. Per ognuna è indicato l'url del proprio sito web e diverse informazioni riguardanti i propri dati finanziari e la loro struttura. L'obiettivo è quello di visitare ogni sito web delle aziende e ricercare al loro interno le seguenti parole-chiavi:

```
1 # Lista delle parole chiave da cercare
```

```
keywords = ["AI","A.I.","I.A.","A.I","I.A","IA","machine␣
    learning","deep␣learning","intelligenza␣artificiale","
    computer␣vision","artificial␣intelligence","LLM","large␣
    language␣models","large␣language␣model","reti␣neurali","
    neural␣networks"]
```

È stato scelto di includere, accanto ai termini in italiano, anche le rispettive versioni in lingua inglese, poiché molti siti aziendali — specialmente quelli orientati a mercati internazionali — utilizzano terminologia tecnica anglofona. Questo accorgimento permette di migliorare la copertura e l'accuratezza dell'analisi, riducendo il rischio di sottostimare la presenza di tecnologie avanzate nei contenuti testuali dei siti web.

3.2 Implementazione

3.2.1 Links

Per effettuare il web scraping abbiamo utilizzato Python, un linguaggio di programmazione particolarmente adatto per questa attività grazie alla disponibilità di numerose librerie specifiche per nell'elaborazione, gestione e manipolazione dei contenuti web. Di seguito è presente un elenco delle pricipali tecnologie e le librerie utilizzate per poter eseguire l'algoritmo, tra cui:

- Pandas: libreria fondamentale per l'analisi e la manipolazione dei dati strutturati. Nel nostro progetto, Pandas è stata utilizzata sia per importare dataset da file Excel, sia per salvare e organizzare i dati ottenuti dallo scraping in tabelle strutturate, facilitandone così l'analisi successiva. La sua capacità di gestire DataFrame in modo semplice ed efficiente ha permesso di mantenere i dati ordinati e facilmente accessibili.

- Threading: si tratta di una tecnica di programmazione che consente l'esecuzione concorrente di più operazioni all'interno dello stesso programma. Attraverso l'uso dei thread (ovvero filoni indipendenti di esecuzione), è stato possibile suddividere il lavoro di scraping in più processi paralleli. Questo approccio ha portato a un significativo miglioramento delle performance complessive del sistema, riducendo drasticamente i tempi di attesa e permettendo il recupero simultaneo di contenuti da più pagine web.

- Requests: una delle librerie più utilizzate in Python per l'invio di richieste HTTP. Nel nostro caso, Requests è stata impiegata per connettersi ai siti target, effettuare richieste GET e scaricare i contenuti HTML delle pagine web da analizzare. La sua interfaccia semplice e intuitiva ha facilitato la gestione delle connessioni e il controllo degli eventuali errori durante il processo di acquisizione.

- BeautifulSoup: libreria utilizzata per il parsing e l'analisi dei documenti HTML. In particolare, BeautifulSoup consente di navigare nella struttura di una pagina web,

permettendo di estrarre facilmente informazioni specifiche tramite tag, classi, ID o altri attributi HTML.

Durante l'analisi approfondita del dataset, è emersa un'evidente eterogeneità nella struttura degli URL presenti nella colonna "Website". Alcuni URL iniziavano con www., altri direttamente con http, mentre in alcuni casi mancavano completamente i prefissi necessari per effettuare correttamente una richiesta web. Questa varietà ha complicato il processo di estrazione delle informazioni, in quanto l'assenza di un formato uniforme rendeva instabile l'invio delle richieste HTTP.

```
# === LEGGI FILE EXCEL ===
df = pd.read_excel(prova_file)
if colonna_siti not in df.columns:
    raise ValueError(f"La colonna '{colonna_siti}' non esiste nel file Excel.")
df = df[[colonna_siti]].dropna()
df[colonna_siti] = df[colonna_siti].astype(str)

# === FUNZIONI ===
def normalizza_url(url):
    url = url.strip()
    if not url.startswith(("http://", "https://")):
        url = "https://" + url
    return url
```

Figura 3.1: Funzione che controlla gli URL

Per ovviare a questo problema e garantire l'affidabilità del processo di scraping, abbiamo sviluppato uno script di pre-processing (come mostrato in Figura 3.1), il cui scopo è quello di normalizzare tutti gli URL presenti nel dataset. Lo script scorre ogni voce della colonna "Website" e applica delle regole correttive, tra cui:

- Se un URL inizia con www. o con http, viene automaticamente modificato affinché cominci con https://.
- Nel caso in cui il protocollo sia assente o non conforme agli standard richiesti, viene anch'esso sostituito con https://.

L'adozione del protocollo https:// non è casuale: questo schema di URL permette infatti di effettuare connessioni sicure e cifrate tra il client e il server. In questo modo, oltre a garantire una maggiore sicurezza nella trasmissione dei dati, si riduce la probabilità che il server rifiuti la connessione a causa di protocolli non sicuri o obsoleti.

3.2.2 Headers e User-Agent

Un ulteriore accorgimento adottato durante la fase di scraping è stato l'inserimento degli header nelle richieste HTTP, con particolare attenzione all'User-Agent, Figura 3.2. Gli header HTTPS sono informazioni aggiuntive inviate al server insieme alla

richiesta, che permettono di simulare il comportamento di un vero browser e di rendere la comunicazione più chiara e conforme agli standard.

```
def cerca(sito):
    try:
        url = normalizza_url(sito)
        print(f"➡️ [MAIN] Richiesta a: {url}")
        response = requests.get(url, headers=headers, timeout=10)
```

Figura 3.2: User-Agents

Tra questi, lo User-Agent gioca un ruolo cruciale: si tratta di una stringa, che identifica il client che sta effettuando la richiesta ad esempio, un browser come Google Chrome, Firefox o Safari. I server web spesso utilizzano questa informazione per decidere se e come rispondere a una richiesta. Se il server rileva che la richiesta proviene da uno script automatico o da un bot, può decidere di bloccarla, restituire una pagina vuota oppure attivare misure di protezione come i CAPTCHA.

```
1 headers = {'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; Win64;
    x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/115.0
    Safari/537.36'}
```

Questa stringa fa apparire la nostra richiesta come proveniente da un browser reale installato su un sistema operativo Windows, aumentando la probabilità di ricevere una risposta valida e completa dal server. Inoltre nella richiesta abbiamo inserito un timeout di circa 10 secondi. Questo parametro indica il tempo massimo che lo script è disposto ad aspettare per ricevere una risposta dal server dopo aver inviato una richiesta HTTP. Se il server impiega più tempo del previsto a rispondere, lo script interrompe l'attesa e genera un'eccezione.

3.2.3 Filtraggio pagine

Quando si effettuano richieste a contenuti web durante il processo di scraping, non si inviano richieste dirette a specifici elementi HTML, ma si recupera l'intero contenuto della pagina, compresi tutti gli elementi caricati nel codice sorgente. Questo approccio, se non opportunamente filtrato, comporta un aumento significativo dei tempi di acquisizione, in quanto vengono scaricati anche contenuti non rilevanti come file multimediali, allegati o documenti esterni. Per ottimizzare le prestazioni del nostro algoritmo e ridurre al minimo i tempi di esecuzione, abbiamo implementato un filtro preliminare che analizza gli URL prima di effettuare una richiesta. L'obiettivo è quello di escludere determinati tipi di risorse non utili all'analisi.

Questa logica ci consente di evitare richieste inutili verso risorse che non contengono informazioni testuali rilevanti, alleggerendo così il carico computazionale e migliorando l'efficenza generale del processo, Figura 3.3.

```
def estrai_link(soup, base_url):
    links = set()
    lingua_valide = {"it-it", "us-en"}
    estensioni_valide = ('.html', '.htm', '.php', '.asp', '.aspx', '.jsp', '.jspx', '.js', '')

    estensioni_escluse = (
        '.jpg', '.jpeg', '.png', '.gif', '.svg', '.webp',
        '.pdf', '.doc', '.docx', '.xls', '.xlsx', '.ppt', '.pptx',
        '.mp3', '.wav', '.mp4', '.avi', '.mov', '.zip', '.rar'
    )
```

Figura 3.3: Filtro

Inoltre, per ottimizzare ulteriormente l'uso delle risorse computazionali e migliorare l'efficienza del processo di scraping, abbiamo implementato uno script di filtraggio avanzato che agisce su due aspetti fondamentali: la lingua dei contenuti e la profondità degli URL. Il web contiene una grande varietà di contenuti in lingue diverse, sopratutto se visitiamo aziende che operano a livello internazionale. Tuttavia, nel contesto del nostro progetto, risultavano rilevanti esclusivamente le informazioni in lingua italiana e inglese. Per questo motivo, abbiamo limitato l'analisi alle pagine che riportano tra i propri meta tag linguistici valori come "it-it" o "us-en". Lo script identifica la lingua di una pagina attraverso il tag e ignora automaticamente tutte le pagine che non corrispondono alle lingue predefinite.

Oltre alla selezione per lingua, è stata introdotta anche una limitazione sulla profondità degli URL. In particolare, abbiamo scelto di esplorare esclusivamente le pagine raggiungibili entro un massimo di due livelli di profondità rispetto alla homepage del sito. Ad esempio , se la homepage è https://esempio.com, verranno analizzati solo gli URL del tipo:

- https://esempio.com/sezione/
- https://esempio.com/sezione/pagina/

URL più profondi, ad esempio *https://esempio.com/sezione/sottosezione/pagina* vengono scartati per evitare un eccessivo numero di richieste e una raccolta di dati troppo dispersiva. Questo approccio, rappresentato schematicamente in Figura 3.4, ci ha permesso di mantenere un bilanciamento ottimale tra qualità e quantità dei dati raccolti, concentrandoci solo sulle sezioni principali e più significative di ciascun sito.

```
    if len(path_parts) >= 1 and path_parts[0].lower() in lingua_valide:
        if len(path_parts) <= 3:
            links.add(full_url)
    elif len(path_parts) <= 2:
        links.add(full_url)
```

Figura 3.4: Livelli

3.2.4 Output

Una volta completate tutte le operazioni di ottimizzazione, l'algoritmo è stato avviato per l'analisi dei dati. I risultati ottenuti sono stati strutturati all'interno di una tabella riepilogativa, costruita in modo da facilitare la lettura e l'interpretazione dei dati raccolti. La tabella contiene una colonna dedicata agli URL di partenza, ossia i link iniziali associati a ciascuna azienda. Per ogni riga, che corrisponde a una specifica azienda, vengono analizzati sia la homepage sia i sottolink (fino a due livelli di profondità, come precedentemente specificato). Per ogni parola chiave di interesse, viene indicato:

- True se la parola è stata individuata all'interno della homepage o in una delle pagine collegate;
- False se la parola non è stata trovata oppure se non è stato possibile completare la ricerca a causa di un timeout troppo restrittivo.

In questo modo, la tabella finale fornisce una mappatura binaria della presenza o assenza delle parole chiave nei siti analizzati, rappresentando un punto di partenza efficace per eventuali analisi statistiche, confronti tra aziende o valutazioni di contenuti web.

Nella Figura 4.4 è mostrato un esempio di output.

Link	AI	A.I.	I.A.	A.I	I.A	IA	machine learning	deep learning	intelligenza artificiale	computer vision	artificial intelligence	LLM	large language models	large language mod
www.archiviozegna.com/it/family_tree/140/detail	False	False	False	False	False	False	False	False	False	False	False	False	False	False
investor.brunellocucinelli.com	False	False	False	False	False	False	False	False	False	False	False	False	False	False
www.twin-set.com	False	False	False	False	False	False	False	False	False	False	False	False	False	False
www.herno.it	False	False	False	False	False	False	False	False	False	False	False	False	False	False
www.crisconf.it	False	False	False	False	False	False	False	False	False	False	False	False	False	False
www.staffinternational.com	False	False	False	False	False	False	False	False	False	False	False	False	False	False
www.fontanamilano1915.it	False	False	False	False	False	False	False	False	False	False	False	False	False	False
www.todsgroup.com	False	False	False	False	False	False	False	False	False	False	False	False	False	False
www.solbiati.it	False	False	False	False	False	False	False	False	False	False	False	False	False	False
www.stoneisland.com	False	False	False	False	False	False	False	False	False	False	False	False	False	False
www.maxmaraorders.com	False	False	False	False	False	False	False	False	False	False	False	False	False	False
www.mnorders.com	False	False	False	False	False	False	False	False	False	False	False	False	False	False
http://www.replay.it/it/replay-info	False	False	False	False	False	False	False	False	False	False	False	False	False	False
www.diesel.com	False	False	False	False	False	False	False	False	False	False	False	False	False	False
www.stefanoricci.com	False	False	False	False	False	False	False	False	False	False	False	False	False	False

Figura 3.5: Tabella di output

CAPITOLO 4

Analisi dei risultati

4.1 Statistiche

A partire dalla tabella generata attraverso l'elaborazione dei vari link, il dataset è stato importato all'interno della piattaforma Qlik. Qlik è uno strumento avanzato di Business Intelligence che consente alle aziende di ottenere una comprensione più immediata, approfondita e interattiva dei propri dati. Grazie al suo motore associativo, Qlik permette un'esplorazione non lineare delle informazioni, facilitando l'individuazione di connessioni, pattern e tendenze che potrebbero rimanere nascoste utilizzando metodi di analisi tradizionali.

4.2 Analisi del dataset

Nel corso dell'analisi è stata realizzata una prima dashboard, Figura 4.1, interattiva volta a visualizzare la distribuzione geografica delle aziende italiane considerate nel dataset, per un totale complessivo di 5.758 imprese. La dashboard si compone di tre elementi principali: un grafico a barre, un indicatore KPI e una mappa tematica.

Il grafico a barre rappresenta il numero di aziende per ciascuna regione italiana, ordinate in maniera decrescente. Le regioni con il maggior numero di aziende risultano essere la Toscana, il Veneto e la Lombardia, ciascuna con circa mille aziende. Seguono le

Marche, la Campania e l'Emilia-Romagna, con valori che si aggirano intorno alle 600 unità. Le regioni con una minore presenza imprenditoriale nel settore analizzato sono invece la Valle d'Aosta, la Liguria, la Basilicata, il Molise e la Sardegna, con meno di 100 aziende ciascuna.

A destra del grafico, la dashboard include un indicatore numerico centrale, che riporta in modo immediato e sintetico il numero complessivo delle aziende analizzate.

Infine, la mappa mostra la distribuzione delle imprese a livello regionale, attraverso una scala di colori che varia in base alla densità di aziende. Le regioni con maggiore concentrazione sono colorate in tonalità più scure, mentre quelle con una presenza più contenuta sono rappresentate in colori più chiari. Questa rappresentazione consente di cogliere visivamente i principali poli territoriali della presenza imprenditoriale, facilitando l'individuazione di eventuali cluster regionali.

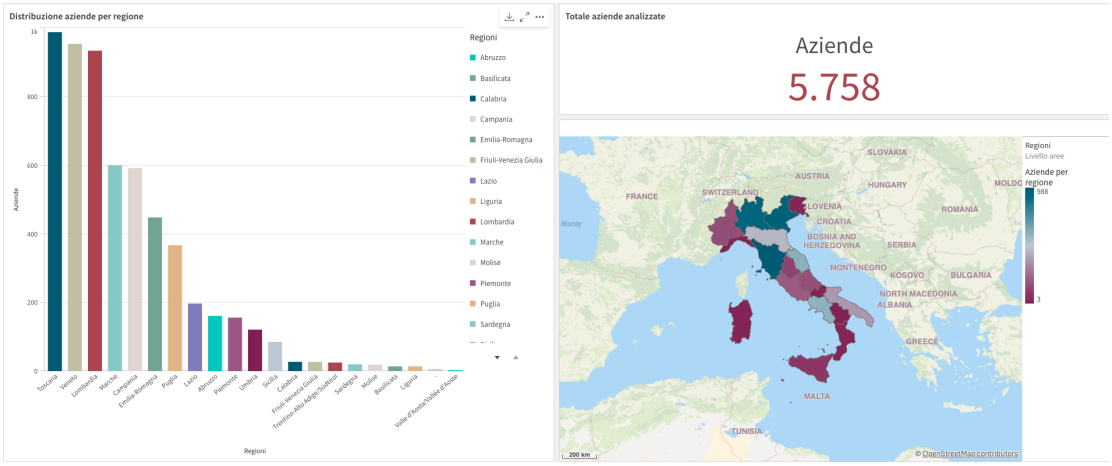

Figura 4.1: Panoramica aziende

Nel contesto dell'analisi settoriale delle imprese italiane, il Codice ATECO rappresenta un sistema di classificazione utilizzato dall'ISTAT per identificare le attività economiche svolte dalle imprese. La nomenclatura ATECO (ATtività ECOnomiche) si basa su una struttura gerarchica articolata su più livelli, che consente una classificazione sempre più dettagliata delle attività produttive.

Ad esempio, il codice ATECO 14 identifica in modo generale il comparto della fabbricazione di articoli di abbigliamento, mentre i sottocodici a due cifre aggiuntive (come 1410, 1411, 1413, ecc.) specificano più precisamente il tipo di attività svolta all'interno del settore.

La tabella 4.1 mostra alcune delle principali attività comprese nel codice ATECO 14, evidenziando la varietà di specializzazioni produttive presenti nel settore moda:

14	-	Fabbricazione di altri articoli di abbigliamento e accessori
	10	Fabbricazione di articoli a maglia e all'uncinetto
	11	Confezione di abbigliamento in pelle e similpelle
	13	Confezione di altro abbigliamento esterno
	19	Confezione di altri articoli di abbigliamento ed accessori
	20	Confezione di articoli in pelliccia
	39	Fabbricazione di pullover, cardigan ed altri articoli simili a maglia

Tabella 4.1: Classificazione delle attività nel settore abbigliamento (codice 14)

La dashboard riportata in Figura 4.2, fornisce una panoramica dettagliata delle imprese attive in Italia nel settore moda, suddivise secondo la classificazione ATECO 2007, con un'attenzione particolare alla loro distribuzione per macro-categoria e sottocategoria.

In alto a sinistra, un primo grafico a barre illustra la ripartizione delle aziende tra i due macro-codici ATECO più rilevanti per il settore:

- il codice 14, relativo alla confezione di articoli di abbigliamento, che rappresenta 3.482 imprese;
- il codice 15, che comprende le attività legate alla fabbricazione di articoli in pelle e simili, con 2.276 imprese.

Questi valori sono inoltre evidenziati da due indicatori numerici nella parte inferiore della dashboard, che offrono una sintesi immediata della composizione del campione in termini di settori produttivi.

Sulla destra, un secondo grafico a barre fornisce un approfondimento sui codici ATECO a 4 cifre, consentendo di identificare con maggiore granularità le specializzazioni delle imprese analizzate. Tra i codici più rappresentativi emergono:

- 1520 (fabbricazione di calzature)
- 1413 (confezione di altro abbigliamento esterno)
- 1419 (confezione di altri articoli di abbigliamento e accessori)

Seguono altri codici rilevanti come 1512, 1410, 1511 e 1439, che completano il quadro delle principali attività industriali nel settore.

Nel complesso, la dashboard consente di analizzare la struttura produttiva del settore moda Made in Italy, mettendo in evidenza la distribuzione delle imprese per attività economica, e supportando eventuali approfondimenti strategici o territoriali sulla base delle specializzazioni rilevate.

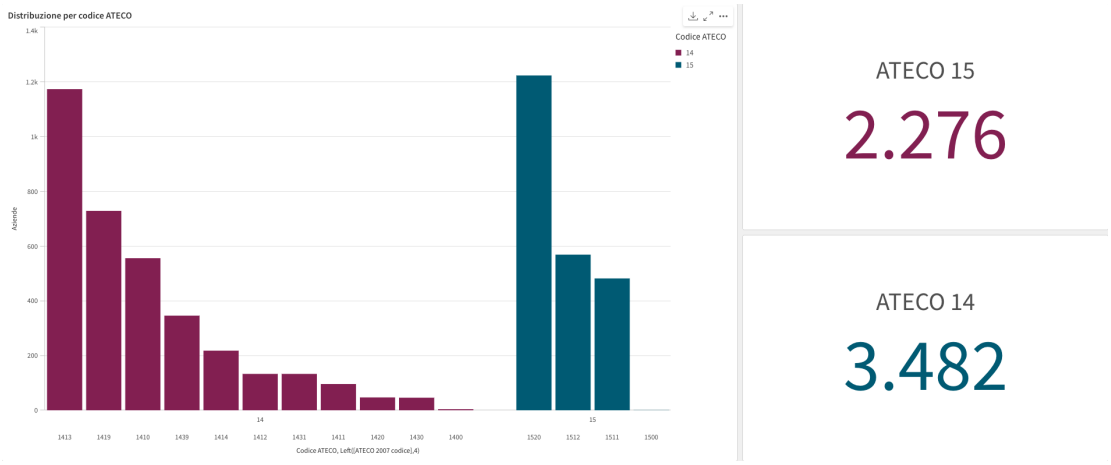

Figura 4.2: Panoramica aziende

4.3 Analisi dei risultati

Nella fase di scraping abbiamo dovuto bilanciare con cura la completezza delle informazioni raccolte con la necessità di ridurre al minimo i falsi positivi. In particolare, la preposizione italiana “ai” rappresentava un’insidia: se trattata come una semplice stringa, avrebbe generato decine di occorrenze non pertinenti. Per questo motivo, nel nostro script, Figura 4.3, abbiamo scelto di riconoscere l’acronimo “AI” solo quando è isolato e maiuscolo, tramite l’espressione regolare `\bAI\b`, escludendo tutte le occorrenze minuscole. Analogamente, la sigla “IA” viene gestita con `\bIA\b`.

Per le forme puntate, `A.I.` , `I.A.` , `A.I` e `I.A` , abbiamo imposto un matching case-sensitive che richiede esattamente la presenza dei punti, in modo da scartare varianti incomplete o in minuscolo. Le espressioni regolari sono costruite senza flag `IGNORECASE` , e al verificarsi di un match si controlla che il gruppo corrisponda esattamente alla keyword (ad esempio “A.I.” e non “a.i.”).

Le parole chiave composte, come "intelligenza artificiale", “machine learning” o “deep learning”, vengono individuate grazie a pattern che accettano uno o più spazi o trattini tra i termini, garantendo flessibilità e coerenza con diverse forme di scrittura(per esempio se viene trovato machine-learning la flag viene impostata a true). Infine, per tutte le altre keyword singole utilizziamo ancora il bordo di parola `\b...\b`, ma con flag `IGNORECASE` per catturare occorrenze indipendentemente dal maiuscolo/minuscolo.

Questo insieme di soluzioni regolari ci ha permesso di eliminare la maggior parte delle occorrenze spurie, concentrando i risultati su quei riferimenti effettivamente legati all’Intelligenza Artificiale.

```python
def contiene(text):
    trovate = []

    for keyword in keywords:
        if keyword in ["AI", "IA"]:
            pattern = r'\b' + keyword + r'\b'
            match = re.search(pattern, text)
        elif keyword in ["A.I.", "I.A.", "A.I", "I.A"]:
            # Cerca la forma con punti obbligatori, solo se tutto maiuscolo
            pattern = r'(?<!\w)' + keyword.replace(".", r"\.") + r'(?!\w)'
            match = re.search(pattern, text)  # case-sensitive: NO flags=re.IGNORECASE
            # Se matcha qualcosa tipo "a.i." (minuscolo), lo scarta
            if match and match.group() != keyword.replace(".", ""):
                continue
        else:
            if " " in keyword:
                parts = map(re.escape, keyword.split())
                pattern = r'\b' + r'[\s\-]+'.join(parts) + r'\b'
            else:
                pattern = r'\b' + re.escape(keyword) + r'\b'
            match = re.search(pattern, text, flags=re.IGNORECASE)

        if match:
            # Solo per "AI" evitiamo falsi positivi con "ai" minuscolo
            if keyword == "AI" and match.group() != "AI":
                continue

            trovate.append(keyword)

    return trovate
```

Figura 4.3: Codice estrazione keywords

Le due rappresentazioni grafiche realizzate, Figure 4.4 e 4.5 offrono uno spaccato interessante sul modo in cui le aziende del settore moda comunicano la propria vicinanza all'Intelligenza Artificiale. Il grafico a barre evidenzia innanzitutto la netta supremazia dell'acronimo "AI": più di trecento imprese lo citano esplicitamente, mentre le formule estese ("intelligenza artificiale" e "IA") rimangono confinate a qualche decina di casi. Questo dato suggerisce che, quando si tratta di mettere in evidenza il legame con l'AI, le aziende preferiscono un termine breve e di immediata comprensione.

Non meno significativo è il dato relativo alle tecnologie più specifiche, come "machine learning", "deep learning", "LLM" o "computer vision", che compaiono solamente in un numero ristretto di siti (da una decina fino a una ventina di occorrenze ciascuna). Ne deriva l'impressione che, sebbene l'adozione di strumenti avanzati sia già presente in alcune realtà, rimanga ancora un fenomeno di nicchia all'interno del panorama moda italiano.

A conferma di quanto appena osservato, il KPI separato ricorda che soltanto 339 aziende — circa il 5,9% del campione di 5758 imprese — dichiarano almeno un riferimento all'AI. In un contesto in cui l'Intelligenza Artificiale è spesso al centro del dibattito tecnologico e mediatico, appare evidente come la sua comunicazione sia per ora circoscritta a una minoranza di operatori, che probabilmente sfruttano questo tema

soprattutto per rafforzare la propria immagine di innovatori.

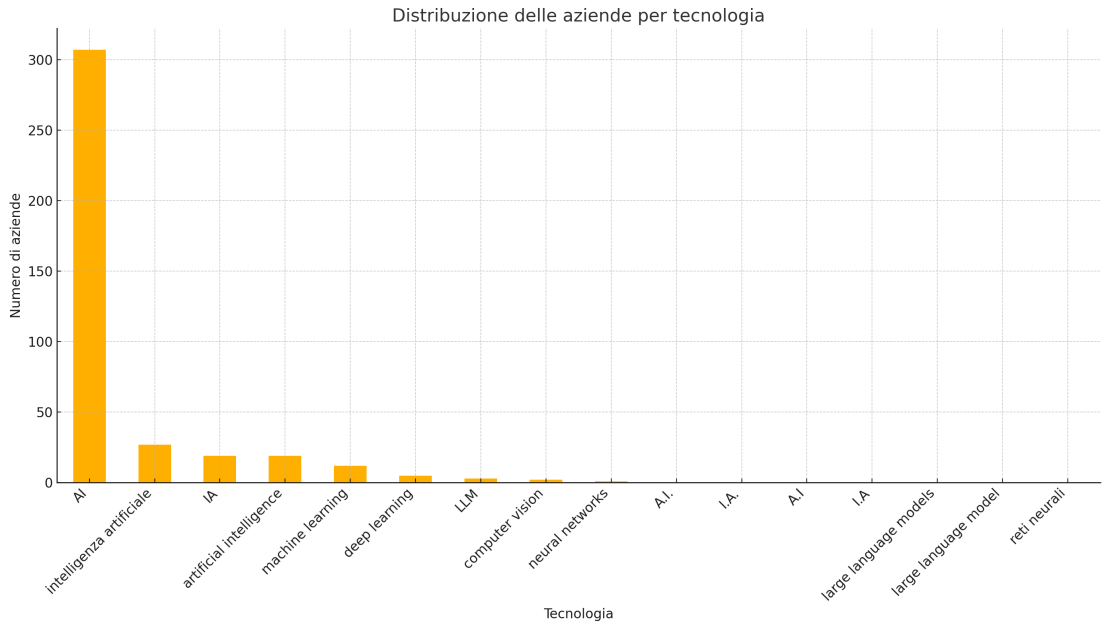

Figura 4.4: Distribuzione delle aziende per tecnologia

Figura 4.5: Aziende che citano l'intelligenza artificiale

4.3.1 Verifica ulteriori falsi positivi

Nonostante lo script riconosca "AI" solo quando isolato e in maiuscolo, è possibile che qualche occorrenza non pertinente (ad esempio acronimi aziendali o altre sigle) sia stata comunque conteggiata. Per ridurre ulteriormente il rischio di falsi positivi, abbiamo isolato il sottoinsieme di imprese che dichiarano "AI" e almeno un'altra tecnologia tra quelle monitorate.

Il conteggio restituisce:

- Aziende che citano solo "AI": **279**
- Aziende che citano "AI" e almeno un'altra tecnologia: **28**

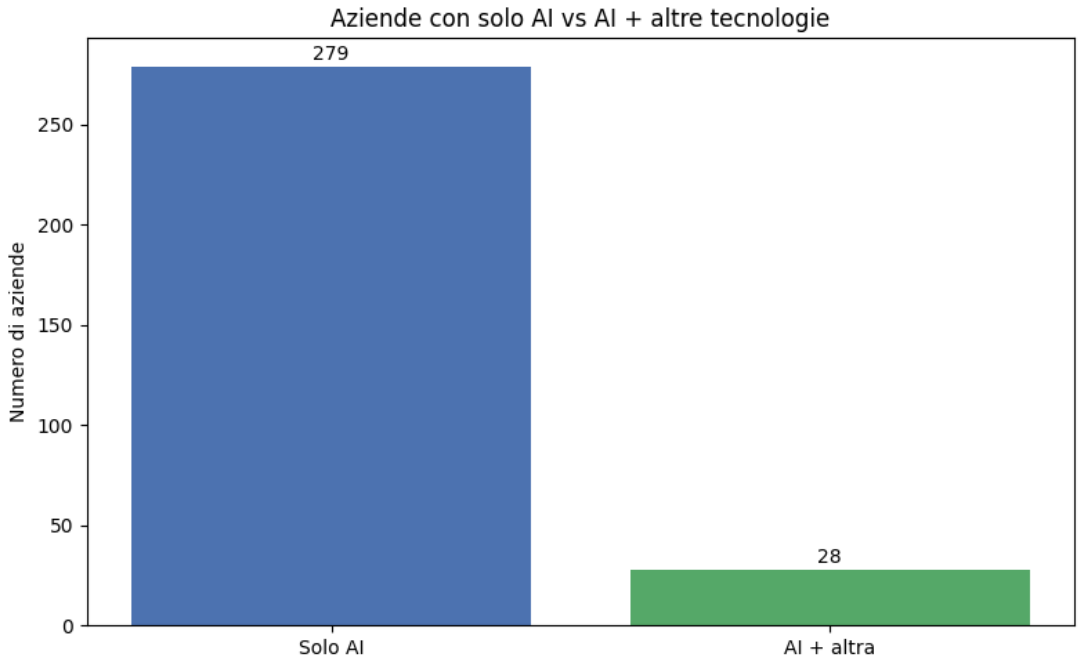

Figura 4.6: Confronto tra aziende con solo "AI" e aziende con "AI" più un'altra tecnologia

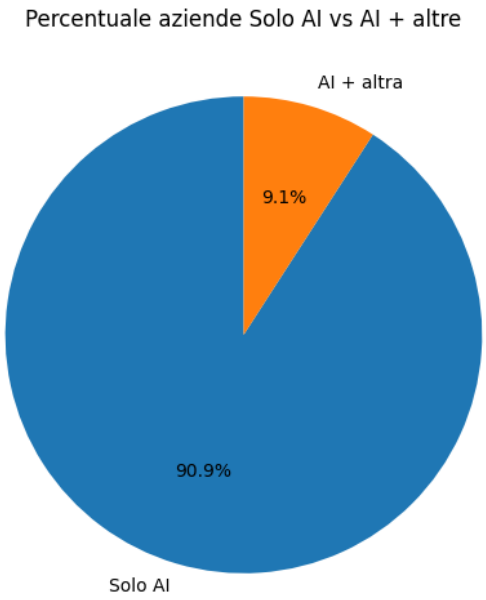

Figura 4.7: Percentuale aziende "Solo AI" vs "AI + altre tecnologie"

Pur avendo adottato filtri rigorosi per riconoscere "AI" solamente come acronimo isolato e maiuscolo, resta il rischio che alcune occorrenze corrispondano alla preposizione italiana "ai" qualora fosse scritta in maiuscolo. La verifica incrociata con la presenza di ulteriori keyword AI-related mostra che solo il 9,1 % delle aziende abbina "AI" a un secondo termine tecnico, confermando in questi casi l'effettivo riferimento all'Intelligenza Artificiale.

Tuttavia, nel 90,9 % dei casi la sigla appare isolata. Nonostante la bassa probabilità che si tratti di una preposizione, questo scenario non ci permette di escluderla a priori in tutti i casi. Di conseguenza, pur rafforzando la nostra fiducia nell'accuratezza del filtro, riconosciamo che un residuo margine di falsi positivi rimane inevitabile.

4.4 Limitazioni

È fondamentale sottolineare il principale limite di questa analisi: il metodo individua unicamente le imprese che scelgono di menzionare l'AI sui propri siti, non quelle che realmente la utilizzano "dietro le quinte". È verosimile che molte aziende impieghino algoritmi di machine learning o sistemi di visione computerizzata senza pubblicizzarlo nei propri canali digitali, soprattutto quando queste applicazioni hanno finalità puramente operative e non di comunicazione. Perciò, i risultati ottenuti riflettono più la

volontà di un certo storytelling tecnologico che il grado effettivo di penetrazione dell'AI nel settore.

CAPITOLO 5

Conclusioni

5.1 Conclusioni

L'analisi condotta ha messo in luce come, nonostante il notevole potenziale offerto da tecnologie quali l'Intelligenza Artificiale, il Machine Learning, la Computer Vision, il Deep Learning, gli LLM e le reti neurali, la loro adozione nel comparto moda italiano risulti ancora circoscritta. Da un lato, questa limitata visibilità è dovuta a una strategia di comunicazione che privilegia pochi player, spesso di grandi dimensioni, in grado di valorizzare l'innovazione per finalità di marketing. Dall'altro, molte realtà – in particolare piccole e medie imprese a conduzione familiare – affrontano barriere culturali e organizzative: processi artigianali radicati, diffidenza verso tecnologie percepite come complesse e costi iniziali elevati. In questi casi, l'effettivo impiego di algoritmi avanzati può restare "invisibile", nascosto dietro la normale operatività, e non tradursi in dichiarazioni pubbliche.

Il contesto di mercato e la pressione dei consumatori globali giocano un ruolo chiave: chi opera soprattutto a livello nazionale potrebbe non avvertire l'urgenza di integrare tecnologie digitali, mentre l'internazionalizzazione e la crescente attenzione a sostenibilità, tracciabilità e autenticità stanno facendo emergere nuove esigenze.

In prospettiva, la diffusione su larga scala di queste innovazioni richiederà:

- **Superamento delle barriere culturali**: promuovere una mentalità aperta alla sperimentazione, anche in realtà più tradizionali, attraverso workshop e casi di successo.
- **Sviluppo di competenze interne**: investire in formazione specializzata per dotare le imprese di risorse in grado di gestire autonomamente progetti di AI.
- **Incentivi e politiche di supporto**: favorire l'adozione tecnologica con agevolazioni fiscali, contributi a fondo perduto e programmi di collaborazione pubblico-privato.

Solo in questo modo il Made in Italy potrà cogliere i benefici derivanti da processi più efficienti, prodotti personalizzati e modelli di business sostenibili. L'integrazione consapevole di AI, Machine Learning, Computer Vision, Deep Learning, ecc..., non rappresenta un tradimento della tradizione artigianale, bensì un'opportunità per rinnovarla, rafforzarla e proiettarla nel futuro con un valore aggiunto riconosciuto a livello globale.