



**UNIVERSITÀ
DEGLI STUDI
DI PADOVA**



**DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE**

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

CORSO DI LAUREA IN INGEGNERIA INFORMATICA

VEICOLI AUTONOMI: VIAGGIO TRA ALGORITMI ED ETICA

Relatore: Prof. Antonio Rodà

Laureanda: Giusi Grotto

Matricola: 1216343

Correlatrice: Prof.ssa Silvana Badaloni

ANNO ACCADEMICO 2021 – 2022

Data di Laurea 22 settembre 2022

Abstract

La tesi ha il primario obiettivo di far luce sui bias negli algoritmi alla base dei veicoli a guida autonoma (VA). Dai vari casi di studio visionati sono stati rilevati chiaramente dei bias in termini di performance nei metodi di riconoscimento di un ostacolo che può essere un pedone, un ciclista o un altro veicolo. Nella seconda parte si discute circa la sorgente di tali bias, le implicazioni etiche che essi hanno, tra cui il concetto di etica delle conseguenze e dei doveri. Si riportano possibili metodi per la correzione ed eventuali ostacoli alla soluzione di questi problemi.

Indice



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

.....	1
CORSO DI LAUREA IN INGEGNERIA INFORMATICA	1
VEICOLI AUTONOMI: VIAGGIO TRA ALGORITMI ED ETICA	1
INTRODUZIONE.....	1
Capitolo 1: I VEICOLI AUTONOMI.....	2
1.1) INTRODUZIONE.....	2
1.2) I LIVELLI	3
1.3) LE OPPORTUNITA'	5
1.4) I RISCHI	6
Capitolo 2: COMPUTER VISION per la RILEVAZIONE DELL'OSTACOLO.....	8
2.1) SVM.....	8
2.2) LE RETI NEURALI.....	9
2.2.1) Il neurone.....	10
2.3) DEEP LEARNING.....	11
2.3.1) CNN.....	11
2.3.2) Il filtraggio.....	13
2.3.3) R-CNN.....	14
2.3.4) Fast R-CNN	14
2.4) PCA	14
2.5) CNN-SVM.....	15
2.6) LM-CNN-SVM	15
2.7) DEEP LEARNING APPLICATO AI VEICOLI AUTONOMI.....	18
Capitolo 3: BIAS ALGORITMICI	21
3.1) BIAS NEI SISTEMI DI RICONOSCIMENTO DEI PEDONI.....	21
3.1.1) L'esperienza	21
3.1.2) Interpretazione dei risultati e motivazione	24
3.2) BIAS DI ETA' E GENERE.....	24
3.2.1) Il caso della Francia	24
3.2.2) La sorgente dei bias	26
3.3) TIPI DI BIAS NEI DATI	27
3.4) CONSEGUENZE ETICHE DEI BIAS	28
3.4.1) Le collisioni inevitabili.....	28
3.4.2) Etica delle conseguenze.....	29
3.4.3) Etica dei doveri.....	30
3.5) ETICA DEONTOLOGICA (DEI DOVERI) NELLA STORIA	31
Capitolo 4: CONTROLLO UMANO E ATTRIBUZIONE DELLA RESPONSABILITA'	33
4.1) CONTROLLO UMANO	33
4.2) IL PROBLEMA DELL'ATTRIBUZIONE DI RESPONSABILITA'	33
4.2.1) Veicoli semi autonomi.....	34
4.2.2) Veicoli L4	34
4.2.3) Veicoli L5	36
Capitolo 5: DESIGN KNOWLEDGE GAP E INCERTEZZA EPISTEMICA	42
5.1) DESIGN KNOWLEDGE GAP	42
5.2) OPACITA' EPISTEMICA	44
5.2.1) I problemi nel sistema.....	44

5.2.2) Il problema dell'ambiente circostante al sistema - statistical only knowledge	44
5.2.3) Decisioni umane incalcolabili.....	46
5.3) STIMA DELL'INCERTEZZA.....	46
5.3.1) Stima dell'incertezza con sampling-free	47
5.3.2) Stima attraverso un approccio bayesiano	48
5.4) DUE TIPI DI INCERTEZZA.....	48
5.5) ALEATORICA VS EPISTEMICA	48
Capitolo 6: OPINIONI DELLE PERSONE E FATTORI DI ADOZIONE	50
6.1) INTRODUZIONE	50
6.2) LE INDAGINI SVOLTE.....	50
6.2.1) <i>Global Automotive Consumer Study 2020</i>	50
6.2.2) <i>Cap Gemini 2017</i>	51
6.2.3) <i>Global Automotive Consumer Study 2022</i>	51
6.3) I FATTORI.....	52
6.3.1) <i>Time benefit</i>	53
6.3.2) <i>Environmental benefit</i>	53
6.3.3) WB.....	53
6.3.4) <i>Social recognition</i>	53
6.3.5) Edonismo	53
6.3.6) Sicurezza della tecnologia	54
6.3.7) Privacy percepita	55
6.3.8) Affidabilità.....	55
CONCLUSIONI.....	56
BIBLIOGRAFIA	57
<i>Ringraziamenti</i>	61

INTRODUZIONE

I veicoli autonomi stanno avendo una grande diffusione in diversi Stati, che cercano di integrarli nella loro rete stradale per i loro numerosi vantaggi per una mobilità futura sostenibile, più sicura e rispettosa dell'utente e dell'ambiente. Ciò può avvenire tramite lo sviluppo di nuovi sistemi di bordo e di regolazione, funzionamento, analisi e controllo del traffico. Ciò consente di raggiungere gli obiettivi di mobilità sostenibile previsti da Agenda 2030 e quelli inclusi negli Obiettivi di Sviluppo Sostenibile dell'ONU. Inoltre, grazie alle applicazioni di Intelligenza Artificiale, deep learning e big data è possibile rilevare la presenza di ostacoli di diverso tipo nell'ambiente circostante al veicolo. Tuttavia, risultano ancora aperte delle questioni di tipo etico e morale, specialmente in situazioni di collisione inevitabile, che obbliga il veicolo a scegliere con chi/cosa collidere, ponendo sul piatto della bilancia la massimizzazione del benessere collettivo da una parte e il senso del dovere dall'altra, aprendo una finestra di discussione sull'attribuzione di responsabilità che, a volte, trova riscontri a livello legale.

Capitolo 1: I VEICOLI AUTONOMI

1.1) INTRODUZIONE

Il VA (Veicolo Autonomo) è un veicolo progettato per essere in grado di circolare da solo, soddisfacendo le capacità di trasporto di una macchina tradizionale, rilevare la presenza di ostacoli nell'ambiente circostante, compresi pedoni e ciclisti senza che sia necessario l'intervento umano. Si tratta, quindi, di una **macchina autonoma** che gode di **autonomia operativa**. (Tamburrini et al., 2021)

È opportuno, ora, capire cosa si intende per macchina autonoma. A tale quesito sono state date tre impostazioni, nelle quali ci si focalizza rispettivamente su:

- i compiti che la macchina deve svolgere senza interventi da parte di agenti esterni (naturali o artificiali)
- le capacità cognitive, percettive e decisionali delle quali essa deve essere dotata;
- la sua architettura e organizzazione funzionale.

La prima opzione può essere perseguita considerando da una parte che cosa un sistema di Intelligenza Artificiale deve fare e dall'altra quali sono le capacità necessarie o sufficienti per farlo a partire dal riconoscimento percettivo di varie classi di oggetti, ad esempio un volto umano.

Nel caso specifico del VA, esso ha il compito di circolare su una parte di rete viaria dei veicoli con conducente e avere delle prestazioni almeno altrettanto buone di quelle di un conducente umano competente. Affinché il veicolo raggiunga l'autonomia nel riconoscimento dell'ostacolo nell'ambiente, è necessario che vengano implementati appositi hardware (es. sensori) e software. Vengono quindi usati dei sistemi di Intelligenza Artificiale ¹, affrontati nel dettaglio nel prossimo capitolo, e di navigazione interna come GPS.

¹ <https://unric.org/it/obiettivo-12-garantire-modelli-sostenibili-di-produzione-e-di-consumo/>

1.2) I LIVELLI



Fig. 1 Livelli di automazione secondo SAE 2016

Considerando il termine “autonomia operativa”, nel caso dei VA, essa dipende dal singolo veicolo, in particolare dal livello di automazione dello stesso.

Per la SAE² esistono sei livelli della scala per la guida autonoma e sono indicati rispettivamente come L0, L1, L2, L3, L4, L5, dove L0 indica che il veicolo non è dotato di alcun tipo di automazione e L5 corrisponde a una situazione di piena automazione. Nel dettaglio:

L0: NESSUNA AUTOMAZIONE. Il pieno controllo dell’automobile è a carico del guidatore umano a bordo, il quale è l’unico a poter gestire la direzione e la velocità.

L1: GUIDA ASSISTITA. Il veicolo è equipaggiato di cruise control o cruise control adattivo. Il guidatore umano può quindi riposarsi, distrarsi perché il veicolo accelera fino al raggiungimento di una velocità impostata, che dovrà mantenere fino a un nuovo ordine da parte del guidatore.

L2: GUIDA SEMI-AUTONOMA. Tale livello prevede l’aggiunta, rispetto a L1, della frenata automatica e il mantenimento della corsia. In tal modo il conducente potrà, in determinati contesti, specie quelli che presentano rallentamento del traffico, non preoccuparsi di freno e volante pur dovendo stare attento a ciò che lo circonda. Praticamente i sistemi mantengono il veicolo al centro della corsia e la distanza di sicurezza dal veicolo che precede.

L3: GUIDA SEMI-AUTONOMA AVANZATA: il veicolo è in grado di gestire freno e acceleratore in modo adattivo, rispettando le distanze di sicurezza e leggendo i cartelli facendo uso di telecamere o cartografia GPS. I sistemi impiegati fanno sì che il veicolo resti al centro della corsia e non esca di

² Society of Automotive Engineer, <https://www.sae.org/>

strada, segua il profilo del manto stradale, esegua curve o cambi di direzione e sorpassi un veicolo che precede azionando l'indicatore di direzione. Tali veicoli sono in grado di circolare su autostrade, zone di rallentamento del traffico o aree di traffico intenso. Il conducente deve comunque esercitare una sorveglianza attiva ed essere pronto a intervenire se necessario.

L4: GUIDA ALTAMENTE AUTOMATIZZATA. Il veicolo con tale livello è in grado di circolare da solo di fronte a gran parte delle situazioni della vita reale. I veicoli L4 sono in grado di comunicare fra di loro per meglio reagire a situazioni pericolose o imprevisti. Il livello di attenzione richiesta al conducente diminuisce notevolmente. Il problema è che gli L4 non sono completamente affidabili quindi al conducente viene ancora chiesto di restare al posto di comando e controllare le azioni del veicolo e l'ambiente circostante per intervenire in caso di necessità.

L5: GUIDA COMPLETAMENTE AUTOMATIZZATA. Il veicolo è in grado di circolare in totale autonomia in tutti i contesti stradali, anche quelli più complessi e affrontare tutte le condizioni reali quotidiane. Gli L5 sono veicoli connessi, in grado di comunicare tra di loro e con le infrastrutture. Ogni situazione sarà gestita e sarà il veicolo a prendere la giusta o la migliore decisione.

1.3) LE OPPORTUNITA'

Vengono riportate in Fig. 2 alcune possibili opportunità offerte dall'utilizzo dei veicoli a guida autonoma³:

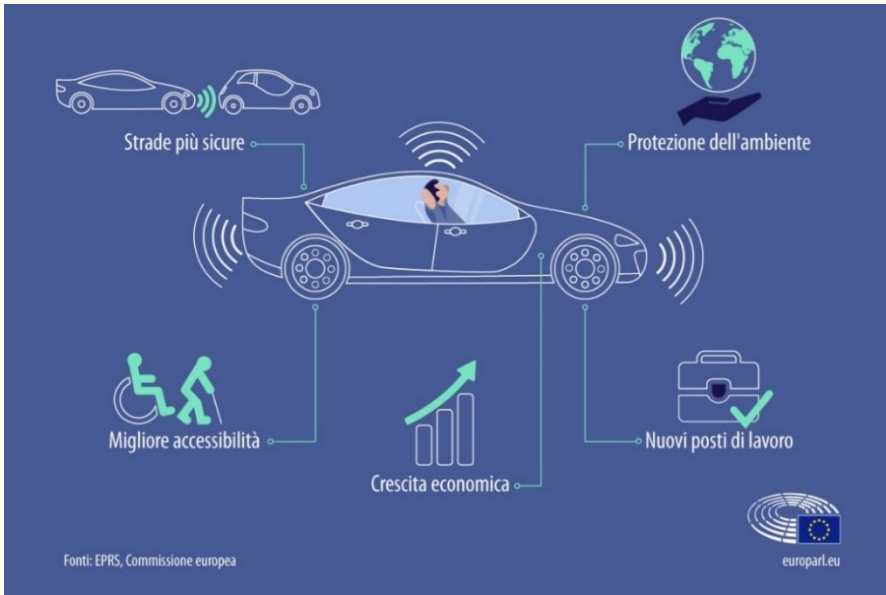


Fig. 2 Opportunità offerte dal VA

- E' ipotizzabile che gli spostamenti in auto diventino più sicuri, con riduzione del numero degli incidenti in quanto il sistema di guida non si stanca e non si distrae, oltre ad essere meno influenzabile da fattori ambientali, quali scarsa luce, nebbia e pioggia;
- Si può sfruttare il tempo trascorso nel veicolo per lavorare, studiare, persino dormire, riposarsi. Il guidatore non ha lo stress della guida;
- È attesa una diminuzione della ownership, ossia del numero di persone con veicoli propri⁴, con possibile riduzione delle emissioni ⁵per la produzione di nuovi veicoli da porre in circolazione;
- Grazie all'esistenza di flotte di fleet management controllate e ottimizzate è pure attesa una riduzione della congestione del traffico, non ci sono code ma i flussi possono aumentare.
- Il VA risulta inclusivo per tutte le persone non dotate di patente, ad esempio, coloro che non hanno ancora raggiunto l'età per possedere la patente di guida o coloro che per scelta non

³ <https://www.europarl.europa.eu/news/it/headlines/economy/20190110STO23102/auto-a-guida-autonoma-in-ue-dalla-fantascienza-alla-realta>

⁴ <https://www.readkong.com/page/the-oslo-study-how-autonomous-cars-may-change-transport-3448047>

⁵ <https://unric.org/it/obiettivo-12-garantire-modelli-sostenibili-di-produzione-e-di-consumo/>

hanno la patente ma hanno la necessità di spostarsi, ma anche patentati che non si sentono più sicuri al volante e preferiscono affidarsi, in questo caso, alla tecnologia;

- Si prevede una crescita sensibile del mercato dei veicoli autonomi con la creazione di nuovi posti di lavoro e al raggiungimento entro il 2025 di profitti di 625 miliardi di euro per il settore automotive e 180 per quello elettronico.

1.4) I RISCHI

Questa tecnologia può comportare anche vari rischi, che hanno un certo impatto dal punto di vista etico.

- Riduzione del numero di posti di lavoro, in quanto le categorie di autisti e tassisti potrebbero essere soppiantati completamente o in parte dai VA.
- Per alcune persone la guida è una forma di piacere e una manifestazione di libertà, che potrebbe essere compromessa. Va considerato che per generazioni la patente di guida ha costituito una sorta di emancipazione e crescita sociale, che verrebbe meno;
- È possibile che il software che gestisce la guida autonoma sia affetto da errori (**bias**) e ciò può causare anche incidenti in situazioni in cui il conducente non è più in grado di riprendere rapidamente il controllo del veicolo.⁶ Il primo report USA afferma che, dal 1 luglio 2021 al 15 maggio 2022, ci sono stati 392 incidenti che hanno coinvolto veicoli con tecnologia avanzata di assistenza alla guida che hanno provocato la morte di sei persone e in cui si contano cinque feriti. Molti di questi incidenti sono dovuti alla mancata rilevazione di ostacoli nell'ambiente circostante.

Ad esempio, nel 2015, un veicolo Tesla livello L2 si è scontrato contro il rimorchio di un camion.

A seguito delle indagini effettuate, sembra che il sistema di riconoscimento del veicolo abbia confuso il cielo luminoso con il colore del camion.

Un altro caso di incidente risale al 2018, quando un veicolo autonomo Tesla Model X ha urtato contro una barriera spartitraffico, causando la morte del conducente. Nei cinque secondi corrispondenti ai centocinquanta metri di visuale prima dell'impatto, l'uomo aveva ripreso la tenuta del volante senza fare delle azioni che potessero evitare l'impatto e la sua conseguente morte. Questo è dovuto al fatto che il sistema automatico di conduzione richiede che la persona a bordo sia in grado di riprendere il controllo dell'auto in caso di particolari eventi.

⁶ <https://www.agi.it/economia/news/2022-06-17/primo-report-auto-guida-utonomo-usa-17124974/>

- In Arizona una donna è stata investita e uccisa da un SUV Volvo in modalità di guida automatica, testato dal servizio di guida Uber. (Kohli & Chadha, 2020)

Questi esempi mettono in discussione la sicurezza del VA. In riferimento agli algoritmi di Intelligenza Artificiale le questioni che sorgono riguardano l'affidabilità degli stessi e in particolare⁷:

- Robustezza: la capacità dell'algoritmo di IA di comportarsi in modo ragionevole in situazioni impreviste, non contemplate dalle specifiche. Situazioni di questo tipo in genere riguardano errori ed eccezioni di varia natura (dati di input scorretti, fallimenti di componenti software o hardware esterni al sistema e interagenti con esso, e così via). Ciò fa sì che i danni non intenzionali siano ridotti;
- aderenza ai principi e valori etici: capacità che ha il sistema di essere il più possibile coerente con il codice etico di quella particolare nazione: ad esempio non dovrebbe fare delle discriminazioni di un gruppo sociale rispetto ad un altro. Nello specifico un buon sistema di riconoscimento deve essere in grado di riconoscere in egual modo uomo e donna, con qualsiasi colore della pelle, età;
- aderenza alla legge: il sistema deve essere coerente con leggi in vigore nello Stato di applicazione, in particolare il Codice della Strada.

FIGURE

Fig.1: <https://www.dariotamburrano.it/wp-content/uploads/2018/10/auto-a-guida-autonoma.png>

Fig.2:

https://www.europarl.europa.eu/resources/library/images/20190114PHT23323/20190114PHT23323_original.jpg

⁷ <https://www.quest-it.com/blog/trustworthy-ai-linee-guida-union-europea/>

Capitolo 2: COMPUTER VISION per la RILEVAZIONE DELL'OSTACOLO

Computer Vision, o visione artificiale, è un campo di studi interdisciplinare che studia algoritmi e tecniche per permettere ai computer di riprodurre funzioni e processi dell'apparato visivo umano, per esempio riconoscere oggetti, persone o animali all'interno di un'immagine singola o in sequenza (video), estrarre informazioni utili per la loro elaborazione, a livelli sempre più alti di astrazione e comprensione.

I metodi di classificazione possono essere “tradizionali”, come Support Vector Machine, oppure basati sulle reti neurali, come le Convolutional Neural Network (CNN).

2.1) SVM

Il suo funzionamento è governato dalla teoria introdotta nel 1965 da Vapnik, e il cui scopo è quello di determinare le superfici di decisione tra le classi. SVM nasce come classificatore binario (due classi) e successivamente viene esteso a più classi. Nel caso di classificazione binaria e pattern linearmente separabili, SVM determina l'iperpiano in grado di separare le classi con il maggior margine possibile, dove il margine viene definito come la distanza minima di punti delle due classi nel training set dall'iperpiano individuato. La massimizzazione del margine è legata alla generalizzazione, perché se il margine è ampio la probabilità che pattern del test set vicini tra le classi siano classificati correttamente è più alta.

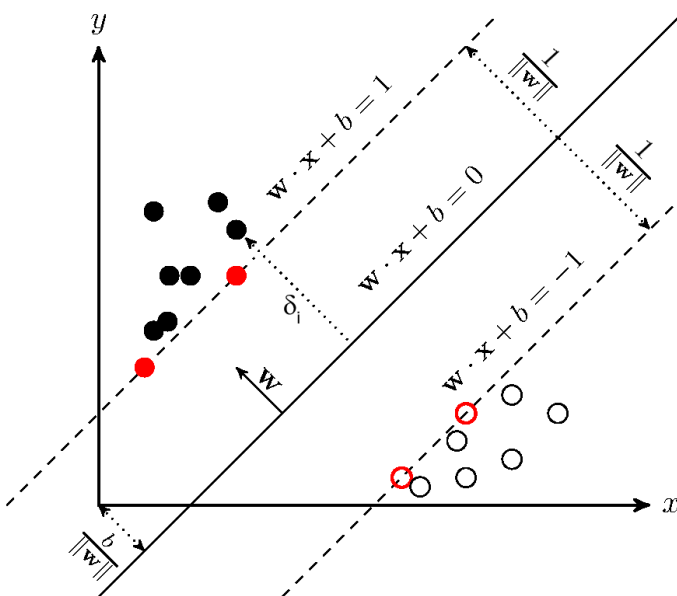


Fig. 3 SVM⁸

⁸ https://miro.medium.com/max/1400/0*OrJFZWuok1v_x9b4

La figura sopra va interpretata nel seguente modo:

- ω è il vettore normale all'iperpiano di equazione $D(x) = wx + b$;
- $\frac{b}{\|w\|}$ la distanza dall'origine;
- $D(x) = 0$ il luogo dei vettori sul piano;
- $\frac{1}{\|w\|}$ la distanza minima su ogni lato (sia dall'iperpiano di equazione $D(x) = 1$ che da quello di equazione $D(x) = -1$;
- Il margine è $\frac{2}{\|w\|}$

L'iperpiano ottimo, ossia la superficie ottima per SVM, è quello che soddisfa i vincoli di separazione dei pattern e massimizza il margine. I pattern del Training Set che giacciono sul margine si chiamano support vector. Il segno di $D(x)$ consente di classificare un pattern generico x .

2.2) LE RETI NEURALI

A partire dall'analogia biologica, sono state create delle reti neurali artificiali atte a riprodurre i meccanismi di scambio delle informazioni tipiche del cervello umano.

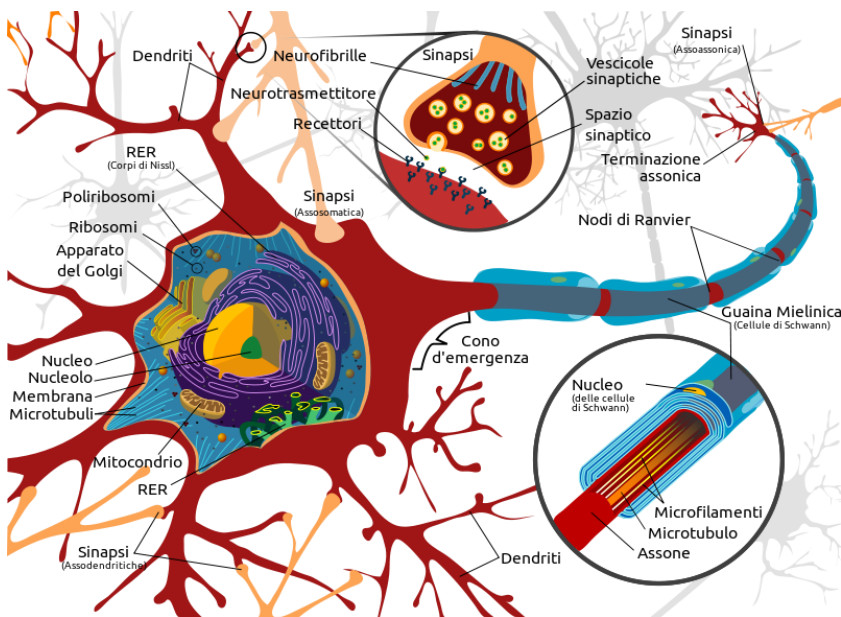


Fig. 4 Un neurone biologico⁹

Le sinapsi costituiscono delle porte di collegamento per il passaggio di informazione tra i neuroni. I dendriti si ramificano dal soma (corpo centrale) del neurone. Attraverso le sinapsi, i dendriti raccolgono input dai neuroni afferenti e li propagano verso il soma. L'assone parte dal soma e si

⁹ https://www.pasquinoni.com/wp-content/uploads/2018/11/800px-Complete_neuron_cell_diagram_it.png

allontana per portare l'output ad altri neuroni. Il passaggio delle informazioni avviene attraverso processi elettro-chimici. La modifica dell'efficacia delle sinapsi è direttamente collegata con processi di apprendimento e memoria. (Schrimpf et al., 2018)

2.2.1) Il neurone

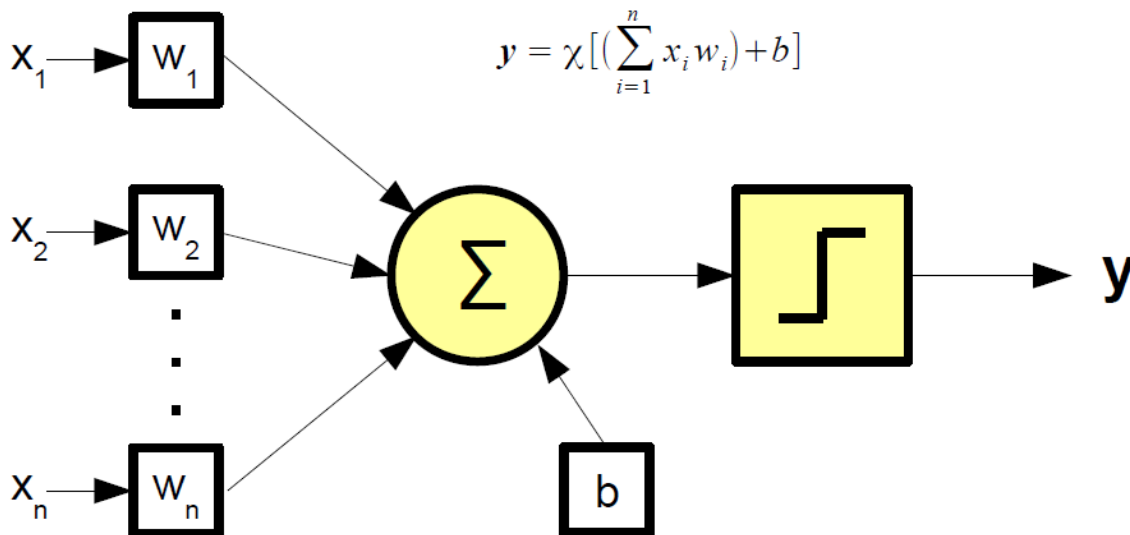


Fig. 5 Neurone secondo McCulloch e Pitz, 1943¹⁰

La figura va interpretata nel seguente modo:

- x_1, \dots, x_n sono gli n ingressi che il neurone riceve da assoni di neuroni differenti;
- w_1, \dots, w_n sono i pesi che determinano l'efficacia delle connessioni sinaptiche dei dendriti e sono i valori che vengono aggiornati durante l'apprendimento;
- b è il *bias*, un ulteriore peso che si considera collegato a un input fittizio e con valore fisso a 1. La sua utilità è quella di "tarare" il punto di lavoro ottimale del neurone; permette di definire la soglia di attivazione del neurone. Più alto è il *bias*, minori saranno i valori degli input necessari ad attivare il neurone e trasmettere l'informazione ai livelli successivi della rete;
- χ è la funzione di attivazione che esprime il comportamento del neurone in funzione del suo livello di eccitazione, che è dato dalla sommatoria (vedi Fig.5) che considera anche il *bias*.

Nei metodi tradizionali, gli input del neurone erano le features, invece con il Deep Learning, la rete neurale è in grado di estrarre le features a partire da un'immagine, della quale i valori di pixel (di RGB, come terna di valori che esprime l'intensità di Rosso, Verde e Blu) sono gli input.

¹⁰ <https://www.gabormelli.com/RKB/images/thumb/3/31/artificial-neuron-model.png/600px-artificial-neuron-model.png>

2.3) DEEP LEARNING

Con il termine DNN (Deep Neural Network) si denotano reti composte da molti livelli (sono “profonde”) organizzati gerarchicamente. Oggi le reti neurali più usate hanno un numero di livelli compreso tra 7 e 50. Il numero di livelli, tuttavia, non è l’unico elemento che caratterizza una *deep neural network*. Alla complessità della stessa partecipa anche il numero di neuroni, connessioni e pesi. Infatti, maggiore è il numero di pesi, maggiore è la complessità di apprendimento. Un elevato numero di neuroni, quindi di connessioni, comporta un maggior costo per la trasmissione delle informazioni in avanti (forward propagation) e indietro (backpropagation). Tra le diverse tipologie di reti neurali si trovano le reti neurali convoluzionali (CNN).

2.3.1) CNN

CNN è un tipo di modello di Deep Learning usato per processare dati strutturati come una matrice (Yamashita et al., 2018), quali le immagini. CNN è tipicamente costituita da tre tipi di livelli: convoluzione, pooling e fully connected. I primi due servono per estrarre le features, mentre il terzo mappa le features estratte nell’output finale. Quest’ultimo livello è un componente fondamentale in una CNN, che include una serie di operazioni, come la convoluzione. Nelle immagini digitali, i valori dei pixel vengono memorizzati in una matrice 2D, e una piccola matrice di parametri, chiamata kernel, viene applicata ad ogni posizione dell’immagine. Il processo di ottimizzazione di tali parametri viene definita training e avviene attraverso algoritmi backpropagation e gradient descent.

Vediamo ora nel dettaglio i livelli inclusi nell’architettura di una CNN, tipicamente formata da una serie di *convolution layers* e un *pooling layer*, seguiti da uno o più livelli *fully connected*.

Il livello di convoluzione contribuisce all’estrazione delle features, che spesso combina operazioni lineari, come la convoluzione, e non lineari, come la funzione di attivazione.

Nella convoluzione, un array di numeri (kernel) viene applicato all’input, che a sua volta è un array di numeri, chiamato tensore. Il kernel viene fatto scorrere sulle diverse posizioni di input e per ognuna di queste viene generato un valore di output, dato dal prodotto scalare tra il kernel e la porzione di input coperta, ottenendo la feature map.

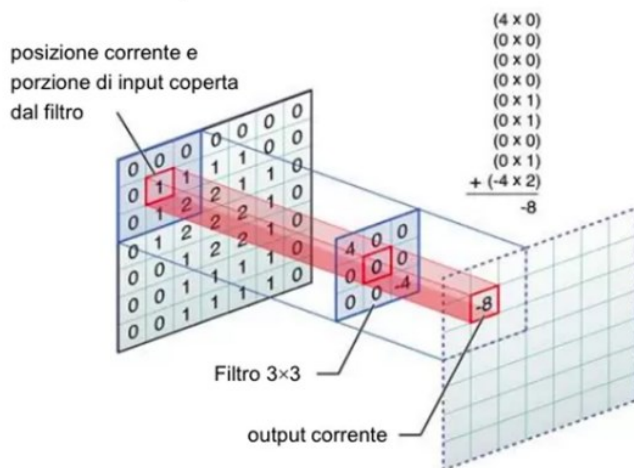


Fig. 6 Esempio del funzionamento della convoluzione in image processing¹¹

La funzione di attivazione non lineare riceve in input i valori di output della convoluzione. Sebbene le funzioni utilizzate maggiormente fossero la sigmoide e la tangente iperbolica in quanto rappresentazioni matematiche del neurone biologico, ora la più comunemente usata è la Rectified Linear units (ReLU), che impiega una funzione di attivazione del tipo:

$$f(x) = \max(0, x)$$

per l'input x .

Nel livello di pooling avviene una riduzione della dimensionalità nel piano delle feature maps, che comporta un decremento del numero di parametri successivi da apprendere. La forma più diffusa di pooling è il Max Pooling che estrae dei percorsi dalle feature maps, fornisce in output il massimo valore in ogni percorso e scarta tutti gli altri valori.

Le feature maps estratte dopo la convoluzione e il pooling, vengono trasformate in un vettore di numeri e connesse a uno o più livelli fully connected, in cui ogni input è connesso con ciascun output attraverso un peso. Una volta che le features sono state estratte attraverso i livelli di convoluzione e sottocampionati con il pooling, esse vengono mappate sugli output finali della rete tramite livelli fully connected, l'ultimo dei quali ha un numero di nodi pari al numero di classi. Ogni livello fully connected è seguito da una funzione non lineare, come ReLU, descritta precedentemente.

¹¹ <https://www.domsoria.com/wp-content/uploads/2019/09/filtroconvoluzione.png>

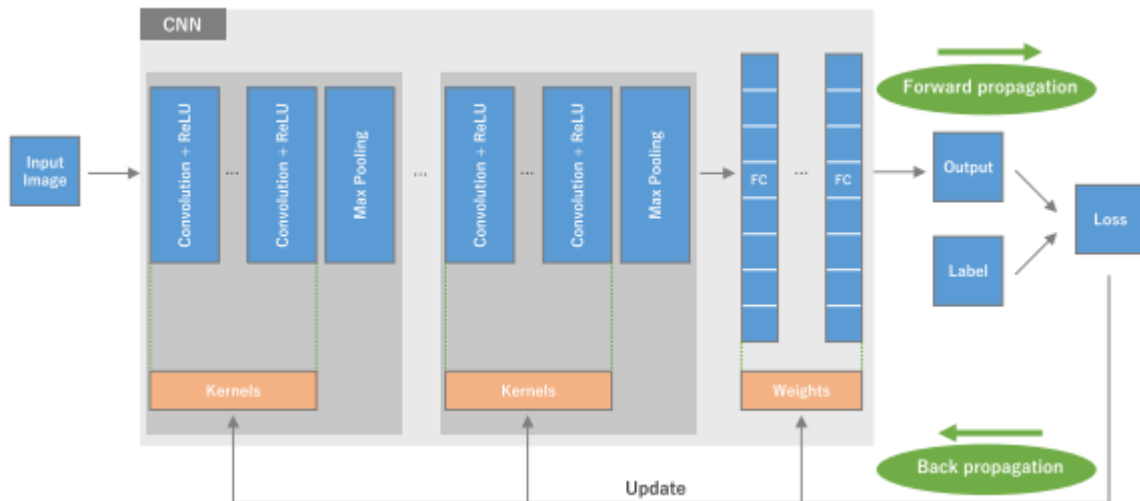


Fig. 7 Struttura di una CNN: una performance del modello con particolari kernel e pesi è calcolata con una loss function attraverso la *forward propagation* su un training set. Kernel e pesi sono aggiornati attraverso *back propagation*.

Una *loss function* misura la compatibilità tra le predizioni sugli output da parte della rete attraverso forward propagation e le date etichette che esprimono la “verità”.

2.3.2) Il filtraggio

Ogni fetta di neuroni opera su una parte di input.

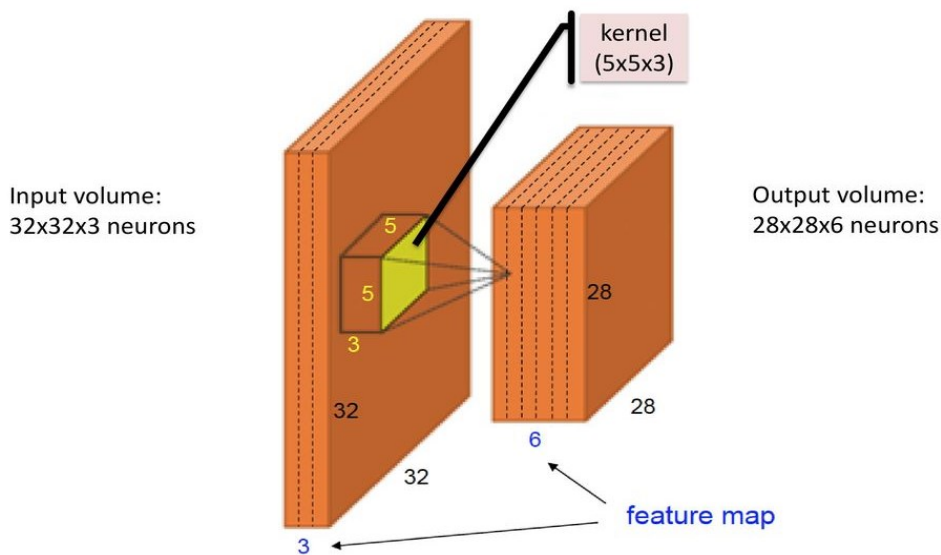


Fig. 8 Filtraggio¹²

¹² <https://slideplayer.com/slide/12213646/72/images/21/Convolution+on+3D+volumes.jpg>

In questo esempio il livello di input è formato da $32 \times 32 \times 3$ neuroni e ogni livello di output, formato da $28 \times 28 \times 6$ neuroni, è connesso a $5 \times 5 \times 3 = 75$ neuroni del livello precedente.

Ogni “fetta” di neuroni rappresenta una feature map. Nell’esempio:

- 3 feature map (dimensione 32×32) del volume di input;
- 6 feature map (dimensione 28×28) del volume di output;

I pesi sono condivisi a livello di feature map. I neuroni di una stessa feature map processano porzioni diverse del volume di input allo stesso modo. Ogni feature map può essere vista come il risultato di uno specifico filtraggio dell’input.

Nell’esempio il numero di connessioni è dato da $28 \times 28 \times 6 \times 5 \times 5 \times 3$ e il numero di pesi da $6 \times (5 \times 5 \times 3 + 1)$, grazie alla condivisione dei pesi. Nell’ultima operazione, 1 indica che nel calcolo si considera la presenza di bias.

Tra le CNN più usate nella rilevazione dell’oggetto ci sono le **R-CNN** e le **Fast R-CNN**.

2.3.3) R-CNN: Data una immagine in input, le Region-CNN (R-CNN) ne estrae le regioni di interesse (ROI) attraverso una ricerca selettiva. Ogni ROI è un rettangolo che può rappresentare il confine di un oggetto nell’immagine. Successivamente ogni ROI passa attraverso una rete neurale di tipo CNN per ottenere l’output. Per ogni ROI la determinazione del tipo di oggetto in esso contenuto avviene attraverso dei classificatori SVM.¹³

2.3.4) Fast R-CNN: Il suo sviluppo inizia da aprile 2015. A differenza della R-CNN, qui la rete neurale computa tutta l’immagine intera. Alla fine viene eseguito un pooling che taglia ogni ROI dal tensore di uscita, lo riforma e lo classifica. Anche qui vengono generate le ROI tramite ricerca selettiva. Nel giugno 2015 si assiste a una modifica: la generazione delle ROI viene integrata nella rete neurale stessa.

2.4) PCA

Un’altra tecnica molto usata in questo ambito è la cosiddetta PCA (Principal Component Analysis) volta a ridurre la dimensionalità, ossia a eseguire un mapping lineare dei pattern, conservando il più possibile l’informazione. Dato un training set di dimensione d , si calcola il vettore medio e la matrice di covarianza. Per un certo k , lo spazio di dimensione k si ottiene dal vettore medio e dalla matrice di proiezione le cui colonne sono formate dagli autovettori che corrispondono ai k più grandi autovalori della matrice di covarianza.

¹³ https://en.wikipedia.org/wiki/Region_Based_Convolutional_Neural_Networks

Nel caso dei veicoli autonomi in cui vengono processate delle immagini, la dimensione d può essere molto elevata e di conseguenza la matrice di covarianza conterrà molti valori, quindi PCA può essere molto utile.

2.5) CNN-SVM

L'architettura di questo modello presenta una rete CNN per l'estrazione delle features e, nella parte finale, un classificatore SVM e non una rete fully connected, come la singola CNN precedentemente analizzata.

2.6) LM-CNN-SVM

Negli ultimi anni è stato introdotto un nuovo modello ibrido, chiamato Local-Multiple CNN-SVM (Uçar et al., 2017), come sistema alternativo alla CNN-SVM singola, con lo scopo di apprendere le features dell'intero oggetto. Il funzionamento può essere descritto come segue:

1. L'intera immagine viene divisa in regioni locali;
2. Viene applicato LM-CNN per l'estrazione delle features delle regioni locali, di dimensione $64 \times 64 \times 3$ successivamente convertite in un'immagine in scala di grigio;
3. L'ultimo livello di output della CNN viene sostituito con un classificatore di tipo SVM perché la CNN ha una performance di generalizzazione più bassa rispetto a SVM;
4. Il livello fully connected della CNN viene espresso come combinazione lineare degli output dei precedenti livelli intermedi con i pesi e il termine bias. Inoltre, l'output di questo livello viene dato in input all'ultimo della CNN, nello specifico dello studio citato, ogni regione locale viene importata in AlexNet;
5. La CNN fornisce le probabilità della classe per ogni immagine di input utilizzando la funzione di attivazione Softmax. Comunemente, infatti, nelle moderne CNN usate come classificatori, viene impiegato un livello finale Softmax, così descritto:

- È costituito da s neuroni (uno per classe) fully connected

La funzione di attivazione (in funzione del livello di attivazione net_k) è:

$$f(net_k) = \frac{e^{net_k}}{\sum_{c=1}^s (e^{net_c})}$$

I valori di questa funzione possono essere interpretati come probabilità, in quanto appartengono all'intervallo $[0,1]$ e la loro somma è 1.

6. Le features ottenute dal livello fully connected vengono salvate;
7. La dimensione delle features salvate viene ridotta tramite PCA;
8. SVM viene impiegato per ridurre il numero di features.

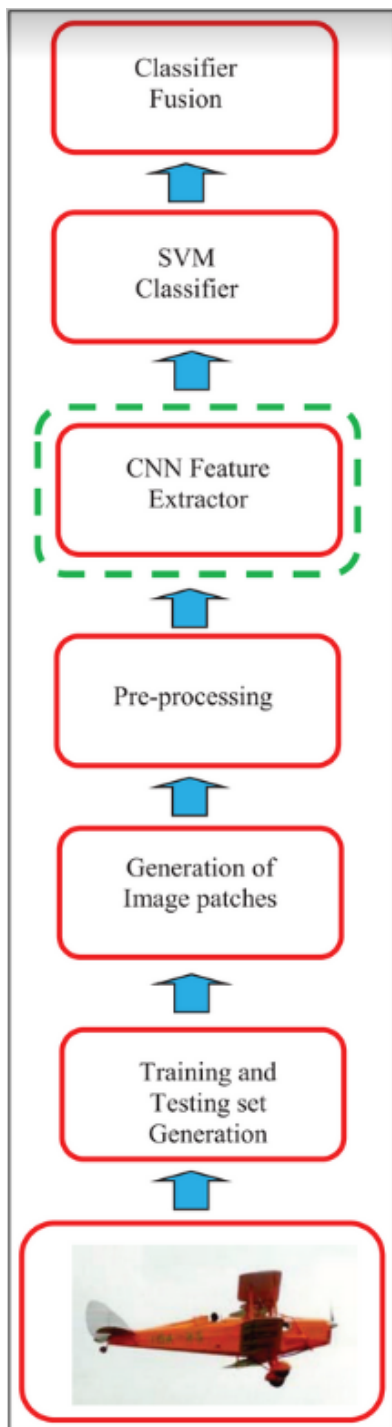


Fig. 9 schema illustrativo dell'algorithmo LM-CNN-SVM

Type	Filter-stride	Feature maps
Input	227×227	1
Convolution 1	$11 \times 11-4$	96
Max Pooling 1	$3 \times 3-2$	96
Convolution 2	$5 \times 5-1$	256
Max Pooling 2	$3 \times 3-2$	256
Convolution 3	$3 \times 3-1$	384
Convolution 4	$3 \times 3-1$	384
Convolution 5	$3 \times 3-1$	256
Max Pooling 5	$3 \times 3-2$	256
Fully connected 6	$6 \times 6-1$	4096
Fully connected 7	1×1	4096
Fully connected 8	1×1	1000

Fig. 10 Alex Net

Lo studio proposto nell'articolo (Uçar et al., 2017), oltre a proporre questo nuovo sistema ibrido per l'apprendimento e la classificazione, si occupa di valutare le performance dello stesso nel rilevare un oggetto. A tal fine i datasets utilizzati sono stati Caltech-101 e Caltech pedestrian. Caltech-101 è formato da 101 categorie di oggetti. Ogni classe ha un numero variabile di immagini compreso tra 31 e 800. Il database include sia oggetti rigidi come aeroplani, macchine, bici, sedie, sia oggetti non rigidi, come pecore, mucche e leoni. Per costruire il training set sono state selezionate casualmente dalle quindici alle trenta immagini per categoria e non più di cinquanta immagini per categoria per il test set. Ogni immagine è stata divisa in nove regioni locali della dimensione di $64 \times 64 \times 3$, poi convertite in scala di grigio. Tutte le immagini sono state localizzate come array in una matrice. Lo Stochastic Gradient Descent applicato prevede che i batch abbiano dimensione 30 e il tasso di apprendimento per i pesi e i bias sia rispettivamente di 0.001 e 0.02. Successivamente sono state estratte le features e poi applicata PCA agli output del livello fully connected finale ottenendo delle features ridotte che vengono date in input al classificatore SVM lineare, i cui output sono stati fusi, utilizzando majority voting rule, una regola di decisione che seleziona la classe che ha ottenuto "più voti". L'esperimento è stato ripetuto dieci volte con immagini di training e di test diverse. Le performance nel nuovo sistema sono migliori rispetto a quelle della singola CNN-SVM. Le features estratte dalla CNN fanno sì che il classificatore SVM sia più accurato.

Il dataset Caltech pedestrian è stato costruito su undici sessioni video fatte da un veicolo che circola nel traffico regolare in un ambiente urbano. Le prime cinque sono state usate per il training set e le restanti per il test set. Le immagini dei pedoni utilizzate sono più grandi di 50 pixel e almeno il 65% delle loro parti del corpo è visibile. Le immagini sono state divise in sedici regioni locali per una corretta rilevazione del pedone dovuta alla piccola dimensione dello stesso. Successivamente sono state rimosse le immagini che non comprendevano pedoni. Dalla valutazione delle performance, che usa come metrica il miss rate medio (in %), di LM-SVM-CNN e CNN-SVM singola sul dataset

Caltech pedestrian, si osserva che il sistema proposto è di gran lunga migliore, con valori di accuratezza significativamente alti, ossia intorno all'89 e 92, rispettivamente per 15 o 30 immagini per classe. Con questo sistema sembrano risolti gli errori nel riconoscimento dell'ostacolo in generale causati da condizioni di scarsa luminosità, ombre e nebbia, che possono provocare incidenti anche gravi. Nella figura sottostante sono riportati graficamente tali risultati.

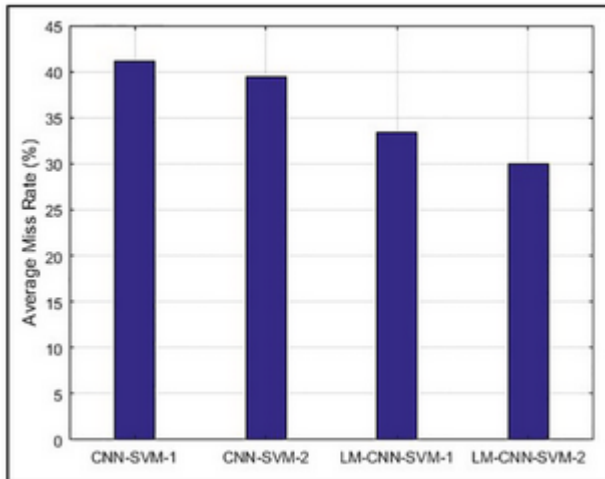


Fig.11 Confronto tra CNN-SVM singola e LM-CNN-SVM in termini di miss rate per la valutazione delle performance

2.7) DEEP LEARNING APPLICATO AI VEICOLI AUTONOMI

Il sistema del veicolo autonomo richiede (Galvao et al., 2021):

- **Percezione**, che fa sì che il VA possa agire in modo sicuro, efficiente e affidabile. Questo “modulo” utilizza i dati derivanti dall’ambiente circostante al veicolo e acquisiti attraverso sensori e dispositivi di comunicazione, che hanno il compito di rilevare, ad esempio, le condizioni della strada. Per fare questo, i recenti veicoli autonomi sono stati sviluppati in due metodi. Tesla, infatti, implementa un sistema che combina l’approccio vision-based puro con la fusione di vari tipi di sensori. Sul cruscotto del veicolo viene montata una telecamera monoculare. Il sistema di percezione del veicolo deve rilevare oggetti statici e non, riconoscere l’informazione fornita dagli oggetti statici e predire il comportamento di quelli non statici. Waymo usa, invece, tecniche di computer vision e sensori;
- **Planning**: una volta percepito l’ambiente, il VA deve decidere cosa fare. Solitamente viene dapprima assegnato un obiettivo (mission), poi si considerano l’interazione con gli altri agenti del traffico e le regole che il veicolo deve rispettare per capire quali sono le azioni che può/dovrebbe fare (behaviour) e infine il motion planner ha il

compito di generare dei percorsi che il VA può fare, possibilmente senza collidere. Questo stadio di planning di solito viene implementato utilizzando tecniche tradizionali come il diagramma di Voronoi¹⁴, che però non è adatto a complessi scenari del traffico e della strada dove è rilevante l'interazione tra il veicolo e gli agenti del traffico. Vengono perciò utilizzate tecniche di machine learning (ML) in particolare CNN e Deep Learning.

- **Act:** in questo modulo avviene l'azione del VA che si realizza attraverso l'invio di segnali di comando agli attuatori dopo che le informazioni raccolte nel planning hanno superato la fase di controllo.

Le tecniche di Deep Learning si sono sviluppate molto negli ultimi anni, superando di gran lunga le tecniche di computer vision tradizionali, soprattutto per quanto riguarda la classificazione e la rilevazione di oggetti. Nel campo dei veicoli a guida autonoma è molto importante capire quali sono le tecniche di Deep Learning applicate, soprattutto nel riconoscimento del veicolo e del pedone.

La rilevazione del veicolo è fondamentale perché, in seguito alla stessa, il VA deve decidere quali azioni intraprendere, quale comportamento adottare. Perciò essa deve essere robusta, veloce, accurata e a basso costo. Per rilevare un veicolo, vengono usate immagini che derivano da telecamere di sorveglianza, piuttosto che telecamere montate sul VA, immagini satellitari. Supponendo di considerare solo veicoli che circolano su strada e immagini derivanti da telecamera montata sul VA, è necessario porre l'attenzione a diversi fattori, principalmente su:

- La classe cui il veicolo appartiene: macchina, bus, camion o altro;
- Veicoli che appartengono alla stessa classe possono avere forma diversa, piuttosto che differente colore, struttura e dimensione;
- Il punto di vista da cui il veicolo viene visto, di fronte, di lato, da dietro;
- Le condizioni meteo variabili;

Per quanto riguarda i pedoni e i ciclisti, essi sono gli utenti della strada più vulnerabili, quindi un veicolo autonomo deve essere in grado di riconoscerli evitando di collidere e lo può fare solo attraverso l'uso di sistemi molto accurati. La rilevazione del pedone rappresenta, di fatto, una sfida per il VA in quanto:

- Il pedone può vestirsi in diverso modo e con colori diversi;
- I pedoni non sono tutti alti uguali;
- La rilevazione del pedone diventa difficile in presenza di illuminazione molto variabile;

¹⁴ https://it.wikipedia.org/wiki/Diagramma_di_Voronoi

- In un'immagine possono esserci più pedoni su scale diverse. Qui il problema è che i pedoni sullo sfondo sono difficilmente riconoscibili;
- I pedoni possono cambiare direzione velocemente.

Molti studi, tra cui (Uçar et al., 2017), (Galvao et al., 2021), (Fang & Cai, 2021), (Hasan et al., 2022), che hanno affrontato e tuttora affrontano questo tema fanno uso di tecniche di Deep Learning, soprattutto per la loro flessibilità che consiste nel fatto che i modelli di DL possono essere riaddestrati su dataset differenti.

Capitolo 3: BIAS ALGORITMICI

Nel capitolo precedente, sono state presentate le principali tecniche di Deep Learning che consentono al VA di rilevare la presenza di ostacoli, tra cui in particolare persone nell'ambiente circostante allo stesso. Si tratta di tecniche via via sempre più diffuse e che sono costantemente oggetto di studio da parte dei ricercatori, che si pongono l'obiettivo di proporre interventi migliorativi delle stesse in termini di performance e robustezza. Tuttavia, nel campo della guida autonoma, si riscontrano spesso dei *bias* legati al mancato riconoscimento del pedone che hanno delle conseguenze anche piuttosto gravi.

Il termine *bias* si riferisce ai casi in cui le decisioni prese da un sistema autonomo e dotato di IA dimostra dei pregiudizi nei confronti di gruppi di persone definite in base a certi attributi, come genere e razza.

La presenza di un *bias* è legata principalmente all'incapacità dell'algoritmo di generalizzare, ovvero di avere delle buone performance quando viene testato su dataset diversi. Il problema è che ciò accade anche quando il Test Set si avvicina molto al dataset su cui l'algoritmo è stato addestrato. Secondo (Hasan et al., 2021), la sorgente di tali *bias* va ricercata nel dataset, in particolare nel modo in cui è stato costruito, quali sono le immagini che lo formano, in quanto l'impiego del Deep Learning implica che l'accuratezza di un algoritmo dipende pesantemente dalla quantità e dalla qualità del dataset.

3.1) BIAS NEI SISTEMI DI RICONOSCIMENTO DEI PEDONI

Nel caso specifico di un sistema atto a rilevare la presenza di pedoni sulla strada, le tre principali limitazioni dei dataset sono:

- Limitato numero di soli pedoni;
- Bassa densità dei pedoni, cioè sono poche, se non rare, le situazioni di affollamento presentate;
- Poca diversità fra i pedoni stessi.

Lo studio sopra citato si occupa di dimostrare la presenza di tali *bias* attraverso un esperimento, descritto di seguito.

3.1.1) L'esperimento

Datasets di riferimento:

- **Caltech**, formato da 13 mila persone estratte da 10 ore di video ripreso da un veicolo a Los Angeles, USA;
- **CityPersons**, dataset differente rispetto a Caltech perché le riprese sono state effettuate in 27 città diverse della Germania e degli Stati limitrofi; il dataset comprende circa 31 mila

bounding box. Il *training*, *validation* e *test set* sono formati rispettivamente da: 2975, 500 e 1575 immagini;

- **EuroCity Person (ECP)** è un nuovo dataset per la rilevazione del pedone ritenuto di gran lunga migliore rispetto ai precedenti in termini di varietà e difficoltà. L'insieme di immagini che lo formano sono state raccolte in 31 città diverse di 12 Stati europei differenti. Ciò che lo rende interessante è la presenza di immagini in condizioni di luce differenti, ossia raffiguranti sia scene diurne che notturne.
- WiderPedestrian e CrowdHuman, che però non sono particolarmente adatti all'addestramento di un sistema di rilevazione in un veicolo autonomo.

Protocollo di valutazione: secondo i protocolli di Caltech, CityPerson, e ECP, largamente adottati nel campo della ricerca, nella valutazione viene confrontato il *log average miss rate* con i Falsi Positivi Per Immagine (FPPI). Il primo si riferisce agli oggetti che non sono stati riconosciuti ed è dato dal rapporto tra il numero dei Falsi Negativi e quello dei Positivi totali (veri e falsi). Il secondo è definito come il rapporto tra il numero dei Falsi Positivi e il totale dei frame testati.

	1	2	3	4	5
Pedone nella realtà	SI	NO	SI	SI	SI
Pedone riconosciuto	NO	SI	SI	SI	NO
	FN	FP	TP	TP	FN

LEGENDA:

FN= FALSO NEGATIVO: il pedone c'è ma non viene riconosciuto

TP=TRUE POSITIVE (VERO POSITIVO): il pedone c'è e viene riconosciuto

FP=FALSO POSITIVO: il pedone non c'è sulla strada ma ne viene comunque rilevata la presenza.

Il log-average miss rate è pari a 2/3 quindi 66% e i FPP risulta uguale a 1/5 cioè il 20%.

I metodi di rilevazione sono stati confrontati usando impostazioni di lavoro similari.

Nella tabella sottostante sono riportati dei numeri per i diversi livelli di occlusione, ossia: Reasonable, Small, Heavy, Heavy* e All che considerano l'altezza e la visibilità del bounding box contenente il pedone, secondo valori tipici del sensore LIDAR e fissati dagli studiosi.(Schiele et al., 2017)

Valutazione del cross-dataset: In questa fase, si indica con A → B un modello addestrato SOLO sul dataset A e testato su B.

Riferimento: Come metodo di riferimento si usa Cascade R-CNN che rappresenta un'estensione della R-CNN e che contiene molte teste di rilevamento in una sequenza, che progressivamente cercano di filtrare i falsi positivi. La rete scelta è HRNet, il cui vantaggio è dato dal fatto che mantiene le feature maps ad alta risoluzione.

Nel caso di Caltech e CityPersons, il riferimento Cascade R-CNN ha delle performance comparabili con gli algoritmi in uso. Tali performance trovano un miglioramento rispetto a metodi in cui il dataset ha una dimensione maggiore. Sono le peggiori sul più piccolo dataset e le migliori sul più grande.

Capacità di generalizzazione. Per vedere quanto gli algoritmi allo stato dell'arte riescono a generalizzare, è stata fatta una valutazione delle performance su cinque rilevatori di pedone e Cascade R-CNN (che non è più il riferimento ma oggetto di valutazione) sui dataset CityPerson e Caltech. Sono stati valutati i recenti alcuni metodi recentemente proposti, quali BGC-Net, CSP, PRNet, ALFNet, FRCNN (adattati per il riconoscimento del pedone). Insieme a Cascade R-CNN, per cui viene fissato il riferimento HRNet, è stato aggiunto Faster R-CNN con il riferimento ResNext-101, una struttura più "portante". Un altro metodo testato è FRCNN, la cui struttura di riferimento è VGG-16, non adattata per rilevazioni specifiche dei pedoni, quali sono la scalatura dell'input e la gestione delle regioni ignorate.

Nella tabella sono riportati i seguenti dati:

Method	Bakbone	CityPersons→CityPersons	CityPersons→Caltech	Caltech→Caltech	Caltech→CityPersons
FRCNN [50]	VGG-16	15.4	21.1	8.7	46.9
Vanilla FRCNN [50]	VGG-16	24.1	17.6	12.2	52.4
ALFNET [26]	ResNet-50	12.0	17.8	6.1	47.3
CSP [27]	ResNet-50	11.0	12.1	5.0	43.7
PRNet [37]	ResNet-50	10.8	10.7	-	-
BGCNet [22]	HRNet	8.8	10.2	4.1	41.4
Faster R-CNN [34]	ResNext-101	16.4	11.8	9.7	40.8
Cascade R-CNN [8]	HRNet	11.2	8.8	6.2	36.5

- Sulla prima colonna il metodo di cui si vuole valutare la performance;
- Sulla seconda colonna il riferimento per ogni metodo, come descritto sopra (es. FRCNN ha riferimento VGG-16);
- Sulla terza colonna il risultato ottenuto quando il metodo viene addestrato e testato su CityPersons;
- Sulla quarta colonna il risultato ottenuto quando il metodo viene addestrato su CityPersons e testato su Caltech;
- Sulla quinta colonna il risultato ottenuto quando il metodo viene addestrato e testato su Caltech;

- Sulla sesta colonna il risultato ottenuto quando il metodo viene addestrato su Caltech e testato su CityPersons;

Si può notare che tutti i metodi presentano un calo delle performance quando sono addestrati su CityPersons e testati su Caltech. Nel caso di Cascade R-CNN le performance rimangono comparabili con il modello addestrato e testato sul set di destinazione. Dal momento che CityPersons è diverso da Caltech, questo peggioramento delle performance non può essere dovuto alla scala del dataset e alla densità della folla. Ciò è una prova del fatto che i sistemi di rilevazione di oggetti in generale hanno una maggiore capacità di generalizzazione rispetto ai sistemi di rilevazione del pedone attuali. Analogamente, dalla tabella sopra si vede che altri metodi hanno migliori performance se addestrati su certi dataset e testati su altri.

3.1.2) Interpretazione dei risultati e motivazione

Questo rappresenta di fatto un *bias* per i sistemi di riconoscimento del pedone dello stato dell'arte che hanno una capacità di generalizzazione parecchio limitata. Inoltre, è molto importante notare che Cascade R-CNN ha delle performance inferiori rispetto agli altri metodi. Ciò è dovuto al fatto che Caltech non presenta immagini molto diverse fra loro e non sono raffigurate folle di persone e ciò giustifica la sua incapacità di generalizzazione. Si tratta nuovamente di un problema nella formazione del dataset. Tuttavia un'altra causa potrebbe essere ricercata nel fatto che spesso i sistemi di riconoscimento di un oggetto si adattano ai sistemi di riconoscimento del pedone, il che pone dei limiti rilevanti alla capacità di generalizzazione. Finora sono stati considerati *bias* correlati con la capacità di generalizzazione molto scarsa, la cui causa va di nuovo ricercata nella formazione del dataset, affermazione che è provata scientificamente anche da altri studi, come (Fabbrizzi et al., 2022), (Danks & London, 2017).

3.2) BIAS DI ETÀ' E GENERE

Analizzando nel dettaglio i *bias*, si vede che essi possono essere di tipo diverso. Diversi studi con l'obiettivo di approfondire il concetto di *bias*, come mancata rilevazione del pedone, hanno messo in luce la presenza di *bias* specifici, per esempio di età e genere, dando ampio spazio a un dibattito su questioni di tipo etico.

3.2.1) Il caso della Francia

Secondo (Brandao, 2019) diversi sistemi di riconoscimento presentano un *bias* nel rilevare la presenza di bambini (persone di età compresa tra 0 e 14 anni). (Brandao, 2019) sviluppa l'analisi

utilizzando come dataset INRIA Person Dataset, costituito da immagini di pedoni sulla strada e altri scenari urbani. Il test set di INRIA Person è stato esteso con etichette “child” o “adult” e “male” o “female”. (Brandao, 2019) si occupa di analizzare la presenza di tali *bias*, considerando diversi metodi disponibili nel Caltech Pedestrian benchmark e valutati solo sui “*bounding box*” etichettati. La metrica per la valutazione dell’accuratezza nella rilevazione è il miss rate medio (in %). Nei grafici sottostanti sono riportati i risultati e in particolare sull’asse orizzontale sono indicati i metodi e sull’asse delle ordinate i valori di miss rate. Dal grafico 1 si osserva che, fatta eccezione per tre metodi, ovvero PoseInV, CJ, FtrMine, il miss rate nella rilevazione di bambini è più elevato, cioè la gran parte dei metodi compie degli errori nella classificazione di pedoni di giovane età. Il grafico 2, invece, suggerisce che la maggior parte dei metodi hanno dei problemi nel riconoscere individui di sesso femminile, anche se non significativamente. Si precisa che (Brandao, 2019) non considera in questa analisi delle performance la combinazione genere-età ma mantiene separate queste classi.

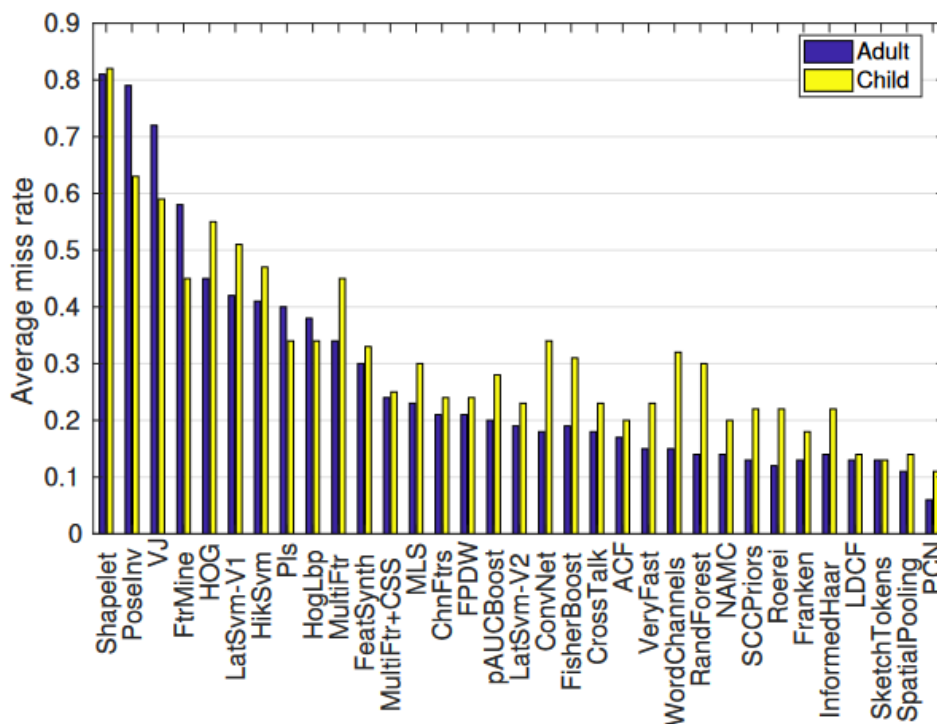


Grafico 1: Confronto miss rate nella classificazione tra “adult” e “child” tra i vari metodi

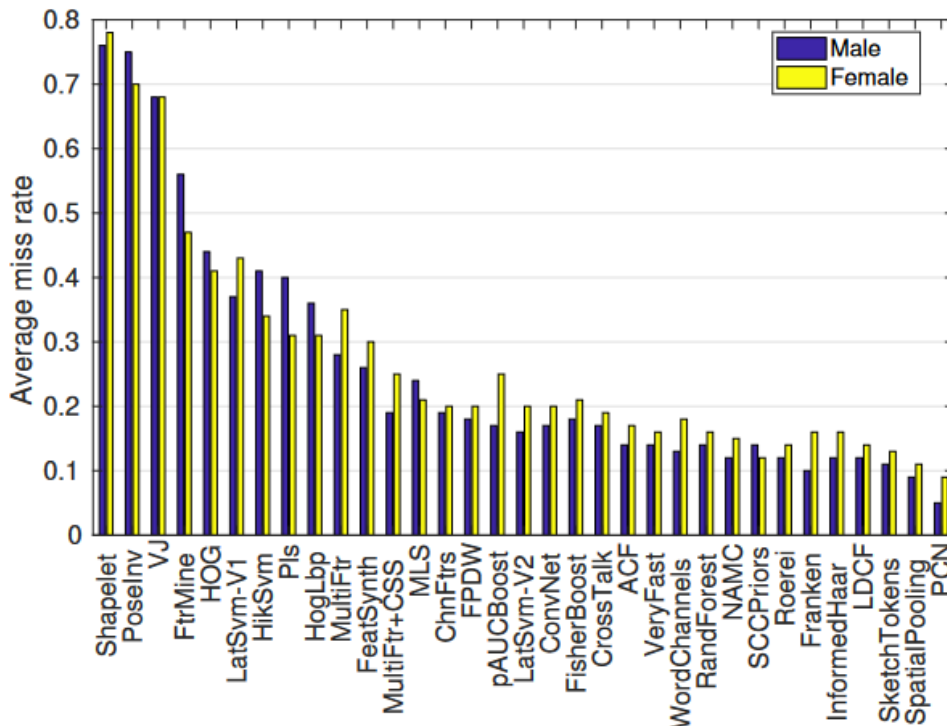


Grafico 2: Confronto miss rate nella classificazione tra “male” e “female” tra i vari metodi

Osservando in modo più dettagliato questi grafici, si vede che i miss rate sono più alti nel riconoscimento di bambini e ciò lo riportano 27 metodi su 33 totali che sono stati considerati, cioè l’82%. Questo calo delle performance può essere numericamente tradotto considerando che il miss rate per i bambini è 1.3 volte più grande se la valutazione viene fatta su tutti i metodi, 1.4 volte maggiore analizzando solo i 24 metodi più performanti.

Come detto precedentemente, per quanto riguarda l’analisi fatta sul genere, la differenza tra il miss rate nella rilevazione di individui di sesso femminile e maschile c’è, anche se non troppo grande. Tuttavia, 24 metodi su 33, cioè il 72% dei metodi, non riescono a classificare correttamente individui “female” e ciò rappresenta un vero e proprio *bias* di performance.

Ciò che desta particolare interesse è il fatto che i 24 metodi con le migliori performance presentano *miss rates* più elevati per i bambini. Secondo (Brandao, 2019), questo è dovuto a un *overfitting* dei metodi, ovvero a un loro adattamento ai dati del *training set* che non implica migliori prestazioni degli stessi quando vengono testati sul Test Set.

3.2.2) La sorgente dei bias

La causa della presenza di *bias* di età va ricercata ancora una volta nella costituzione del dataset, INRIA Person, le cui immagini ritraggono scenari francesi e pedoni che vivono in Francia, dove, nel 2018, il 18% degli abitanti è coperto da bambini/e e ragazzi/e nella fascia di età 0 -14 anni, che al momento delle riprese si trovavano a scuola e quindi non sono presenti nelle immagini del dataset.

Un'altra causa può essere il fatto che i “child” occupano un'area più piccola dell'immagine rispetto agli adulti e come già detto gli algoritmi di riconoscimento attuali hanno diverse difficoltà nel rilevare *bounding box* di dimensione più piccola, oltre al fatto che sono molto più veloci nel muoversi e cambiare direzione. Infatti solitamente corrono, saltano, giocano.

(Ntoutsi et al., 2020) conferma quanto appena affermato, identificando due modi in cui i *bias* possono trovarsi all'interno dei dati. Il primo è collegato con le correlazioni e influenze casuali sugli attributi e le caratteristiche dei soggetti. Il secondo riguarda la mancanza di rappresentazioni di gruppi “protetti”.

3.3) TIPI DI BIAS NEI DATI

I *bias* all'interno dei dati possono manifestarsi come:

- *Selection bias*: qualsiasi disparità o associazione creata come risultato del processo mediante il quale i soggetti sono inclusi in un set di dati visivo;

I sistemi di guida autonoma possono essere influenzati da *selection bias*, dal momento che, ad oggi, la costruzione di un dataset completo, accurato, contenente immagini e riprese che possano coprire tutti gli scenari del traffico con tutte le condizioni di luce e ombre, rimane una sfida; come già anticipato, se un sistema di riconoscimento viene addestrato su dataset che contengono immagini di una sola città o di poche città con scenari molto simili, è alquanto improbabile che possa avere delle buone prestazioni quando viene testato su dataset diversi e contenenti immagini di città molto diverse da quelle dell'addestramento e raffiguranti pedoni che differiscono nel modo di vestire, piuttosto che nel fatto che si trovino da soli o in gruppo. Questa influenza dei *selection bias* nei veicoli a guida autonoma è molto rischiosa, dal punto di vista delle conseguenze, soprattutto quando ad essere coinvolti sono i sistemi di rilevazione del pedone. Un esempio di *selection bias* sono quelli di “razza”, come affermato da (Wilson et al., 2019) che ha evidenziato una sottorappresentazione di individui di pelle scura già ribadito in (Buolamwini & Gebru, 2018);

- *Framing bias*: qualsiasi associazione o disparità che può essere usata nella diffusione di messaggi e che può essere ricondotta al modo in cui è stato composto il contenuto visivo; un esempio è dato dagli annunci pubblicitari, in cui parti del corpo di una donna sono utilizzate spesso al fine di pubblicizzare un prodotto, provocando un effetto desumanizzante sulla donna stessa;

- *Label bias*: qualsiasi errore nell'etichettatura dei dati visivi, rispetto a qualche verità fondamentale, o l'uso di categorie semantiche mal definite o inadeguate.

3.4) CONSEGUENZE ETICHE DEI BIAS

Il mancato riconoscimento di un ostacolo, nel caso specifico di un pedone o di un ciclista, può comportare situazioni di rischio elevato per l'incolumità della persona, ma anche di difficile studio dal momento che diventa poco immediata l'attribuzione della responsabilità, visto che ad un veicolo non è possibile conferire alcun tipo di colpa secondo i canoni riservati alle persone. Il veicolo, in questi casi, si trova di fronte a una cosiddetta **collisione inevitabile** e ad una scelta da fare.

Ecco che la combinazione tra etica e diritto cerca di dare una risposta a queste importanti questioni.

3.4.1) Le collisioni inevitabili

La collisione inevitabile rappresenta una situazione in cui il veicolo autonomo non può più evitare l'ostacolo. Tuttavia potrebbe avere il tempo per scegliere l'azione da compiere. Nonostante gli algoritmi di riconoscimento dell'ostacolo siano in continuo e incrementale sviluppo, essi presentano *bias* piuttosto importanti che possono causare tali situazioni, già riscontrate di recente. Si ricordano i sinistri in cui sono stati coinvolti sia i VA Tesla che le Google Cars, oltre ad un caso avvenuto in Arizona, dovuto al fatto che il collaudatore della società Uber di VA non è riuscito a riprendere il controllo del veicolo in tempi sufficienti da evitare la ciclista.

Secondo (Tamburrini et al., 2021) il problema della collisione inevitabile trova la sua origine nell'intersezione tra delega e autonomia operativa. Infatti, a mano a mano che il livello di autonomia del veicolo cresce nella scala L1-L5, il guidatore-passeggero conferisce al veicolo sempre maggior autonomia operativa, che gli consente di non dover necessariamente intervenire nelle azioni svolte. Queste azioni però non sono solo del tipo parcheggiare, proseguire su una strada o scegliere di girare a destra piuttosto che a sinistra, ma anche scelte che hanno delle implicazioni a livello etico e morale che coinvolgono persone, le quali possono essere sia il passeggero, o i passeggeri a bordo, sia gli altri utenti della strada. Il veicolo potrebbe scegliere la manovra migliore per la protezione dei suoi passeggeri, compiendo delle azioni più velocemente rispetto a un conducente umano che può essere distratto, stanco e con riflessi più lenti. La valutazione morale riguarda azioni ancora da compiere automaticamente e per mezzo della tecnologia e dei sistemi a disposizione in essa compresi. Non è dunque da chiedersi se la scelta

fatta sia stata giusta o sbagliata, sviluppando quindi una analisi post evento, ma va fatta una valutazione di tipo predittivo, con lo scopo di scegliere l'azione, o le azioni, da compiere. Ciò non è sicuramente così immediato. Si parla di "dilemma morale" (Tamburrini, 2020). Per farne un esempio pratico, si considerano due scenari, S1 e S2, che somigliano al caso del carrello ferroviario Trolley Problem (Thomson, 1985) se pur con delle differenze. Infatti, nel caso del carrello non si pone il problema dell'iterazione della scelta e dei suoi effetti sulle scelte successive. Nel primo scenario ci sono due ciclisti, uno con casco l'altro senza, che attraversano la strada in direzione ortogonale rispetto al VA. Nell'ipotesi che il VA si trovi in uno stato di collisione inevitabile e abbia ancora tempo per decidere chi investire, quale scelta dovrà fare il VA? Nel secondo scenario il veicolo autonomo, con a bordo due passeggeri, una donna con il/la figlio/a che devono arrivare a scuola, si trova in stato di collisione inevitabile nei confronti di un gruppo di pedoni che stanno attraversando la strada. Il VA dovrà scegliere se sterzare in maniera brusca per cercare di evitare i pedoni, mettendo però a rischio la vita dei passeggeri a bordo, oppure investire i pedoni cercando di proteggere l'incolumità dei suoi passeggeri.

3.4.2) Etica delle conseguenze

L'etica delle conseguenze si focalizza sulle ricadute che una decisione o scelta del VA può avere a livello etico e morale, non tenendo in considerazione il rispetto di alcuni doveri. Assume un ruolo importante all'interno dell'indagine etica e filosofica, che viene messa in atto quando si cerca di dare una risposta a domande del tipo: "Cosa deve fare un VA?".

È necessario, a questo punto, capire in base a quali criteri una conseguenza possa essere giudicata moralmente accettabile o meno. Uno di questi criteri può essere il benessere collettivo e supporre che il benessere di ogni utente della strada abbia lo stesso peso. Volendo applicare quest'etica allo scenario 1, verrebbe subito da dire che la scelta migliore da fare, per minimizzare il danno, è investire il ciclista con casco. Questa rappresenta sicuramente la soluzione più accettabile se si valuta il singolo caso, l'evento unico. D'altra parte, considerando un arco temporale più ampio, potrebbe venir meno la regola sull'uso del casco. Infatti, se ogni volta che si verifica una situazione del genere viene investito il ciclista con il casco, si correrebbe il rischio che le persone non lo indossino più perché si convincono che rappresenta un incremento di rischio e la regola potrebbe essere meno rispettata. Tutto ciò porta a conseguenze tutt'altro che banali, dal momento che si assisterebbe a una riduzione della popolazione con casco, con danni dal punto di vista del benessere collettivo. Ecco perché si parla di "conseguenzialismo delle regole", ossia quell'approccio secondo cui l'azione va fatta non solo in base alle conseguenze ma anche alle

regole, attraverso l'applicazione delle quali si hanno le migliori conseguenze dal punto di vista del benessere collettivo, e di "conseguenzialismo dell'atto" che, invece, richiede di prendere la decisione sull'azione da compiere tenendo in considerazione unicamente le conseguenze del singolo atto. Secondo quanto detto, quindi, il VA dovrebbe investire il ciclista con casco se si considera un orizzonte di tempo ristretto, altrimenti non sempre dovrà investire quello con casco. Simili osservazioni possono essere fatte per lo scenario 2. Considerando un ristretto orizzonte temporale, la scelta del VA dovrebbe essere quella di evitare di investire i pedoni per una minimizzazione dei danni "del singolo momento". Ampliando l'orizzonte temporale sorgono dei dubbi se questa sia davvero la decisione migliore da prendere. È vero che l'alternativa sarebbe mettere a repentaglio la vita del passeggero, scelta che a livello etico sarebbe molto influente, oltre a destare non pochi dubbi a potenziali acquirenti sull'acquisto di un VA, che si trovano a riflettere sulle promesse di questa nuova tecnologia che presenta comunque dei limiti. Queste perplessità hanno la loro influenza anche a livello economico, perché potrebbero compromettere gli obiettivi commerciali delle case automobilistiche.

3.4.3) Etica dei doveri

L'etica delle conseguenze traslascia, come già detto, il rispetto o la violazione di un dovere nel momento in cui una scelta viene fatta. In questo paragrafo viene affrontato l'argomento dell'etica dei doveri. Per introdurlo si consideri il caso di un genitore che porta il/la figlio/a a scuola, usando come mezzo di trasporto un VA e la cui priorità è la protezione del/della figlio/a durante il trasporto casa-scuola. Questa priorità per il genitore è un dovere che va rispettato senza tenere conto delle conseguenze, quindi senza l'obiettivo di minimizzare i danni a persone coinvolte, in caso di collisione inevitabile. L'etica dei doveri, dunque, ha il primario obiettivo di fare ciò che è giusto in base al rispetto di certi doveri (regole). In questo caso specifico, il dovere è un dovere di ruolo, quello del genitore che vuole proteggere il/la proprio/a figlio/a. Altri doveri di ruolo si ritrovano in ambito lavorativo, professionale, dove ogni persona ha un ruolo, una posizione, dei compiti da rispettare nel suo campo di occupazione, più in generale nel mondo del lavoro. Si può dire che secondo l'etica dei doveri, nello scenario 2 il veicolo dovrebbe scegliere di investire i pedoni. A questo punto la deontologia (etica dei doveri è detta anche etica deontologica) pone al centro della riflessione un conflitto tra concreto e astratto, tra la dimensione dei ruoli e la razionalità umana. Ciò lo si può rilevare se si considera che, se da una parte un genitore non potrebbe mai perdonarsi di aver sacrificato la vita del/della proprio/a figlio/a nel tragitto da casa a scuola, dall'altra si assiste alla negazione dei diritti degli utenti della strada che potrebbe

provocare lo scatenarsi di una serie di tensioni sociali, prima di tutto perché se la progettazione di un VA è volta a soddisfare il dovere prioritario del genitore, allora molti pedoni verranno investiti e gli abitanti, per far fronte a ciò, si potrebbero unire in un gruppo da “effetto NIMBY” che ha, tra i suoi scopi molteplici, quello di opporsi ad una circolazione di VA progettati considerando il dovere del genitore. NIMBY è l’acronimo di Not In My Back Yard che in italiano si traduce come “Non Nel Mio Cortile” e indica un gruppo di persone che da diversi anni si oppongono alla realizzazione di opere di interesse comune, come depuratori di acque nere, grandi vie di comunicazione, cave, bruciatori di rifiuti solidi urbani, impianti industriali inquinanti, che possono avere effetti negativi su un territorio di interesse comune per gli oppositori, che sarebbero però a favore della costruzione di queste opere su un ambiente non interessante per loro.

In diverse interviste e studi svolti emerge, in maniera evidente, che gli stessi intervistati che non vogliono viaggiare a bordo di un VA progettato per soddisfare i diritti dei passeggeri, sono gli stessi che preferiscono veicoli autonomi realizzati con lo scopo di minimizzare i danni alla collettività. Si può quindi affermare che le preferenze di un potenziale utente del VA oscillano tra la massimizzazione del benessere collettivo e il rispetto dei diritti dei passeggeri, ossia tra quello che è l’obiettivo dell’etica delle conseguenze e quello dell’etica dei doveri. Ciò avviene a seconda del ruolo che la persona ricopre come utente, pedone o passeggero.

3.5) ETICA DEONTOLOGICA (DEI DOVERI) NELLA STORIA

L’etica dei doveri o etica deontologica (da *deon-ontos* che significa “cosa occorre fare” e *logos* ossia scienza) è oggetto di riflessione e approfondimento nel mondo occidentale fin dai tempi di Socrate, ma soprattutto nei tempi dell’Illuminismo, in cui emerge il filosofo Immanuel Kant (Kant, 1788). Nel mondo orientale invece, il filosofo più importante che si è dedicato allo studio dell’etica è il cinese Confucio. In oriente, dove non esiste una netta separazione tra religione e filosofia, un forte impulso viene dato dal Buddismo e dal Taoismo. Tra tutti i deontologi, il più famoso è Kant.

Kant esclude dal campo della morale tutto ciò che è esperienza. Il bisogno di agire secondo regole deriva dalla duplicità della natura umana: l’uomo è sia ragione che sensibilità. Per questo, la legge morale si configura come un dovere, assumendo il carattere di necessità incondizionata e si presenta come comando imperativo. La morale kantiana prevede che ci siano principalmente due tipi di imperativi, quelli ipotetici, che comandano un’azione come mezzo per qualcosa, per raggiungere uno scopo, come nel caso del genitore che ha il fine di accompagnare il/la figlio/a a scuola (già esplicitato nel paragrafo precedente), seguendo regole di prudenza e abilità alla guida, e quelli

categorici, che prescrivono un'azione da compiere senza condizioni, per se stessa e, nel campo dei VA, sono più interessanti perché determinano una serie di principi sulla base dei quali un'azione può essere giudicata buona o meno, senza tenere conto dell'azione stessa ma considerando l'intenzione con cui un'azione viene fatta, come nell'esempio già citato e, in questo caso: "DEVO portare il/la figlio /a a scuola". Nella formulazione della deontologia, Kant rifiuta tutto ciò che è empirico, perché cerca un sistema etico che dipenda da una logica non confutabile, così che la correttezza etica di un comportamento sia un dovere assoluto e innegabile. L'etica non riguarda una categoria ristretta di persone ma l'unanimità in generale.

Capitolo 4: CONTROLLO UMANO E ATTRIBUZIONE DELLA RESPONSABILITÀ

4.1) CONTROLLO UMANO

In caso di collisione inevitabile, il guidatore umano non riesce più a riprendere il controllo diretto del veicolo. A questo punto nasce una questione che riguarda la centralità dell'uomo nelle azioni che un VA dovrebbe svolgere in modo autonomo, in base al suo livello di autonomia (L1-L5). In questo contesto si propone l'idea di Controllo Umano Significativo (CUS), che deve affrontare delle sfide di carattere tecnico e teorico, sia per la creazione di interfacce volte ad agevolare l'interazione tra essere umano e veicolo, sia per le necessità di massimizzare le prestazioni dei VA.

Il CUS assume tre ruoli principali:

- *Meccanismo di salvaguardia* per cercare di impedire che un comportamento imprevisto del veicolo abbia un impatto negativo sulla società;
per svolgere tale funzione, il controllore umano deve avere la possibilità di esprimere un giudizio sulle azioni che il VA sta compiendo o dovrebbe compiere e di avere il tempo per intervenire in caso di azioni impreviste; deve quindi prendere familiarità con i limiti della tecnologia in uso e delle relative cause;
- *Catalizzatore di responsabilità* che consente di individuare il soggetto responsabile in caso di eventi dannosi;
- *Garante del rispetto della dignità degli esseri umani*, nonché del diritto umano di prendere decisioni sulla vita e sull'integrità fisica delle persone coinvolte, ad esempio, in un sinistro.

4.2) IL PROBLEMA DELL'ATTRIBUZIONE DI RESPONSABILITÀ

Nel caso di sinistro stradale in cui sono coinvolti un VA e delle persone, la domanda da porsi è: "Chi è responsabile?".

Il tema dell'attribuzione della responsabilità rappresenta un problema di tipo morale e legale, che nasce al confine tra autonomia operativa e controllo umano. Maggiore è l'autonomia del VA, maggiore è l'entità del problema, essendo sempre minore il controllo diretto che l'uomo esercita sul veicolo nello svolgere determinate azioni.

Analizzando le figure e i ruoli delle persone che possono essere coinvolte, esse sono molte: programmatori, produttori, passeggeri, oltre a tutti i possibili utenti della strada. In questa molteplicità di soggetti diventa difficile rispondere a questa domanda, in quanto nessuno di questi esercita un controllo diretto e completo sul sistema. Idealmente l'unica entità a cui il controllo si vorrebbe delegato è il VA stesso; tuttavia è un paradosso attribuire ad esso la responsabilità, dal

momento che i sistemi tecnologici attuali non hanno responsabilità giuridica. Ciò provoca incertezze di attribuzione della responsabilità e di retribuzione morale, quest'ultima intesa come la possibilità di attribuire con giusta causa la responsabilità di un danno morale a qualcuno.

4.2.1) Veicoli semi autonomi

Se si considerano i veicoli aventi un livello di autonomia fino a L3 compreso, ossia vetture semi autonome, si ritiene che sia sempre possibile attribuire la responsabilità ad un soggetto umano, identificato nel conducente, per danni derivanti da malfunzionamenti. Infatti, in queste auto ci sono dei compiti che devono essere svolti automaticamente e altri manualmente, perciò il conducente deve monitorare costantemente l'azione del veicolo ed avere sempre la giusta attenzione durante la guida, per poter intervenire prontamente in caso di comportamenti inopportuni. Questo soggetto viene chiamato "conducente potenziale", il quale è ritenuto avere le stesse responsabilità di un conducente tradizionale, affinché possa riprendere il controllo del veicolo in qualunque momento. Entrambi i conducenti ricoprono un ruolo di garanzia, sia nei confronti degli utenti della strada, sia rispetto alle persone a bordo del veicolo.

4.2.2) Veicoli L4

Fino a che il livello di automazione del VA è basso, il conducente è chiamato a partecipare in modo attivo e costante all'attività del veicolo, rimanendo responsabile per gli eventuali danni provocati.

Col crescere del livello di automazione, le richieste fatte al "conducente potenziale" calano gradualmente a livello di entità e viene chiamato a effettuare una mera sorveglianza dell'attività del mezzo, senza che siano necessari interventi attivi su lunghi tratti stradali o in archi temporali più lunghi. Supponendo che il passeggero imposti il proprio mezzo per un viaggio lungo, mentre lui si riposa, si dedica alla lettura o a lavorare e che ad un certo punto del viaggio il veicolo provochi un sinistro, diventa difficile parlare di colpa commissiva, ma omissiva. Questo è dovuto al fatto che il conducente non ha svolto azioni di sorveglianza. Differente è il caso in cui il conducente, pur sorvegliando sulle azioni del veicolo, non ritenga opportuno prendere il controllo dello stesso e, eventualmente, intervenire. In tal caso si può parlare di commissione e non di omissione perché volontariamente non ha compiuto l'azione corretta. Dal modello di evento dannoso provocato commissivamente per colpa del conducente del veicolo, si passa a un modello alternativamente commissivo e omissivo, in quanto il conducente non può essere incolpato per aver direttamente violato una regola cautelare, ma per la mancata o mal osservanza dell'obbligo

di sorveglianza sul comportamento del veicolo, che non dovrebbe incorrere in errori o malfunzionamenti. Il conducente rimane pertanto il “capro espiatorio” cui addossare qualsiasi responsabilità.

Seppure nel pieno della quarta rivoluzione industriale, il cui tema principale è la connessione tra sistemi fisici e informatici, che possono collaborare strettamente con altri sistemi cyberfisici (Quarta, E., & Trezza, R., 2021), non è un azzardo affermare che stiamo vivendo la società del rischio 4.0, alquanto perplessa nei confronti dell’abbandono della figura del conducente nella sua figura di “capro espiatorio, parafulmine capace di attirare su di sé le colpe per qualunque evento occorso indipendentemente dalla sussistenza di un reale e ragionevolmente esigibile potere di controllo degli eventi”.

In questo contesto, non sembra ragionevole attendersi che ad un conducente possa venire chiesto di avere un grado di sorveglianza minore dal momento che ciò potrebbe avere delle ricadute in termini di responsabilità. In tale scenario normativo appare irrealistico supporre l’introduzione di limitazioni legislative della punibilità dei conducenti di auto semi autonome, nonostante esse siano prossime alla completa automazione. In tutto ciò la problematica principale è data dal *control dilemma*: nonostante le potenzialità tecnologiche offerte da un sempre crescente livello di autonomia nella guida, il conducente non viene mai liberato da responsabilità, anche in caso di eventualità lesive imprevedibili e non governabili. Ciò fa rilevare un contrasto con il principio di colpevolezza ma prima ancora con una ragione di carattere pratico. Da un lato si assisterebbe alla perdita di tutte le comodità e benefici connessi a una circolazione di veicoli con tecnologia avanzata, dall’altro lato si tratta di una soluzione che provocherebbe rischi peggiori e più numerosi rispetto a quelli che l’obbligo di sorveglianza è effettivamente in grado di prevenire. Sono molti gli studi che hanno evidenziato come la capacità e la rapidità di reazione umane appaiono essere limitate dallo svolgimento della mera sorveglianza, diversamente da quanto avviene quando il conducente copre un ruolo attivo, ovvero con la guida manuale. Si giungerebbe al paradosso per cui conducenti sempre meno abituati a guidare manualmente, vengono chiamati a riprendere il controllo del veicolo in presenza di situazioni più critiche. Infine, va tenuto in considerazione anche il caso dei “falsi pericoli”, in cui il conducente interviene, convinto della presenza di un pericolo che, in realtà, non esiste, provocando un sinistro che altrimenti non si sarebbe verificato.

4.2.3) Veicoli L5

L'attribuzione di responsabilità nel caso di VA di livello L5 presenta problematiche ulteriori. Dal momento che di conducente non si può più parlare, la responsabilità in caso di sinistro potrebbe essere attribuita al passeggero del mezzo (inteso come possessore dello stesso), ma solo se la sua mancata manutenzione ha causato l'evento dannoso. La problematica riguarda gli errori di malfunzionamento che sono dovuti a un'evoluzione anomala dell'algoritmo, errori che non possono essere imputati a un conducente. Chi è il responsabile, quindi?

Sono percorribili due ipotesi: la prima è quella di attribuire la responsabilità al veicolo; la seconda è quella di imputarla al programmatore e al produttore.

Prima proposta: Veicolo soggetto di diritto

Anche in sede europea si è discussa la possibilità di attribuire la responsabilità civilistica a una personalità elettronica. Ciò apre una finestra di riflessione dal punto di vista giuridico ed etico, in quanto porrebbe fine alla quarta rivoluzione (Crosato, 2021). Attraverso le quattro rivoluzioni, infatti, c'è stata una progressiva decentralizzazione dell'uomo e considerare la macchina soggetto di diritto (e imputabile) sembrerebbe chiudere il cerchio. L'idea di uomo al centro del mondo come prescelto da Dio, inizia a venire meno nel 1543, con l'avvento della teoria di Copernico, il quale pubblica un trattato con cui sostiene la rotazione dei pianeti intorno al Sole. Ecco che il ruolo dell'uomo viene ripreso in considerazione e la sua posizione non è più al centro dell'Universo, pur rimanendo al centro del mondo biologico. Tale linea però si dissolve con *L'Origine delle Specie* di Darwin (Popper, 1978), il quale afferma che l'uomo è frutto di un processo evolutivo, tramite una selezione naturale. Tuttavia, la specie umana rimane padrona della propria mente e dei suoi contenuti, attraverso l'affermazione *cogito ergo sum* di Cartesio. Freud smentisce tale teoria, a sostegno del pensiero che la nostra mente non è completamente nostra in quanto tale, ma perché condizionata dal nostro passato. A questo punto della riflessione si può ancora affermare che l'uomo occupa un posto speciale nel cosmo per le sue capacità di pensiero superiori. L'Intelligenza è qualcosa di prezioso, che ogni uomo deve custodire.

Le radici della quarta rivoluzione vanno ricercate nel 1600, quando Hobbes afferma che pensare equivale a ragionare e ragionare è calcolare, tralasciando però alcune oscurità non ancora percepite e chiarite successivamente con Alan Turing (Turing, 1950), che è detto il padre della quarta rivoluzione, con il test di Turing. È proprio con Turing che viene abbandonata l'idea dell'uomo al centro dell'Universo, della sua capacità di agire in modo intelligente. Si tratta di un passaggio sofferto avvenuto nel tempo. Infatti, pur essendo affascinato e curioso dell'autonomia di macchine da esso stesso creato, l'uomo accetta con fatica che alcune funzioni che da sempre

gli sono appartenute, vengano svolte dalla macchina. Attribuire la responsabilità al veicolo potrebbe porre fine a un processo per cui l'uomo è sempre stato al centro del diritto, oltre a creare un effetto di deresponsabilizzazione nei confronti dell'agente umano.

Hallevy è il principale sostenitore dell'attribuzione della responsabilità a una macchina per danni creati da anomalie nel software e individua tre modelli di responsabilità applicabile ai sistemi di IA.

1. *The perpetration through another*: il fatto di reato provocato dall'azione o omissione di un diverso soggetto viene attribuito al soggetto che utilizza l'autore materiale del crimine come mero strumento, ferma restando la responsabilità dell'autore materiale, qualora esso sia perseguibile. Secondo lui la macchina è totalmente priva di capacità cognitive. Il programmatore viene individuato come penalmente responsabile;
2. *The Natural Probable Consequence*, in cui i programmatori sono individuati come penalmente responsabili del reato commesso dal prodotto IA, nel caso in cui esso sia la conseguenza naturale di un loro atto doloso. A tale scenario risponde l'approccio *Etic By Design*. Come dimostrato precedentemente, in fase di addestramento del software, vengono inseriti dataset che non sono abbastanza rappresentativi e diversi *bias* si trovano all'interno degli algoritmi. Questo è un elemento che sposta la responsabilità sul programmatore. Affinché egli sia esente da questa colpa è necessario che intervenga per correggere tali *bias* attraverso l'utilizzo di un approccio etico alla base della progettazione dell'algoritmo, che dovrebbe essere addestrato su dataset più inclusivi.
3. *The Direct Liability*, modello che dimostra che non esiste alcuna dipendenza tra la condotta del programmatore e quella posta in essere dalla macchina. Ritiene che la piena responsabilità sia da attribuire al veicolo.

Diversi esperti in IA si mostrano contrari alla teoria di Hallevy in quanto sostengono che attribuire la responsabilità al veicolo sia una scelta non appropriata dal punto di vista etico e legale. I punti sviluppati dagli esperti sono:

1. Si attribuisce dignità e integrità al prodotto IA, contrariamente a quanto previsto dalla Carta Europea dei Diritti dell'uomo e alle diverse Convenzioni atte a proteggere i diritti umani e le libertà fondamentali;
2. La scelta non può essere desunta dal modello legale di identità perché ciò implicherebbe l'esistenza di una persona umana dietro una legale, che la rappresenti e la diriga;

Seconda proposta: Responsabilità del produttore e al programmatore

Con la Risoluzione del febbraio 2017, relativa ai danni causati dai prodotti IA, il Parlamento Europeo sostiene di estendere le disposizioni già esistenti, con riferimento alla direttiva 85/374/CEE relativa al ravvicinamento delle disposizioni legislative, regolamentari e amministrative degli Stati Membri della Comunità Europea, alla materia delle responsabilità per danno da prodotti difettosi.

Due sono gli approcci al fine di inquadrare la responsabilità dell'*uomo dietro la macchina* per i danni provocati dalla stessa.

Il primo è rappresentato dall'ipotesi di *Strict Liability*, ossia di responsabilità oggettiva per la quale è richiesta solo la prova del danno avvenuto e l'individuazione del nesso di causalità che collega l'azione lesiva posta in essere dalla macchina al danno subito dalla parte lesa (Passagnoli, 2019).

Il secondo consiste in un approccio di gestione dei rischi. L'attenzione va sul soggetto in grado di minimizzare i rischi e affrontare l'impatto negativo.

Così come un incidente è causato dall'intervento di molti agenti, così la responsabilità può essere distribuita su di essi. Il rischio che si corre però è che non sia giusta ed equa, il problema delle molte mani, che genererebbe una sorta di vuoto di responsabilità. Come colmare questo vuoto è una questione che accompagnerà l'integrazione dei sistemi autonomi e intelligenti nelle società umane e nelle città. Segue un'analisi dei regolamenti che diversi stati hanno adottato per la circolazione dei VA.¹⁵

GERMANIA

La Germania è il primo stato europeo favorevole all'integrazione dei VA nelle sue città.

Nel 2012 un rapporto pubblicato da *Highway Research Institute* afferma che sono compatibili con le leggi in vigore tutti i livelli di guida autonoma, tranne quello di completa autonomia, in quanto la loro circolazione sarebbe in contrasto con due disposizioni del codice della strada tedesco, in particolare con l'Art. 1 in cui è richiesta espressamente attenzione da parte del guidatore e con la sezione 18 che stabilisce che il guidatore è tenuto a risarcire il danno scaturente dall'incidente a meno che non si dimostri non imputabile a sua colpa.

In ultimo nel maggio 2017 la Germania ha ufficialmente approvato una legge che consente i test di auto a guida autonoma sulle strade pubbliche.

Tale legge prevede però alcune condizioni per l'esperimento dei test tra cui:

- La presenza nel veicolo di un sistema di registrazione dei parametri di funzionamento dello stesso;

• ¹⁵ https://www.dirittoconsenso.it/2021/05/14/guida-autonoma-tra-progresso-e-profili-di-responsabilita/#_ftn6

- La presenza di un operatore sempre in grado di riassumere la guida manuale del veicolo.

La normativa prevede ulteriormente un regime di responsabilità oggettiva, addossando in caso di incidente la responsabilità al proprietario del veicolo.

Nel 2021 il Bundestag (parlamento federale tedesco) decide che dal 2022 i VA possono circolare su alcune strade pubbliche, soprattutto quelle che collegano gli aeroporti con le grandi città. La nuova legge, approvata con il contributo del senato federale, autorizza la circolazione dei VA con livello di autonomia L4 che può gestire qualsiasi tipo di soluzione, ma con l'ausilio di segnaletica stradale e con la limitazione che la modalità autonoma non deve essere attivata in caso di maltempo o altre condizioni estreme di guida.

ITALIA

Ad oggi la normativa italiana non consente ancora la circolazione di VA di livello superiore a L2. Analizzando il Codice della Strada italiano, che definisce i veicoli come tutte le macchine che circolano su terra guidate da un uomo (Art. 46 co.1 del CdS), emerge che al guidatore umano viene trasferita la responsabilità di supervisionare il sistema e di intervenire nelle situazioni critiche, di pericolo o quando richiesto.

Lo Stato italiano applica la Convenzione di Vienna del 1968 sul traffico stradale, grazie alla quale in Italia si sono svolti diversi test, tra cui quello del 1998 in cui una Lancia Thema ha percorso 2000 Km circa sulla penisola, in modo autonomo per il 94% del tratto totale.

Il 28 febbraio 2018, con il decreto Smart Road, sono state introdotte nuove regole per la sperimentazione dei VA su strade pubbliche. Il testo identifica il VA come veicolo dotato di tecnologie in grado di adottare e attuare comportamenti di guida senza l'intervento attivo del guidatore, in determinati ambiti stradali e condizioni esterne e ha istituito anche un apposito osservatorio tecnico mirato a coordinare le varie iniziative e sperimentazioni, anche a supporto di studi, ricerche e approfondimenti sul tema della sicurezza.

Il decreto consente la circolazione di VA solo se presente un supervisore a bordo, preparato a passare dalla modalità automatica a manuale in qualsiasi momento. Inoltre, ogni 15 giorni il soggetto che effettua la sperimentazione deve produrre e consegnare un rapporto su eventi o problematiche che possono avere impatti sulla sicurezza.

Un'altra novità introdotta dal decreto è l'obbligo, nel caso in cui la sperimentazione venga portata avanti da soggetti diversi dal costruttore del mezzo, di presentare un **nulla osta rilasciato da chi ha costruito il veicolo**.

FRANCIA

In **Francia**, attraverso un provvedimento con data 3 agosto 2016 è stata avviata la regolamentazione della sperimentazione dei veicoli autonomi su strade pubbliche.

Nel febbraio 2017, l'Ispettorato Generale ha pubblicato un documento che programma un'azione coordinata al livello governativo, amministrativo al fine di dare impulso allo sviluppo tecnologico e giuridico così da facilitare l'ingresso dell'auto a guida autonoma nel mercato.¹⁶

La Francia ha adottato un regolamento per i VA, inserendolo nelle modifiche 2022 al Codice della Strada. I cambiamenti nel codice, in vigore dal settembre 2022, sono da applicare su veicoli di trasporto autonomi (o dotati di determinati dispositivi) su strade e zone di città specifiche. L'introduzione sarà molto graduale e le regole non saranno valide su tutta la rete stradale francese. Il regolamento prevede il sollevamento del guidatore dalla responsabilità in caso di incidente con guida autonoma attivata, modalità che va utilizzata in modo appropriato. La responsabilità sarà attribuita all'auto e, di riflesso, al costruttore.

Tra i sistemi considerati non c'è il cruise control adattativo, ma, ad esempio, il mantenimento di corsia attivo, atto a mantenere l'auto all'interno della corsia, correggendo la traiettoria.

La svolta nel campo della guida autonoma è data dall'introduzione di dispositivi che consentono all'auto di raggiungere una destinazione impostata regolando autonomamente acceleratore, freno e svolte agli incroci.

A Rouen, ad esempio, si potranno vedere su strada le prime auto "legittimamente" autonome, mentre in alcuni quartieri di Parigi e a Montpellier verranno testati gli autobus o i furgoni senza conducente.

STATI UNITI

Dal punto di vista giuridico-internazionale, la mancata adesione alla Convenzione di Vienna del 1968 complica il quadro sulla circolazione dei veicoli in territorio americano. Si richiede pertanto ai Governi dei singoli Stati federali di risolvere i principali problemi giuridici legati ai veicoli a guida autonoma, tra cui protezioni dati personali e responsabilità. Secondo un rapporto del 2020, già dal 2012 almeno 41 Stati, assieme a Washington D.C., consideravano la possibilità di introdurre un

• ¹⁶ <https://it.motor1.com/news/518158/francia-guida-autonoma-regolamento-2022/>
• <https://www.quattroruote.it/news/sicurezza/2021/07/05/guida-autonoma-francia-codice-della-strada.html>

regolamento apposito per i VA. Da allora, il numero degli Stati che hanno effettivamente adottato delle norme per le *self-driving cars* è in costante aumento: allo stato attuale 29 Stati americani, in aggiunta al Distretto di Washington, hanno promulgato le proprie leggi per i VA. Tra i restanti 11, lo Stato della California è stato tra i primi a dotarsi di una legge specifica per i VA: il Governo californiano inizia ad apportare degli emendamenti al proprio Codice della Strada già dal 2012, con la Legge n. 1298, per modificare la sezione 38750 del *California Vehicle Code*, che descrive i veicoli a guida autonoma come “*ogni veicolo dotato di una ‘tecnologia automatizzata che è stata integrata nel veicolo stesso’*”. Vengono esclusi dalla categoria in esame tutti quei veicoli che, seppur dotati di sistemi elettronici di supporto alla guida, come ad esempio cruise control e assistenza al parcheggio, non sono in grado di condurre l’autovettura senza il controllo attivo o il monitoraggio di un supervisore. A tal proposito, occorre sottolineare come la Legge non fa riferimento alla figura del “conducente”, quanto piuttosto a quella dell’operatore, indentificandolo in colui che siede al posto del guidatore ovvero, in mancanza, in colui che provoca l’attivazione della modalità guida autonoma. Riguardo ai profili di responsabilità, la Legge si limita a prevedere che “*il costruttore della tecnologia autonoma installata su un veicolo deve fornire una comunicazione scritta all’acquirente che descriva quali informazioni sono raccolte dalla tecnologia autonoma di cui l’autovettura è provvista*”. Tra i requisiti obbligatori di un VA che circola sulle strade californiane è infatti contemplato un meccanismo che cattura e conserva i dati che i sensori ricevono nei trenta secondi antecedenti una eventuale collisione. I dati devono essere archiviati all’interno di tale sorta di scatola nera per un periodo massimo di tre anni dalla data in cui si è verificato l’incidente e sono accessibili esclusivamente mediante un meccanismo esterno designato appositamente. Dal quadro finora tracciato sembrerebbe dunque poter ammettere che nelle ipotesi in cui il sistema di *storing* dei dati dimostri che la causa della collisione non sia riconducibile ad un difetto dell’Intelligenza artificiale della vettura, il produttore sia esonerato dalla responsabilità per i danni eventualmente prodotti a terzi. La legge tutela quest’ultimo prevedendo tutta una serie di certificazioni, le quali non solo provano la dotazione di tutti gli elementi meccanici del veicolo, ma anche il rispetto degli standard di sicurezza e prestazione dettati a livello nazionale e federale, nonché l’esistenza di un vincolo assicurativo, il cui importo è innalzato nei casi di veicoli a guida autonoma rispetto ai massimali previsti dalla legge per i veicoli per così dire comuni. Qualora dunque il produttore sia regolarmente in possesso di tutte le certificazioni del caso, senza le quali tra l’altro non potrebbe avviare né la sperimentazione né l’utilizzo su strada del suo veicolo a guida autonoma, egli risulterebbe esonerato da qualsiasi tipo di responsabilità.

Capitolo 5: DESIGN KNOWLEDGE GAP E INCERTEZZA EPISTEMICA

5.1) DESIGN KNOWLEDGE GAP

Nel discutere dei problemi etici che riguardano i sistemi di IA, il dibattito riguarda principalmente cosa un sistema dovrebbe o non dovrebbe fare (Héder, 2020). I possibili esiti sono definiti e pochi, spesso due (decisione binaria).

Un buon esempio è il *MIT Moral Machine*¹⁷, una piattaforma online con lo scopo di raccogliere il punto di vista umano in merito a decisioni morali, che vengono prese da un sistema IA, come il veicolo autonomo.

In questa piattaforma vengono mostrati alcuni “dilemmi” morali, tra cui quello in cui il VA deve scegliere se investire cinque pedoni o mettere a rischio la vita dei due passeggeri. Il partecipante al sondaggio dovrà porsi come osservatore esterno e giudicare quale sia, secondo lui, l’esito più accettabile, rivelando così preferenze e pregiudizi di conservazione tra giovani e anziani vittime di incidenti stradali, maschi o femmine, eccetera. Sono state rilevate 40 milioni di decisioni in dieci lingue diverse, da milioni di persone provenienti da 233 Stati e territori.

(Awad et al., 2018) analizza i risultati dell’esperimento secondo i seguenti punti:

- 1) Riassunto delle preferenze morali globali;
- 2) Documentazione delle variazioni individuali, in base ai dati demografici degli intervistati;
- 3) Studio della variazione etica interculturale e suddivisione dei Paesi in tre gruppi;
- 4) L’analisi delle differenze che porta a concludere che esiste una correlazione tra queste e le istituzioni moderne e tratti culturali.

La percezione dell’esperimento è evidente: (Saxena et al., 2019) afferma che è un input rilevante su ciò che la gente, che può essere distinta anche per cultura, in generale giudica equo. Anche altri, come (Kaplan & Haenlein, 2019) sostengono che tali risultati sono dei veri indicatori. Oltre a MIT Moral Machine tra le rappresentazioni dei “dilemmi” morali, si trova anche (Goodall, 2014) e sono utili per accertare le preferenze morali delle persone e quindi in grado di orientare le decisioni.

Basato su oltre 400 mila soggetti, *MIT Moral Machine* ha fatto luce sulle differenze culturali tra Paesi e regioni nel loro approccio morale e nella valutazione del valore.

¹⁷ <https://www.moralmachine.net/>

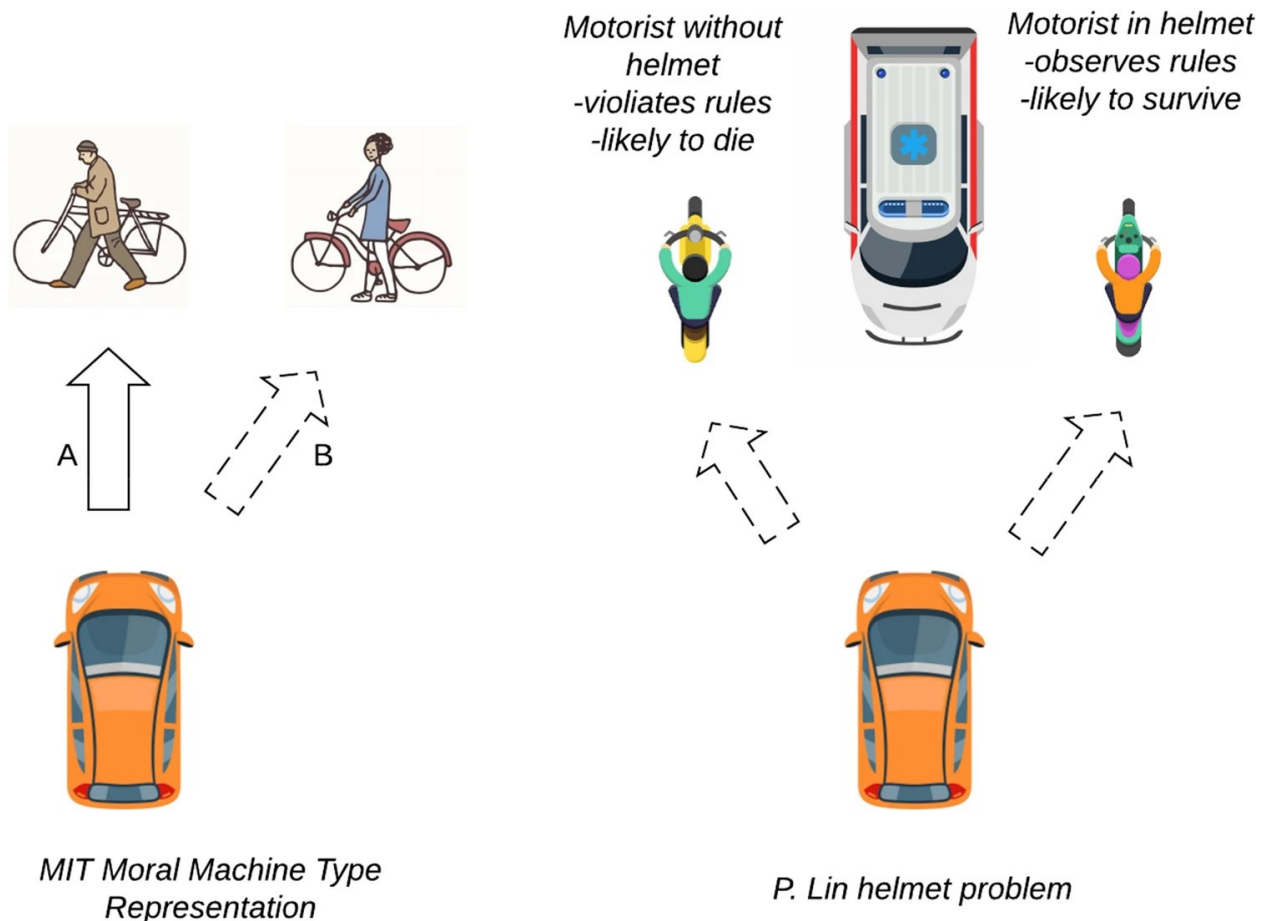


Fig.12 Rappresentazione del MIT Moral machine, anche nel “dilemma” morale in cui il VA deve decidere se investire la persona con o senza casco.

(Héder, 2020) afferma che queste rappresentazioni, pur portando a dei risultati, la cui analisi può far concludere che esiste un legame tra le differenze tra persone di gruppi di Paesi diversi e la cultura e le istituzioni, non aiutano direttamente il programmatore o, in generale, il gruppo di ingegneri che formano il “design team”, che si occupano della progettazione del veicolo, perché i risultati dell’esperimento sono accertati ma non direttamente implementabili. Il problema sta nel fatto che il programmatore non può anticipare con certezza ciò che il veicolo farà in una data situazione, anche perché il numero di alternative può essere molto alto e non è noto in alcun momento. Il “design team” si trova in una situazione di incertezza epistemica. Il *design-time knowledge gap* è il problema per cui, sebbene i progettisti in fase di design abbiano un elevato grado di potere nei confronti di ciò che creano, non sono mai abbastanza in controllo e ciò consentirebbe loro di codificare semplicemente un comportamento utilitaristico, deontologico o qualsiasi altro tipo di comportamento morale. Questo per dire che i dilemmi morali non esistono in fase di progettazione di un algoritmo, ma quando l’algoritmo viene testato, quando, nel caso specifico dei VA, essi vengono messi in strada (dove consentito) ed esso si trova di fronte ad una situazione in cui deve scegliere cosa fare. In fase di design

time, molti scenari e condizioni del traffico e della strada vengono previsti, ma certamente questi fanno parte di un sottoinsieme di tutte quelle situazioni che si possono verificare nella realtà.

5.2) OPACITA' EPISTEMICA

Due sono le possibili motivazioni che stanno alla base del gap in questione. La prima riguarda il sistema, la seconda l'ambiente che lo circonda.

5.2.1) I problemi nel sistema

Tra questi vi è *l'embodiment effect*, di cui si propone una descrizione nel caso dei VA.

Si supponga che in un VA di colore giallo vivace, la tecnologia venga testata e realizzata e che includa la macchina, i componenti hardware e software a essa connessi. Un altro VA viene prodotto con le stesse caratteristiche del primo, ad eccezione del colore, che questa volta è grigio pallido. Il veicolo grigio è coinvolto in un numero minore di incidenti dell'altro, a parità di condizioni. Entrambi, infatti, sono stati posti in circolazione in un traffico ricco di pedoni. Lo studio afferma che questa differenza è dovuta al colore diverso dei due veicoli. Il grigio pallido è infatti molto meno visibile del giallo vivace. Questo fatto dovrebbe far sì che i programmatori intervengano a livello software, i colori diversi dovrebbero essere testati e compresi in un range definito e l'utente dovrebbe essere avvisato che la garanzia decade se applicato un colore fuori range. Tuttavia, ad oggi non esiste una regola per il colore del veicolo per la presenza di un soggetto umano imputabile. Infatti, anche se l'auto grigia investe più pedoni, alle spalle c'è comunque una persona, che sia essa il conducente o il proprietario a cui può essere addossata la responsabilità del danno provocato.

5.2.2) Il problema dell'ambiente circostante al sistema - *statistical only knowledge*

L'imprevedibilità del sistema autonomo, anche per la conoscenza imperfetta dell'ambiente circostante, determina molte incertezze.

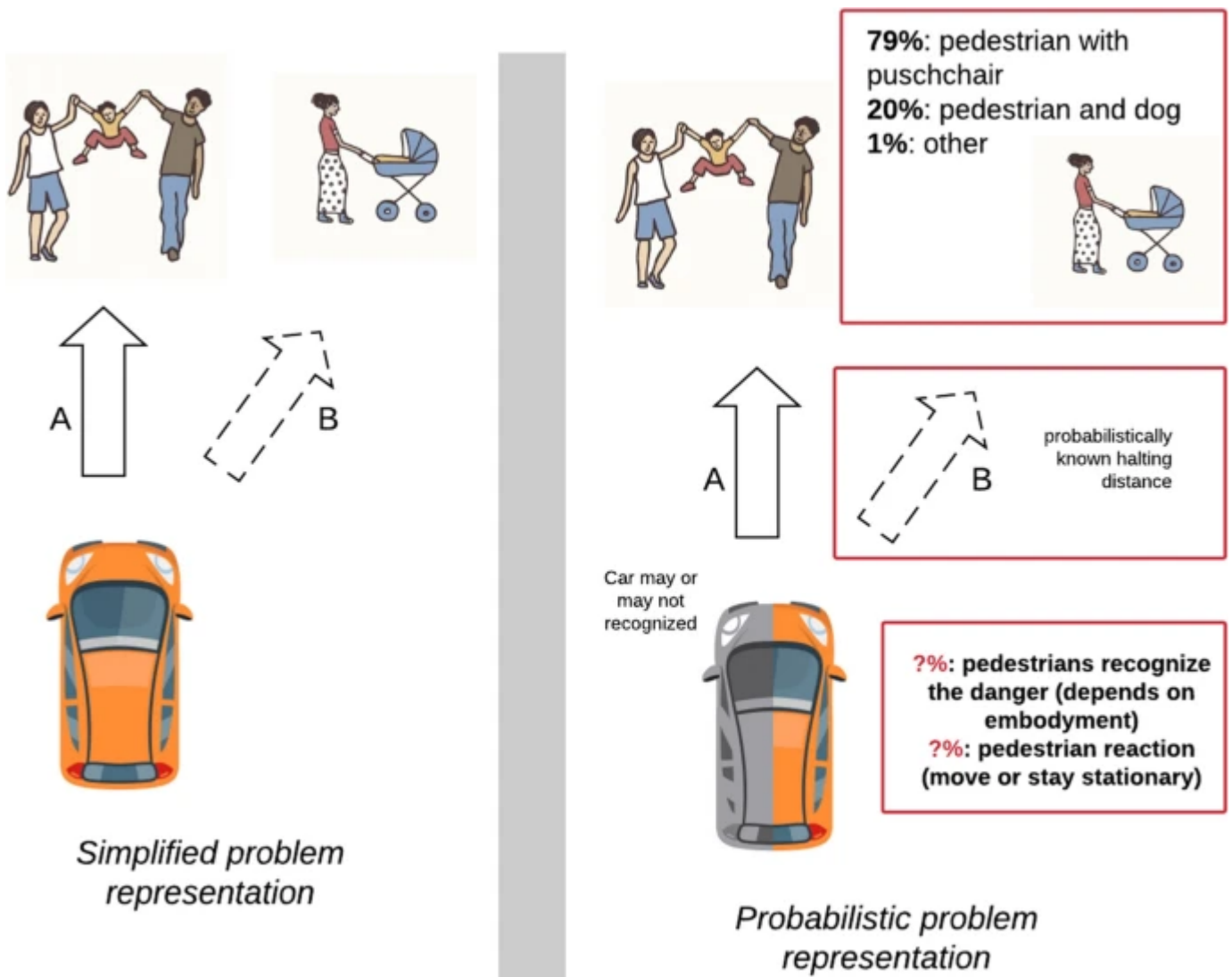


Fig.13 Poiché la situazione stessa non è nota con perfetta precisione, anche l'input della presunta decisione morale è incerto.

Un buon esempio è il software Auto-Pilot di Tesla. È noto che viene addestrato sulla base di filmati di guida reali: anche in modalità Non Auto-Pilot, il software presta attenzione a ciò che fa il conducente e calcola cosa farebbe lui stesso nella stessa situazione e rileva, apprende dalle differenze. Ciò significa che i pregiudizi sono incorporati su almeno due livelli: il primo, la guida umana che rappresenterà le abitudini di guida del gruppo che può permettersi l'auto; il secondo, le strade e l'ambiente di formazione saranno dei Paesi che avranno una maggiore adozione di Tesla. Quindi, nel complesso, si può dire che la rappresentazione automatica di situazioni di importanza morale non è solo probabilistica, ma le probabilità stesse si basano solo su dati di addestramento nel migliore dei casi, approssimando l'ambiente reale e completamente diversi nel peggiore dei casi.

5.2.3) Decisioni umane incalcolabili

Un'altra spiegazione dell'opacità epistemica va ricercata nella reazione degli umani e di altri VA presenti nella scena. Ad esempio, un VA si trova di fronte a un pedone, il quale ha ancora tempo per saltare a destra o a sinistra o rimanere fermo. Anche il VA può decidere se sterzare a destra o a sinistra. Il pedone non è al sicuro perché potrebbe saltare proprio nella direzione verso cui il VA sterza.

Un altro tipo di decisione umana non calcolabile può essere data da atti malvolenti nei confronti del sistema, in questo caso VA, e l'ambiente circostante. (Carfagna B., 2018) afferma che una ragazza di Singapore è stata arrestata per aver "alterato i segnali stradali" applicandoci degli adesivi con delle frasi, con scopo creativo ma consapevole di commettere un'azione non legale.

La figura sottostante rappresenta il gap epistemico e morale tra la fase di progettazione e la fase di esecuzione.

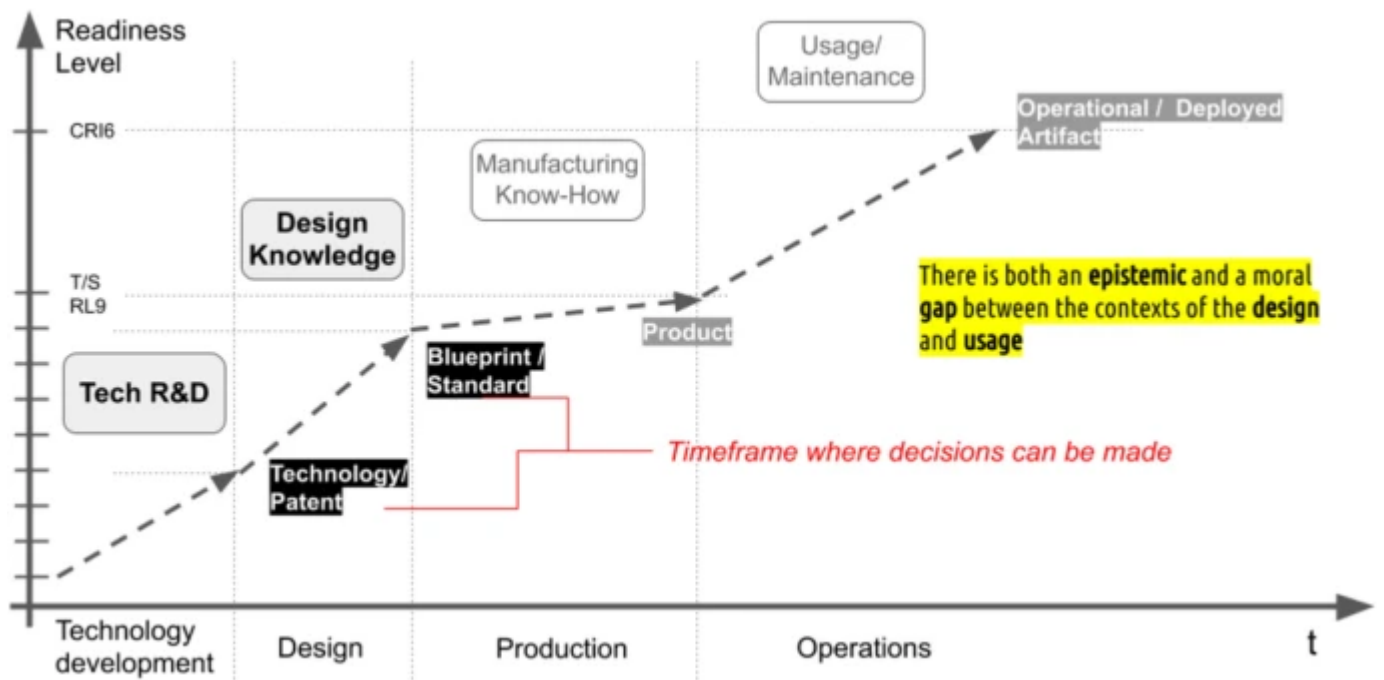


Fig.14 Design-Time VS Run-Time in AI

Le ragioni alla base del gap sono dupplici: da una parte, l'ambiente futuro circostante non può essere predetto, dall'altra le azioni del sistema stesso non possono essere note anticipatamente, perciò considerate per la loro opacità.

5.3) STIMA DELL'INCERTEZZA

La stima dell'incertezza è un tema molto importante nell'ambito del Deep Learning, in particolare nella distribuzione di reti neurali in applicazione *safety-critical* perché rappresenta quanto la predizione sui nuovi dati sia affidabile.

5.3.1) Stima dell'incertezza con sampling-free

(Postels et al., 2019) propone un metodo *sampling-free* al fine di fornire una stima dell'incertezza epistemica di una rete neurale. Il metodo, che si chiama OUR, lascia l'addestramento invariato e usa la propagazione dell'errore, il cui funzionamento è descritto nella figura seguente e in cui l'errore coincide con la varianza.

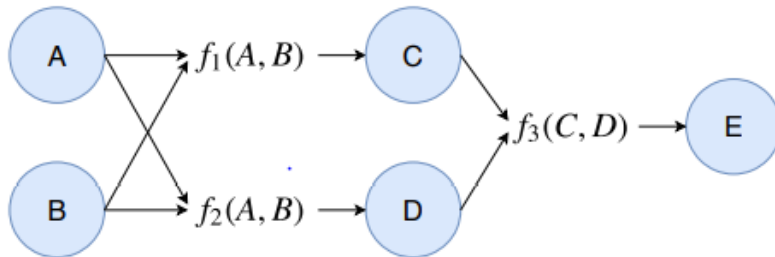


Fig.15 Propagazione dell'errore

L'errore integrato nella rete neurale viene trattato come errore sui valori di attivazione. Addestrandosi con l'integrazione dell'errore, la rete cerca di minimizzare gli errori accumulati sulla distribuzione dei dati di addestramento, in quanto grandi errori corrispondono a grandi segnali di perdita. Lo studio sopra citato ha sperimentalmente provato che il metodo OUR può produrre stime di incertezza veloci e accurate in caso di classificazione e regressione, attraverso l'applicazione dello stesso a un dataset sintetico. Per svolgere l'esperimento è stato creato un dataset di regressione, consistente in un singolo input e un singolo output. L'input è uniformemente distribuito nell'intervallo $[0,20]$. La "target function" è il seno dell'input più il rumore gaussiano di media nulla e sigma 0,3. La rete neurale usata è formata da tre livelli intermedi, ognuno con cento neuroni. Prima il Test Set è stato popolato con valori che non appartengono all'intervallo di input, poi affiancati da campioni che derivano dalla distribuzione del Training Set. Lo scopo è quello di valutare la varianza e, così, la deviazione standard. I risultati dell'esperimento sono riportati graficamente in figura.

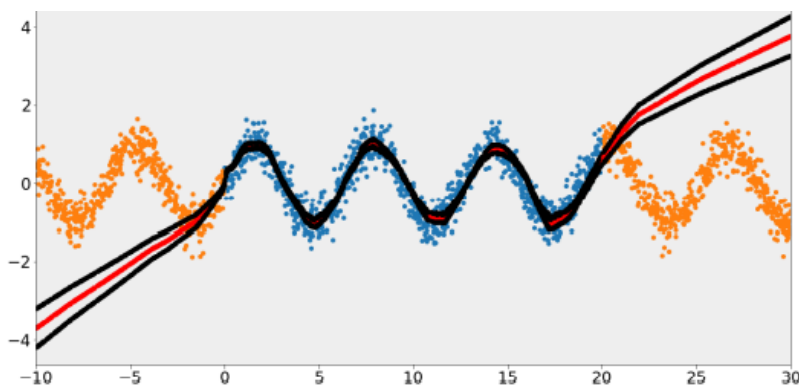


Fig.16 Risultati dell'esperimento, in cui la deviazione standard viene approssimata secondo OUR

- In blu sono stati plottati i campioni appartenenti alla distribuzione dei dati di training;
- In arancione quelli che non appartengono al training set;
- In rosso la predizione;
- In nero la predizione più/meno la deviazione standard.

La deviazione standard della predizione fuori dai dati appartenenti al Training Set aumenta sensibilmente per l'incertezza epistemica.

5.3.2) Stima attraverso un approccio bayesiano

(Harakeh et al., 2019) propone un approccio, chiamato *BayesOD* per la stima dell'incertezza nell'output degli algoritmi di rilevazione dell'oggetto. Questo approccio è di tipo bayesiano ed è stato applicato a quattro tipi di dataset differenti. Si dimostra che *BayesOD* offre delle stime di incertezza meglio correlate con l'accuratezza degli algoritmi di rilevazione, con una riduzione significativa dell'errore di incertezza minimo (MUE), valutato rispetto al prossimo metodo migliore e su tutti i diversi dataset. Ciò è emerso tramite il confronto di *BayesOD* con i metodi già esistenti quali *Black Box*, *Sampling Free*, *Anchor Redundancy* e *Joint Aleatoric Epistemic*.

La metrica utilizzata è MUE: *Minimum Uncertain Error*, usato per determinare la capacità di un rilevatore di distinguere tra *True Positive*, cioè entità correttamente classificate come appartenenti a quella classe e *False Positive*, ossia classificate come appartenenti ad una classe a cui, in realtà, non appartengono.

5.4) DUE TIPI DI INCERTEZZA

È opportuno distinguere due tipi di incertezza, quella aleatoria e quella epistemica.

Quella aleatoria si riferisce alla nozione di casualità, che è la variabilità degli esiti di un esperimento, dovuta alla presenza di effetti random in esso; quella epistemica è dovuta alla mancanza di conoscenza. In altre parole, si riferisce all' "ignoranza" del soggetto che deve prendere una decisione. A differenza di quella aleatoria, l'incertezza epistemica può essere corretta con l'aggiunta di informazioni. Nel Machine Learning e nel Deep Learning questi tipi di incertezze sono spesso confuse. Per questo, in precedenza si è parlato di incertezza, in via generale.

5.5) ALEATORICA VS EPISTEMICA

Per una guida autonoma sicura si ritiene importante "catturare" entrambe le incertezze. A tal fine, (Feng et al., 2018) propone dei metodi pratici che hanno il compito di catturare le incertezze di un rilevatore di veicoli 3D per nuvole di punti Lidar. I contributi dati dallo studio sono tre:

- *Estrazione dell'incertezza del modello e dell'osservazione*; sono state misurate rispettivamente l'incertezza epistemica e aleatoria. Per quella epistemica sono state usate funzioni Softmax più o meno complesse e l'entropia di Shannon che cattura l'incertezza nell'output di predizione e che raggiunge il suo picco quando la probabilità che un campione oggetto sia classificato come veicolo è pari a 0,5. Si precisa che quest'ultima viene calcolata come probabilità condizionata e attraverso una funzione Softmax. Inoltre, si cerca di dare una stima dell'incertezza spaziale, attraverso un'approssimazione della posizione del bounding box, le cui coordinate vengono trasformate in una cornice di coordinate LIDAR. Questa stima fa uso principalmente della varianza totale della matrice di covarianza calcolata considerando i parametri sopra citati; per quella aleatoria viene usata principalmente una funzione softmax piuttosto semplice e i rumori di osservazione sono stati ottenuti aggiungendo un livello di regressione in output;
- *Analisi della differenza tra incertezza epistemica e aleatoria*, quest'ultima influenzata dalla distanza del veicolo. Esse si comportano in maniera differente l'una dall'altra. I risultati dell'esperimento condotto in (Feng et al., 2018) mostrano che l'incertezza epistemica è correlata con l'accuratezza del riconoscimento. La rete mostra un'elevata incertezza epistemica con campioni che sono molto diversi dal Training Set, come per esempio oggetti fantasmi, grandi veicoli, o veicoli con regressioni del *bounding box* anormali. Ad esempio, quando un rilevatore di veicoli addestrato sull'autostrada, viene impiegato in area urbana, è necessario che esso venga riadattato al nuovo ambiente. Applicando l'incertezza epistemica il rilevatore può essere migliorato coinvolgendo in modo attivo pedoni e ciclisti che sulle autostrade non esistono.
L'incertezza aleatoria è influenzata invece dalla distanza di rilevazione e dall'occlusione, in quanto veicoli distanti o parti "occluse" del veicolo possono contenere alti rumori di osservazione; l'incertezza aleatoria mostra le limitazioni di un sensore e può essere applicata al fine di migliorare il tracciamento della posizione di un veicolo.
- Lo studio dimostra che modellando l'incertezza aleatoria è possibile avere un miglioramento delle performance del rilevatore dell'1%-5%.

Capitolo 6: OPINIONI DELLE PERSONE E FATTORI DI ADOZIONE

6.1) INTRODUZIONE

Alla fine di questo “*viaggio tra algoritmi ed etica*” dedico un capitolo all’analisi delle opinioni delle persone sulla guida autonoma, ritenuta da alcune aziende automobilistiche il futuro della mobilità. Lo sviluppo di diversi sondaggi ha consentito di raccogliere informazioni significative sull’utilizzo, l’acquisto, i vantaggi e svantaggi dell’uso del veicolo autonomo ed individuare i fattori che incidono su tali preferenze.

6.2) LE INDAGINI SVOLTE

Il capitolo si focalizza dapprima sull’approfondimento delle principali indagini che hanno lo scopo di capire i “sentimenti” delle persone verso la guida autonoma. Alcuni esperti del settore affermano che questo tipo di tecnologia avanzata dominerà il futuro del settore della mobilità. Tra le persone comuni, invece, esse possono essere suddivise in due gruppi, quello dei favorevoli e quello dei contrari.

6.2.1) *Global Automotive Consumer Study 2020*

Il primo studio proposto, il *Global Automotive Consumer Study* è stato condotto da *Deloitte* nel 2020. Ha coinvolto 35.000 persone di 20 Paesi diversi in un sondaggio sulla guida autonoma che ha confermato una maggiore tendenza positiva nell’accettazione della tecnologia da parte degli utenti.

Dai risultati emerge che gli italiani sono tra i più favorevoli e meno perplessi rispetto all’arrivo dei VA. Dal 2018 al 2020 si assiste a un calo delle persone preoccupate per la sicurezza. Si passa dal 30% nel 2018 al 25% nel 2020. Inoltre, si assiste mediamente a un calo dei diffidenti rispetto agli anni precedenti, a prova di un’accettazione collettiva nei confronti della tecnologia. I più scettici nel 2018 sono i giapponesi, tra i quali solo il 43% dichiara di essere favorevole all’eventuale adozione del VA. Rispetto all’anno precedente 2 persone su 10 hanno cambiato idea, optando per essere favorevoli. Tale trend viene confermato negli anni successivi.

Nel 2020, a presentare le maggiori perplessità sul tema della sicurezza è l’India, in cui il 58% dei partecipanti al sondaggio dichiara di essere preoccupato circa la sicurezza.

È in Cina dove si registra il più alto numero di consensi nel 2018 e dove è possibile assistere a un netto calo dei diffidenti rispetto all’anno precedente. Il tasso dei contrari passa, infatti, dal 62% del 2017 al 26% del 2018. I consumatori cinesi giustificano questo fatto ritenendo che l’utilizzo del VA potrebbe contribuire alla riduzione del numero di incidenti, attualmente troppo alto. Tuttavia, il trend non viene confermato nel 2020, con un tasso del 35%.

Nello stesso anno è in Italia la percentuale più alta di favorevoli. I risultati sono visibili graficamente nella figura sottostante:

Percentuale di consumatori secondo i quali i veicoli a guida autonoma non saranno sicuri

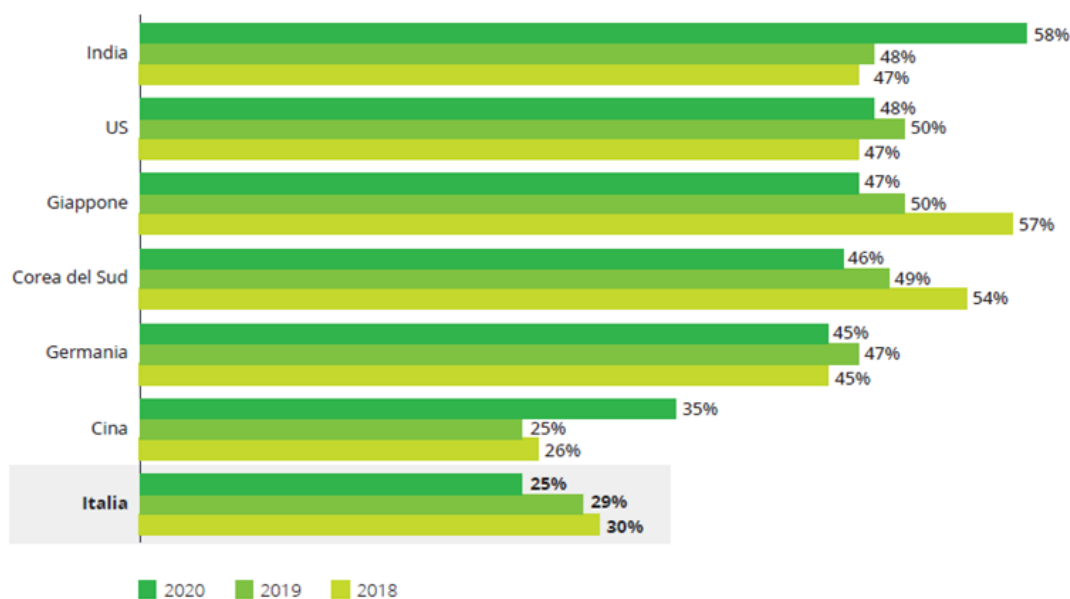


Fig.17 Rappresentazione grafica dei risultati del *Global Automotive Consumer Study, Deloitte 2020*¹⁸

6.2.2) *Cap Gemini 2017*

Il secondo studio, denominato *Cap Gemini*, è stato condotto nel 2017 su circa otto mila mercati chiave e dimostrano che i vari optional della guida autonoma costituiscono dei *selling point* per i consumatori. L’81% degli interessati si mostra disposto a pagare di più per avere in dotazione la funzionalità di guida autonoma. Si registra una maggiore credibilità e affidabilità nei confronti di alcuni assi della tecnologia, quali Google e Apple. In ciò la sicurezza informatica ricopre un ruolo chiave sulle future decisioni di acquisto, dal momento che gli utenti richiedono maggiore controllo sui dati trasmessi e sul loro utilizzo da parte delle aziende.

6.2.3) *Global Automotive Consumer Study 2022*

La figura sottostante riferita all’anno 2022 riporta, per ogni categoria di funzioni di tecnologia avanzata e per ogni Stato considerato, la percentuale di abitanti non disposti a pagare più di 500 dollari americani. L’analisi è stata effettuata attraverso il *Global Automotive Consumer Study* da parte di *Deloitte*. Osservando, nel caso dell’autonomia, la quarta riga della tabella, si può notare che

¹⁸ <https://www.missionline.it/mission-fleet-post/guida-autonoma-e-sicurezza/>

globalmente circa la metà delle persone è disposta a pagare più di 500 dollari americani per avere un veicolo dotato della funzione di guida autonoma.

Advanced technology category	US	Germany	Japan	Rep. of Korea	China	India	Southeast Asia [†]
Safety	56%	70%	66%	58%	31%	48%	59%
Connectivity	65%	77%	83%	72%	39%	48%	65%
Infotainment	69%	82%	86%	78%	39%	57%	72%
Autonomy	61%	69%	56%	42%	31%	37%	48%
Alternative engine solutions	53%	56%	57%	41%	31%	35%	46%
Unwilling to pay more than...	\$500	€400	¥50,000	₩500,000	¥2,500	₹25,000	Local currencies [‡]

Fig.18 Percentuale di consumatori non disposti a pagare più di 500 dollari americani per le diverse funzioni di tecnologia avanzata secondo *Global Automotive Study (2022)* ¹⁹

6.3) I FATTORI

(Meyer-Waarden & Cloarec, 2022) propone il modello UTAUT (Unified Theory Of Acceptance and Use of Technology) per individuare i fattori che incidono sull'adozione e l'accettazione del VA.

Dapprima viene impiegato UTAUT per comprendere l'aspettativa dei potenziali utenti sulle performance della nuova tecnologia e sulla sua facilità di utilizzo, poi lo studio procede considerando anche variabili come l'edonismo e il riconoscimento sociale. Il terzo contributo dato dallo studio riguarda il benessere degli utenti e il quarto si occupa di approfondire i fattori cognitivi, legati alla sicurezza, alla privacy e il loro impatto nei confronti dell'utilizzo della tecnologia dei VA.

I modelli più usati per lo studio dell'accettazione della tecnologia sono TAM e UTAUT. Tuttavia, lo studio effettua le analisi utilizzando UTAUT e non TAM, perché UTAUT è più recente e più sviluppato, dal momento che è frutto del consolidamento di più teorie.

¹⁹ <https://www2.deloitte.com/global/en/pages/consumer-business/articles/global-automotive-consumer-study.html>

6.3.1) *Time benefit*

Nel dettaglio, la ricerca svolta suggerisce che più le persone percepiscono il beneficio derivante dall'uso del veicolo, per esempio trovando vantaggioso riposare durante la guida, guardare un film, o lavorare, sfruttando al massimo il tempo a disposizione (*time benefit*), più sono interessate ad adottare queste funzionalità offerte dalla tecnologia.

6.3.2) *Environmental benefit*

Un altro tipo di beneficio preso in considerazione è quello dell'ambiente (*environmental benefit*) e consiste nella riduzione del consumo di combustibile e dei tempi di viaggio, che spingono il consumatore a esplorare queste funzionalità, rappresentando un approccio di miglioramento dell'efficienza della guida.

6.3.3) WB

Il *wellbeing* (WB) è definito come il livello soggettivo con cui i consumatori percepiscono le esperienze di guida in modo più o meno positivo, attraverso crescenti giudizi su basi cognitive sensoriali e reazioni affettive, senza giudizi oggettivi.

Può essere correlato per esempio con il buon umore, le emozioni positive, la condizione fisica e mentale. Tutte queste condizioni si riferiscono a una buona qualità della vita e a soddisfazioni personali.

La ricerca dimostra che il WB è correlato con le scelte e l'utilizzo della tecnologia da parte dei consumatori. Lavorando sull'incremento della sicurezza, sul miglioramento della qualità dell'aria nel veicolo, sulla riduzione dell'impatto ambientale negativo, si può incrementare il WB del consumatore, nonché incidere positivamente sulla sua salute psicologica e fisica.

6.3.4) *Social recognition*

Social recognition ha un impatto sull'adozione del veicolo autonomo. Essa è definita come il livello con cui l'uso di un prodotto o di una nuova tecnologia tende a rinforzare lo stato sociale di una persona all'interno di un gruppo. L'adozione di un VA come acquisizione di un ruolo nella società deriva dalla necessità di appartenere ad un gruppo e di essere identificati all'interno del gruppo stesso, invitando il consumatore ad essere sempre aggiornato e ad adottare le ultime tecnologie.

6.3.5) Edonismo

Esso è definito come atteggiamento in cui il conseguimento del piacere nell'uso di una nuova tecnologia, come VA, è ritenuto essenziale. Alcune ricerche precedenti hanno dimostrato che la

sensazione di piacere e divertimento può costituire una motivazione per l'utilizzo della tecnologia. Di conseguenza guidare un VA dovrebbe far vivere al consumatore un'esperienza divertente e piacevole.

Si ipotizza che l'edonismo abbia un effetto positivo sulle aspettative da parte dei consumatori in termini di performance e sul benessere (WB).

6.3.6) Sicurezza della tecnologia

Per quanto riguarda questo tema lo studio individua cinque categorie di rischio come fattori che possono influenzare l'acquisto del VA considerato che le conseguenze non sono note a priori con certezza.

I tipi di rischio individuati sono:

- 1) Rischio funzionale: il VA non si comporta come atteso dall'utente;
- 2) Rischio fisico: il VA rappresenta una minaccia per l'incolumità dell'utente;
- 3) Rischio economico: il VA non vale il prezzo pagato;
- 4) Rischio psicologico: il VA ha un impatto su incertezze e salute mentale dell'utente;
- 5) Rischio di privacy: la divulgazione dei dati dal VA minaccia la vita privata e il benessere dell'utente.

Dal punto di vista delle persone il rischio percepito deriva dal fatto che la loro sicurezza è affidata al VA, in quanto nuova tecnologia promettente ma in via di sviluppo, con la paura che accadano imprevisti o si verifichino malfunzionamenti del sistema.

La possibile perdita del controllo del VA e il rischio di hackeraggio sono degli elementi che preoccupano i potenziali consumatori che possono avere dei dubbi e delle paure su questa tecnologia. Le maggiori preoccupazioni vanno ricercate quindi nella mancanza di fiducia nei confronti dei VA. Molti guidatori non sono disposti a delegare compiti al mezzo di trasporto e perciò ad adottarlo come mezzo abituale.

Lavorare sulla sicurezza nei VA è fondamentale per due aspetti: il primo è legato alla sua interazione con l'uomo, nel senso che una guida autonoma più sicura contribuisce alla salute sia fisica che mentale del consumatore, trattandosi di un veicolo pensato per reagire più velocemente di un umano, non essendo distratto e stanco e auspicabilmente non influenzato da scarsa visibilità, dovuta a estreme condizioni meteo o altri fattori che possono ostacolare la visione e il riconoscimento sensoriale dell'ambiente circostante da parte dell'uomo; il secondo è legato alla protezione dei dati degli utenti: è necessario lavorare per ridurre in maniera significativa gli attacchi alle transazioni di dati e reti o accessi non autorizzati ad account di altri utenti.

È atteso che la sicurezza abbia un impatto sulle attitudini dell'utente e sul controllo: se gli utenti ritengono la guida autonoma un mezzo per ridurre gli errori umani in situazioni impreviste, ciò avrà un impatto positivo sulla fiducia nei confronti della tecnologia. Diversamente, tale impatto può diventare negativo nel momento in cui i consumatori percepiscono la tecnologia VA come un rischio, una minaccia, un pericolo per chi sta a bordo e per gli altri utenti della strada (Klobas et al., 2019).

Si ipotizza quindi che la sicurezza percepita possa avere complessivamente un impatto positivo sull'accettazione del VA.

6.3.7) Privacy percepita

Alcuni studi, tra cui (Gurumurthy & Kockelman, 2020), hanno evidenziato l'importanza della privacy nei VA circa la loro accettazione, in quanto i dati che vengono trasmessi sono molti e tra questi una grande quantità contiene informazioni personali che vengono continuamente memorizzate e trasmesse. È importante che nell'operare con un VA siano implementati meccanismi di protezione dei dati, che farebbe sentire più tutelato un potenziale consumatore perché di esso vengono trasferite informazioni circa il suo stato di salute, i suoi comportamenti abituali, le attività quotidiane. Di fronte ad una minaccia di questo tipo, il consumatore può avere delle ricadute negative dal punto di vista del benessere mentale.

Si ipotizza che la questione della privacy possa influenzare negativamente l'adozione del VA.

6.3.8) Affidabilità

Lo studio dimostra che la fiducia nella tecnologia VA rappresenta un fattore tridimensionale nell'accettazione del veicolo.

La prima dimensione riguarda la trasparenza del sistema, ovvero come un VA opera, la seconda è relativa alle competenze tecniche, la terza alla gestione delle situazioni, quindi alla possibilità di riassumere il controllo del veicolo in ogni momento.

La ricerca ha individuato generalmente due tipologie di fiducia nella tecnologia: human-like e system-like. La prima è correlata con la competenza, l'integrità mentre la seconda con la funzionalità e l'affidabilità (Liu et al., 2019).

Si ipotizza che più gli utenti si fidano della tecnologia, più positivo è l'impatto sul benessere della persona.

CONCLUSIONI

Ho compiuto un percorso che mi ha portato a toccare argomenti e discipline diverse e ciascuna mi ha offerto degli spunti di riflessione. Via via ho potuto constatare che vi era la necessità di raggiungere un livello di interrelazione fra le diverse materie per cercare di ottenere un risultato il più possibile condiviso e multifunzionale, apprezzabile sia da coloro che si occupano della tecnica sia da coloro che con l'etica e la morale misurano il consenso ai progetti.

I sistemi di riconoscimento presentano dei *bias* di età, di genere, di razza, che derivano dalla composizione del dataset su cui l'algoritmo viene addestrato. È possibile ridurre l'impatto etico e morale intervenendo a livello software per far sì che gli algoritmi possano apprendere a partire da dataset rappresentativi della società e quindi più inclusivi rispetto a quelli attuali. È importante una cooperazione e un lavoro in sinergia tra le varie parti e figure, nonché ruoli, che sono coinvolti in un progetto così promettente e che potrebbe in futuro rappresentare uno sviluppo importante nella tecnologia e, in particolare, nella mobilità.

BIBLIOGRAFIA

- Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J.-F., & Rahwan, I. (2018). The Moral Machine experiment. <https://doi.org/10.1038/s41586-018-0637-6>
- Brandao, M. (2019). *Age and gender bias in pedestrian detection algorithms* arXiv. <http://arxiv.org/abs/1906.10490>
- Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of the 1st Conference on Fairness, Accountability and Transparency* <https://proceedings.mlr.press/v81/buolamwini18a.html>
- Carfagna B. (2018). *Democrazia digitale, la seconda fase*. <https://www.tau.ac.il/~liort/images/Public%20Appearance/Democrazia-digitale-la-seconda-fase.pdf>
- Crosato, C. (2021). Introduzione: Se la rivoluzione informazionale si rivela troppo conservativa. <https://doi.org/10.17454/pam-2012>
- Danks, D., & London, A. J. (2017). Algorithmic Bias in Autonomous Systems. *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence* <https://doi.org/10.24963/ijcai.2017/654>
- Fabbrizzi, S., Papadopoulos, S., Ntoutsis, E., & Kompatsiaris, I. (2022). *A Survey on Bias in Visual Datasets* arXiv. <http://arxiv.org/abs/2107.07919>
- Fang, R., & Cai, C. (2021). Computer vision based obstacle detection and target tracking for autonomous vehicles. *MATEC Web of Conferences*. <https://doi.org/10.1051/mateconf/202133607004>
- Goodall, N. J. (2014). Ethical Decision Making during Automated Vehicle Crashes. *Transportation Research Record: Journal of the Transportation Research Board*, <https://doi.org/10.3141/2424-07>

- Gurumurthy, K. M., & Kockelman, K. M. (2020). *Modelling Americans' autonomous vehicle preferences: A focus on dynamic ride-sharing, privacy & long-distance mode choices*.
<https://doi.org/10.1016/j.techfore.2019.119792>
- Harakeh, A., Smart, M., & Waslander, S. L. (2019). *BayesOD: A Bayesian Approach for Uncertainty Estimation in Deep Object Detectors* arXiv. <http://arxiv.org/abs/1903.03838>
- Héder, M. (2020). The epistemic opacity of autonomous systems and the ethical consequences. *AI & Society*. <https://doi.org/10.1007/s00146-020-01024-9>
- Kant, I. (1788). *Critica della ragion pratica*. LaTerza.
- Kaplan, A., & Haenlein, M. (2019). Rulers of the world, unite! The challenges and opportunities of artificial intelligence. <https://doi.org/10.1016/j.bushor.2019.09.003>
- Klobas, J. E., McGill, T., & Wang, X. (2019). How perceived security risk affects intention to use smart home devices: A reasoned action explanation. *Computers & Security*.
<https://doi.org/10.1016/j.cose.2019.101571>
- Kohli, P., & Chadha, A. (2020). Enabling Pedestrian Safety Using Computer Vision Techniques: A Case Study of the 2018 Uber Inc. Self-driving Car Crash. *Advances in Information and Communication*. https://doi.org/10.1007/978-3-030-12388-8_19
- Liu, P., Yang, R., & Xu, Z. (2019). Public Acceptance of Fully Automated Driving: Effects of Social Trust and Risk/Benefit Perceptions. <https://doi.org/10.1111/risa.13143>
- Meyer-Waarden, L., & Cloarec, J. (2022). “Baby, you can drive my car”: Psychological antecedents that drive consumers’ adoption of AI-powered autonomous vehicles. *Technovation*.
<https://doi.org/10.1016/j.technovation.2021.102348>
- Ntoutsis, E., Fafalios, P., Gadiraju, U., Iosifidis, V., Nejdil, W., Vidal, M.-E., Ruggieri, S., Turini, F., Papadopoulos, S., Krasanakis, E., Kompatsiaris, I., Kinder-Kurlanda, K., Wagner, C., Karimi, F., Fernandez, M., Alani, H., Berendt, B., Kruegel, T., Heinze, C., ... Staab, S.

- (2020). Bias in data-driven artificial intelligence systems—An introductory survey. *WIREs Data Mining and Knowledge Discovery*, <https://doi.org/10.1002/widm.1356>
- Passagnoli, G. (2019). *Ragionamento giuridico e tutela nell'Intelligenza Artificiale*. <https://core.ac.uk/download/pdf/301579331.pdf>
- Popper, K. (1978). *Natural Selection and the Emergence of Mind*. laboratoriogene.info/Ciencia_Hoje/Popper1978.pdf
- Postels, J., Ferroni, F., Coskun, H., Navab, N., & Tombari, F. (2019). Sampling-Free Epistemic Uncertainty Estimation Using Approximated Variance Propagation. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, <https://doi.org/10.1109/ICCV.2019.00302>
- Saxena, N., Huang, K., DeFilippis, E., Radanovic, G., Parkes, D., & Liu, Y. (2019). *How Do Fairness Definitions Fare? Examining Public Attitudes Towards Algorithmic Definitions of Fairness* <http://arxiv.org/abs/1811.03654>
- Schiele, B., Zang, S., & Benenson, R. (2017). *CityPersons: A Diverse Dataset for Pedestrian Detection*.
- Schrumpf, M., Kubilius, J., Hong, H., Majaj, N. J., Rajalingham, R., Issa, E. B., Kar, K., Bashivan, P., Prescott-Roy, J., Schmidt, K., Yamins, D. L. K., & DiCarlo, J. J. (2018). *Brain-Score: Which Artificial Neural Network for Object Recognition is most Brain-Like?* *bioRxiv*. <https://www.biorxiv.org/content/10.1101/407007v1>
- Tamburrini, G. (2020). *Etica delle Macchine*. Carocci editore.
- Tamburrini, G., Fossa, F., & Schiaffonati, V. (2021). *Automi e persone. Introduzione all'etica dell'intelligenza artificiale e della robotica*. Carocci editore. <https://www.ibs.it/automi-persone-introduzione-all-etica-libro-vari/e/9788829011704>
- Thomson, J. J. (1985). The Trolley Problem. *The Yale Law Journal*.

Turing, A. M. (1950). I.—Computing machinery and intelligence. *Mind*.

<https://doi.org/10.1093/mind/LIX.236.433>

Uçar, A., Demir, Y., & Güzeliş, C. (2017). Object recognition and detection with deep learning for autonomous driving applications. *Simulation*. <https://doi.org/10.1177/0037549717709932>

Wilson, B., Hoffman, J., & Morgenstern, J. (2019). *Predictive Inequity in Object Detection*

<http://arxiv.org/abs/1902.11097>

Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks: An overview and application in radiology. *Insights into Imaging*,

<https://doi.org/10.1007/s13244-018-0639-9>

Ringraziamenti

A conclusione della trattazione, vorrei dedicare qualche riga a tutti coloro che mi hanno accompagnato in questo percorso formativo e di crescita personale.

Un sentito grazie al mio relatore, prof. Antonio Rodà e alla correlatrice, prof.ssa Silvana Badaloni per la loro disponibilità e tempestività ad ogni mia richiesta. Grazie per avermi fornito ogni materiale utile alla stesura dell'elaborato.

Voglio ringraziare i miei genitori per il loro supporto e vicinanza, soprattutto nei momenti più difficili.

Grazie di cuore a mio fratello Andrea, mia sorella Sara per avermi sempre sostenuto e per aver creduto in me.

Un ringraziamento speciale va a Francesco, per avermi trasmesso la sua forza e il suo coraggio.

Ringrazio Elena e Anna per aver condiviso momenti di studio e per essermi state accanto in questo periodo intenso e per gioire, insieme a me, dei traguardi raggiunti.

Grazie a tutti

Giusi