UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Diritto Pubblico, Internazionale e Comunitario

Dipartimento di Matematica “Tullio Levi-Civita”

Corso di Laurea in Diritto e Tecnologia

Analisi dei Sistemi di Guida Autonoma Attraverso un Approccio Neuro-Simbolico

Relatore: Prof. Roberto Confalonieri

Laureando: Simone Zanetti

Matricola 2019845

Anno Accademico 2022/2023

Sommario

CAPITOLO 1
INTRODUZIONE

1.1 Motivazione

Negli ultimi anni, il notevole sviluppo dei modelli di Deep Learning ha portato a una vera e propria rivoluzione nell'ambito dell'Intelligenza Artificiale. Tuttavia lo sviluppo di questi modelli è incentrato sugli aspetti commerciali e industriali, senza considerare le importanti questioni relative alla salute e ai diritti degli utenti. Particolarmente critico è il caso dei sistemi per la guida autonoma. Ho scelto di condurre una ricerca in questo campo per comprendere le principali sfide e problematiche, con la speranza che future ricerche nell'ambito Neuro-Simbolico possano apportare miglioramenti significativi nei processi decisionali delle auto a guida autonoma. Confido che lo sviluppo di nuove e più raffinate tecniche possa accrescere la fiducia degli utenti, migliorare l'usabilità, la robustezza e la spiegabilità di questi sistemi, contribuendo così alla comprensione delle scelte che questi stessi prendono facilitando anche l'applicazione delle leggi pertinenti in ambito legale.

1.2 Obiettivi

Partendo dall'analisi dei sistemi attuali di Intelligenza Artificiale, questo lavoro si propone di mettere in evidenza le principali problematiche, concentrandosi soprattutto sui sistemi di Machine Learning (ML) e, in particolare, sulla sottocategoria del Deep Learning (DL). Una volta compreso perché tali sistemi possono risultare opachi nelle decisioni che prendono, l'obiettivo del lavoro è introdurre l'approccio Neuro-Simbolico come una possibile soluzione a questi problemi. Inoltre, questa ricerca mira a dimostrare come tale approccio non solo aumenti la robustezza e la capacità di generalizzazione dei sistemi di AI, ma anche apporti benefici nel campo dell'Explainable AI (XAI), migliorando il rapporto di fiducia tra utenti e macchine, particolarmente rilevante per la diffusione su larga scala dei sistemi di guida autonoma. Si vuole anche evidenziare come, nel caso di dispute e questioni legali che spesso riguardano la determinazione della responsabilità in caso di incidenti in relazione all'utilizzo di questi sistemi, l'uso di metodi simbolici possa migliorare l'interpretazione del processo decisionale delle auto a guida autonoma, riducendo l'effetto "black-box" associato alle reti neurali artificiali e consentendo di stabilire l'attribuzione della responsabilità in maniera equa.

1.3 Approccio

Verrà adottato un approccio multidisciplinare, che combina analisi di sistemi di auto a guida autonoma, explainable AI e metodo neuro-simbolico per affrontare il problema della comprensibilità e dell'interpretabilità delle decisioni prese dai veicoli autonomi.

Per quanto riguarda i sistemi di auto a guida autonoma analizzati nella tesi, è necessario sottolineare che si tratta di framework o lavori recentemente proposti e che sono stati sviluppati in ambienti simulati di guida. I dati utilizzati vengono riportati nella tabella (fig.20). Questi sistemi utilizzano un approccio neuro-simbolico per modellare il processo decisionale del veicolo autonomo, combinando elementi di reti neurali e rappresentazioni simboliche per creare un modello ibrido che sia in grado di apprendere dai dati e di spiegare le decisioni in modo più trasparente. Attraverso l'analisi di questi sistemi cercherò di fare emergere i contributi che sistemi ibridi possono apportare in termini di spiegabilità ai sistemi neurali.

È importante osservare come questo approccio sia limitato dal fatto che i sistemi analizzati non sono attualmente adoperati in auto a guida autonoma, ma si tratta di proposte che mirano a risolvere problemi legati ad ambienti di guida non ancora testate in ambienti fisici. Inoltre, è importante notare che l'approccio neuro-

simbolico, sebbene promettente, può presentare alcune sfide in termini di complessità computazionale e di addestramento del modello e che la totale integrazione dei sistemi è tuttora un argomento in via di sviluppo e tema di ricerca trattato da molti esperti del settore (sezione 2.4).

1.4 Struttura

Nel capitolo 2, viene fornita una panoramica generale di Machine Learning (ML), Deep Learning (DL) e Intelligenza Artificiale (AI), con una distinzione tra AI simbolica e neurale, evidenziando i loro vantaggi e svantaggi. L'AI Neuro-Simbolica viene introdotta come un approccio ibrido che unisce queste due tecniche. Si utilizza la tassonomia di Kautz per categorizzare i sistemi AI e si discute l'importanza della Explainable AI (XAI) nell'ambito della guida autonoma. In ultimo si esplorano le sfide legali e regolamentari legate alle tecnologie di guida autonoma, focalizzandosi sulle differenze tra l'Europa e gli Stati Uniti. Si introduce brevemente il tema dei dilemmi etici e morali associati alle decisioni autonome dei veicoli e si propone l'dea di un ipotetico approccio neuro simbolico come soluzione a problemi etici.

Nel capitolo 3, si analizzano alcuni sistemi che adottano un approccio Neuro-Simbolico nell'ambito della guida autonoma, evidenziando in singole sezioni le caratteristiche principali e riflettendo sulle potenzialità di miglioramento. Successivamente, nella sezione 3.6, viene presentata una tabella valutativa che compara tra loro i sistemi. Infine, nel capitolo 4, vengono presentate le conclusioni e le possibili direzioni che potrebbero prendere futuri lavori.

CAPITOLO 2

RAGIONAMENTO E APPRENDIMENTO NEURO-SIMBOLICO PER XAI

2.1 AI, Machine Learning e Deep Learning

Per procedere all'analisi successiva dei sistemi di veicoli a guida autonoma mediante un approccio neuro-simbolico, ritengo sia utile innanzitutto acquisire una comprensione delle distinzioni tra modelli di natura simbolica e sotto-simbolica o neurale. In questa sezione, intendiamo innanzitutto soffermarci sulla disamina delle differenze tra Intelligenza Artificiale, Machine Learning e Deep Learning.

AI. La questione iniziale che ci poniamo richiede una veloce riflessione al fine di comprendere la natura dell'Intelligenza Artificiale e determinare le sue differenze rispetto al Machine Learning. Per affrontare questa questione, possiamo utilizzare tre semplici equazioni, sfruttando la loro natura simbolica:

- AI vs ML
- AI = ML
- AI ≠ ML

Ci si chiede se l'Intelligenza Artificiale sia in conflitto con il Machine Learning, se sia equivalente al Machine Learning o, infine, se sia qualcosa di distintivo rispetto al Machine Learning. La risposta a questo interrogativo può essere risolta attraverso una nuova equazione, la quale oltre a fornire una possibile risposta, introduce anche ad una definizione di Intelligenza Artificiale:

- AI ≥ 

Fondamentalmente l'obiettivo dell'Intelligenza Artificiale è superare o almeno eguagliare le capacità e abilità dell'essere umano. Lo studio e lo sviluppo dell'Intelligenza Artificiale mirano a ricreare la capacità di ragionamento e l'intelligenza umana, consentendo l'esecuzione di una vasta gamma di compiti (IBM s.d. a).

Machine Learning. Per Machine Learning si intende la capacità di effettuare predizioni e prendere decisioni basandosi su vaste quantità di dati (IBM s.d. b). Tale processo si basa su un'analisi dettagliata su un ampio insieme di informazioni. Il Machine Learning si suddivide ulteriormente in tre principali categorie a seconda del metodo di apprendimento: supervisionato, non supervisionato e con rinforzo. La distinzione primaria tra apprendimento supervisionato e non supervisionato risiede nel coinvolgimento umano. Nel Machine Learning supervisionato, esperti del settore etichettano i dati su cui il modello sarà successivamente addestrato, consentendo al modello di apprendere dalle informazioni fornite e di eseguire le attività apprese. Al contrario, nel Machine Learning non supervisionato, il modello è in grado di lavorare e svolgere compiti su dati che non sono stati etichettati, ovvero dati su cui il modello non possiede informazioni preliminari. Infine una spiegazione semplificata per la terza categoria è che in questo caso l'agente intelligente prende una decisione, analizza il cambiamento dello stato dell'ambiente valutando i feedback tramite una funzione di rinforzo. Attraverso 'ricompense' o 'penalizzazioni l'agente apprende cosa è giusto o sbagliato fare: obiettivo dell'agente risiede nella massimizzazione della funzione di rinforzo.

Deep Learning. Il Deep Learning è una sottocategoria del Machine Learning che incorpora le reti neurali. Si può parlare di 'deep' nel momento in cui i dati in entrata vengono processati attraverso più strati di elaborazione, in maniera che tra lo strato di dati in entrata (input layer) e risposta in uscita (output layer) del sistema, ci sia almeno uno strato nascosto di elaborazione di informazioni (fig.1). Le reti neurali sono algoritmi di apprendimento utilizzati all'interno dei modelli di Deep Learning. Tali algoritmi di apprendimento si basano su nodi interconnessi attraverso relazioni statistiche. Ciascun nodo ha associato un peso proprio,

il quale rappresenta il livello con cui influenzerà a livello locale la scelta finale del sistema. I pesi vengono modificati dalla rete al fine di apprendere evidenze statistiche specifiche dai dati, con l'obiettivo di emulare il processo con cui il nostro cervello elabora le informazioni ed eseguire compiti specifici del problema preso in considerazione.

Durante un processo di addestramento a partire da dati grezzi, le reti neurali artificiali acquisiscono conoscenze specialistiche sul dominio del problema e la capacità di generalizzare tali conoscenze a situazioni simili ma precedentemente non incontrate, in un modo che spesso supera le abilità degli esperti umani (Bader e Hitzler 2005).

Il procedimento descritto sopra viene effettuato su dati di addestramento che possiamo considerare come degli esempi dati in 'pasto' al modello con lo scopo di apprendere da questi dati e perfezionare la precisione nel tempo. Una volta ottimizzati per l'accuratezza, questi algoritmi di apprendimento si rivelano strumenti potenti nell'ambito della computer science e dell'IA, consentendo la rapida classificazione e raggruppamento di dati.

Possiamo trovare un'analogia tra questo processo e il funzionamento biologico del cervello umano (fig. 2). Ogni nodo corrisponde a un neurone, i dendriti ricevono input da altri neuroni sotto forma di impulsi elettrici, mentre attraverso il corpo cellulare ("cell body") gli input vengono elaborati e viene presa una decisione sull'azione da intraprendere. Infine, gli assoni trasmettono l'output in forma di impulso elettrico.

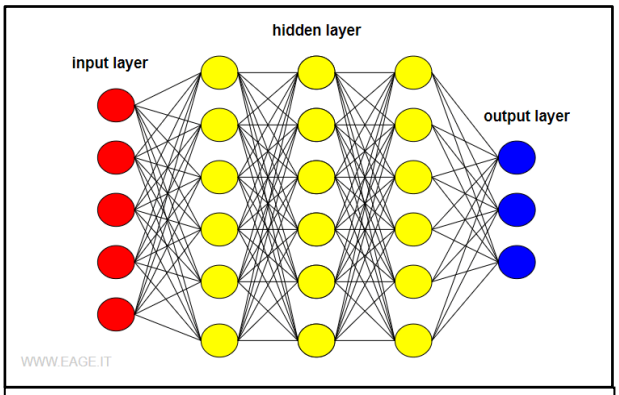

Fig. 1 – Artificial Neural Network.

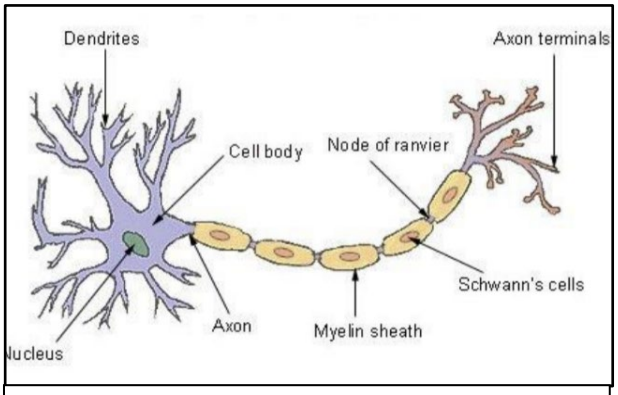

Fig. 2 – Biological Neuron.

Di seguito alcune applicazioni dell'Intelligenza Artificiale, (Zorzi 2022) che meritano di essere prese in considerazione:

- **Elaborazione del Linguaggio Naturale (Natural Language Processing).** Si concentra sulla comprensione e la generazione del linguaggio umano attraverso algoritmi e modelli computazionali.
- **Sistemi di Visione (Vision Systems).** Si riferisce a sistemi in grado di elaborare informazioni visive e distinguere oggetti e pattern visivi, consentendo la presa di decisioni informate.
- **Computer vision.** Consente di utilizzare immagini e filmati per eseguire analisi estremamente complesse che l'occhio umano non sarebbe mai in grado di fare. Le applicazioni comprendono la 'detection', applicabile per esempio al controllo degli accessi in azienda, ma anche la classificazione degli oggetti e l'analisi delle loro condizioni in ottica di manutenzione predittiva.
- **Trasformazione Testo-Voce (Text-to-Speech).** Coinvolge la capacità di convertire il testo scritto in output vocale, contribuendo all'interazione naturale tra umani e sistemi AI.

- **Movimento.** Questa abilità è ampiamente utilizzata nel campo della robotica, con l'obiettivo di replicare le capacità umane di movimento e interazione.
- **Sistemi di Guida Autonoma.** Questi sistemi sono progettati per consentire a veicoli di operare in modo autonomo, senza l'intervento diretto di un conducente umano. Essi sfruttano una combinazione di sensori, algoritmi e reti neurali per percepire l'ambiente circostante, interpretare segnali stradali, rilevare ostacoli e prendere decisioni per guidare in sicurezza e in conformità alle regole stradali.
- **Rappresentazioni della Conoscenza.** Creazione e utilizzo di strutture organizzate per rappresentare informazioni e conoscenza in modo che possano essere comprese e utilizzate da sistemi informatici. Le rappresentazioni della conoscenza consentono ai sistemi di ragionare, inferire e prendere decisioni basate su informazioni complesse, spesso utilizzando ontologie, grafi e altre strutture formali.

Guardando il diagramma (fig. 3), osserviamo che il Machine Learning è sottoinsieme dell'Intelligenza Artificiale. Quando utilizziamo tecniche di Machine Learning, stiamo effettivamente facendo Intelligenza Artificiale. Lo stesso vale per gli altri sistemi menzionati, che sono implementazioni di Intelligenza Artificiale.

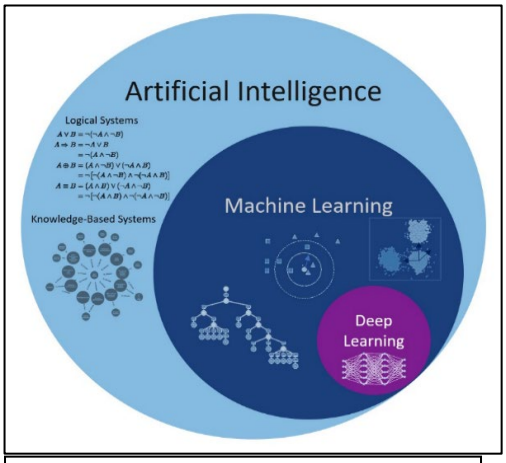

Fig. 3 – Diagramma di Veen su AI, ML e DL.

- AI vs ML
- AI = ML ⟹ ML ⊆ AI
- AI ≠ ML

2.2 AI simbolica e AI neurale. Approccio Top-Down e Bottom Up, verso una AI meno opaca

Il campo dell'intelligenza artificiale si suddivide in due macrocategorie con visioni contrastanti: l'IA simbolica e l'IA neurale. Un'analisi delle due categorie può aiutare a comprendere i possibili vantaggi offerti da un modello ibrido.

2.2.1 Ai Simbolica, approccio Top-Down

L'approccio dell'IA simbolica si basa sull'utilizzo di modelli simbolici per rappresentare, esplorare e acquisire conoscenza. Questo approccio segue una metodologia "top-down" (Minsky 1991), cui obiettivo è imitare la capacità umana di comprensione attraverso la creazione di regole e la definizione di concetti, con lo scopo di creare un approccio basato sulla logica classica, vista come metodo per rendere un sistema 'intelligente' attraverso il ragionamento. Si basa su una rappresentazione dichiarativa dell'intelligenza umana, fondandosi su fatti e regole. Alcuni esempi di applicazioni di questo approccio includono i sistemi esperti, i "decision support systems" e le "knowledge bases". In questa prospettiva, l'insegnamento di informazioni ad un sistema di intelligenza artificiale richiede di fornire ogni singola caratteristica, dettaglio o regola che il sistema utilizzerà per elaborare in modo corretto le informazioni presentate.

*Il punto fondamentale è che questi metodi, basati sulla logica e su modelli formali che fanno uso di simboli, sono in grado di fornire una rappresentazione della conoscenza facilmente interpretabile da parte degli esseri umani. Tutto ciò agevola il processo di acquisizione di nuova conoscenza attraverso strategie logiche quali deduzione, induzione e abduzione. T*ra i vantaggi più significativi emergono: la dichiarativà, l'osservabilità,

l'interpretabilità e l'affidabilità. Tuttavia sussistono anche svantaggi rilevanti, come la complessità nell'acquisizione delle regole e le limitazioni in termini di scalabilità.

2.2.2 Ai sub-simbolica, approccio Bottom-Up

La seconda corrente invece, fondata sull'utilizzo di reti neurali e metodi statistici, ha il grande vantaggio di poter acquisire la conoscenza da esempi, di essere scalabile e di riuscire a rappresentare basi di conoscenza complesse e altamente imprecise. Tali sistemi lavorano secondo un approccio 'Bottom-Up'. Questa modalità è basata su reti neurali addestrate con enormi quantità di dati e sull'assunzione che è possibile modellare i processi del cervello umano. Una rete neurale usa un elevatissimo numero di istanze annotate per scovare delle relazioni e creare un modello corrispondente; in questo modo l'intelligenza della rete è costruita sui dati e non su regole logiche.

Come viene sottolineato da Lake (Lake et al. 2017), questi sistemi riportano alcune mancanze, una volta eliminate le quali diventerebbero più efficienti e meno opachi. Nel suo lavoro le ritroviamo elencate in questo modo:

1) necessità di costruire modelli causali del mondo che sostengono spiegazioni e comprensione, piuttosto che risolvere semplicemente problemi di riconoscimento di pattern;
2) ancorare l'apprendimento in teorie intuitive di fisica e psicologia per supportare ed arricchire le conoscenze apprese;
3) sfruttare la composizionalità e l'apprendimento dell'apprendimento per acquisire rapidamente conoscenze e generalizzare a nuovi compiti e situazioni.

Reti neurali, modelli di insieme, modelli di regressione, alberi decisionali, macchine vettoriali di supporto sono alcuni dei modelli AI subsimbolici più popolari.

2.2.3 Simbolico e sub-simbolico. Verso un modello ibrido

La prima ondata di IA negli anni '80 era basata su logica simbolica e programmazione logica, e successivamente su reti bayesiane; la seconda ondata di IA negli anni 2010 era neurale (o connessionista), basata sull'apprendimento profondo. Avendo vissuto entrambe le ondate e avendo visto i contributi e gli svantaggi di ciascuna tecnologia, sosteniamo che è giunto il momento per la terza ondata di IA: l'IA neurale-simbolica (Garcez e Lamb 2023).

Per quanto riguarda l'implementazione dell'IA simbolica, uno dei linguaggi di programmazione logica più antichi, ma ancora tra i più popolari, è 'Prolog'. 'Prolog' ha le sue radici nella logica del primo ordine, una logica formale. Ogni spiegazione è corredata da una serie di fatti e di regole che hanno portato alla risposta finale. La trasparenza è parte fondante nel meccanismo stesso di computazione.

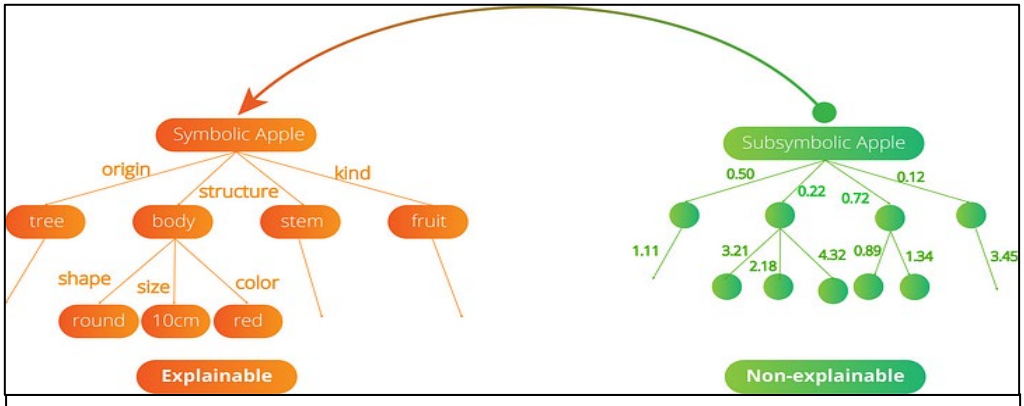

Fig. 4 - Differenze AI simbolica-Sub simbolica (XAI). Fonte (Yalçın 2021).

Osservando la figura 4, per quanto concerne il modello simbolico, la mela viene rappresentata come un grafo contenente caratteristiche e regole relazionali tra entità. Riusciamo così a definire una mela attraverso le sue caratteristiche. Se ad esempio chiedessi al sistema 'cosa è una mela'? Otterrei come risposta che si tratta di un 'frutto', che è 'verde', 'grande'... Il modello darà risposte sulla base di descrizioni simboliche correlate all' entità presa in considerazione. Qui ogni spiegazione è corredata da una serie di fatti e di regole che hanno portato alla risposta finale. La trasparenza è parte fondante nel meccanismo stesso di computazione. Secondo il modello sub-simbolico o neurale invece, si tratta solo di operazioni matematiche e statistiche che il sistema apprende nel corso del tempo al fine di riuscire a rappresentare una mela, cecando di estrarre il miglior modello possibile dai dati presentati in training. I 'pesi' ottenuti saranno associati al concetto di mela appreso dal sistema, rendendo il procedimento di interpretabilità e leggibilità di tali sistemi più opaco rispetto ai sistemi simbolici (Yalçın 2021).

Già nel 1988 Bazzocchi metteva in rilievo i vantaggi e gli svantaggi dei due approcci (Bazzocchi 1988):

Vantaggi	
BOTTOM UP	**TOP DOWN**
Riprende l'evoluzione naturale.	Grossi risultai pratici, anche se parziali.
Presenta 'solide' basi.	Simula comportamenti umani sofisticati.
Non ha bisogno di teoria ambiziose.	Pretende di spiegare l'intelligenza.
Costruisce 'autonomamente' livelli ulteriori.	Deve solo 'rafforzare' certe posizioni.

Svantaggi	
BOTTOM UP	**TOP DOWN**
Il processo può essere lento.	Risultati non siginificativi a livello di AI.
Legato al processo tecnico dell'hardware.	Utilizza metodi diversi dal comportamento simulato.
Non fornisce teorie esplicative.	All'aumento di prestazioni implica crescita esponenziale complessità, tempo elaborazione e memoria.
Sistemi numerici, spesso non spiegabili.	Necessita di teorie.

Fig. 5 – Pro e contro, Top Down vs Bottom Up. Elaborazione personale.

Più dettagliatamente le principali differenze tra ML simbolico e l'apprendimento profondo (Deep Learning e modelli sub-simbolici) sono:

1) La scelta della rappresentazione – Logica localista nel ML simbolico, distribuita nel caso del DL.
2) L'assenza di algoritmi di apprendimento basati su gradienti nel ML simbolico.

L'IA simbolica non riguarda solo regole di produzione scritte a mano e una definizione corretta di IA comprende anche la rappresentazione e il ragionamento sulla base di conoscenze, i sistemi autonomi multi-agente, la pianificazione e l'argomentazione, oltre all'apprendimento automatico (Garcez e Lamb 2023).

L'interrogativo che si presenta in questo momento è il seguente: come possiamo instaurare una comunicazione efficace tra questi sistemi al fine di sfruttare appieno il potenziale di entrambe le parti e amalgamarle in un singolo sistema ibrido? Quali sono le sfide e le complessità che emergono quando si cerca di unire tali sistemi? Inoltre, quali strategie e tecniche sono disponibili per ottimizzare l'efficienza di tale fusione?

2.3 Tassonomia ‘AI neuro simbolica’

Considerate le fondamentali disparità tra l'Intelligenza Artificiale di natura simbolica e quella sotto-simbolica e avendo consapevolezza che l'adozione congiunta delle loro caratteristiche possa costituire una buona soluzione per un'Intelligenza Artificiale scalabile, dotata di maggiore spiegabilità e interpretabilità, emerge ora la necessità di allargare l'orizzonte per quanto riguarda questo nuovo paradigma "neuro-simbolico". A tal fine, occorre condurre una classificazione adeguata di tali sistemi, al fine di sondarne la prospettiva di sviluppo futura elaborando una tassonomia di sistemi neuro-simbolici. In particolare, voglio soffermarmi su due approcci riguardo la visione di sistemi neuro-simbolici: gli studi di Garcez (Garcez et al. 2019) e il lavoro di Kautz (Kautz 2022).

Sintetizzando il più possibile quanto scritto da Garcez, tenendo in considerazione anche gli studi affrontati ad Gibaut nell’articolo ‘Neurosymbolic AI and its Taxonomy: a survey’(Gibaut et al. 2023), possiamo affermare che metodi riconducibili alla famiglia della NSC possono essere categorizzati sulla base del loro approccio alle seguenti sfide (fig. 6):

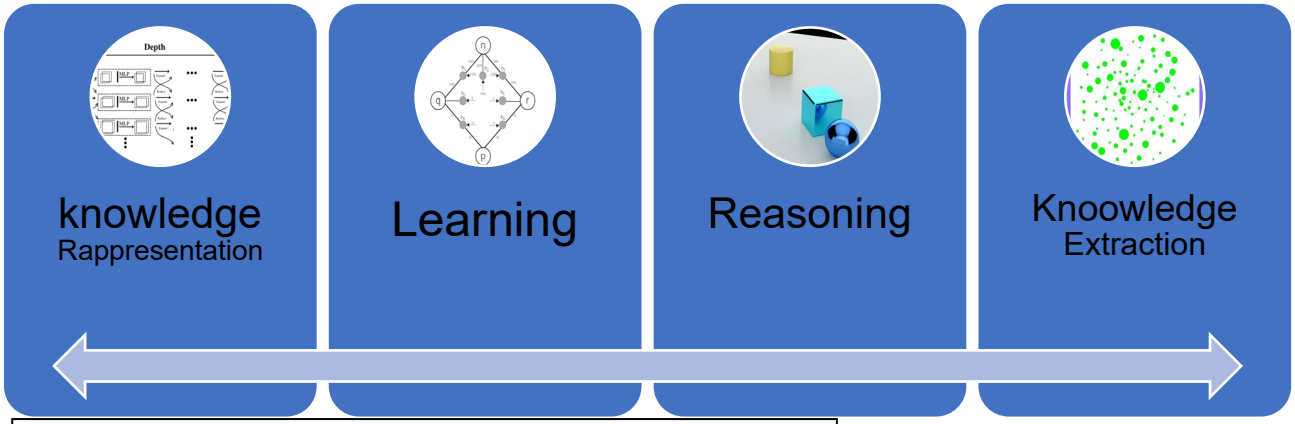

Fig. 6 – Sfide NSC. Elaborazione personale.

(1) **Rappresentazione della Conoscenza**. Viene affrontato il problema di come rappresentare la conoscenza basata su simboli e di come gestire la corrispondenza tra la conoscenza simbolica e le rappresentazioni neurali sub-simboliche o neurali.
(2) **Apprendimento**. L’attenzione è rivolta a comprendere come le componenti simboliche e sub-simboliche siano integrate nel processo di acquisizione di conoscenza.
(3) **Ragionamento**. Si riferisce alla capacità di una rete neurale di effettuare inferenze, simili ai motori logici tradizionali.
(4) **Estrazione della Conoscenza**. Capacità di un sistema NSC di convertire la conoscenza da un formato sub-simbolico a uno simbolico.

Kautz (Kautz 2022) invece categorizza i modelli di integrazione neurale-simbolica per tipi, a partire da quelli più semplici che operano su input e output simbolici fino a quelli più avanzati che cercano di integrare il ragionamento simbolico all'interno delle reti neurali:

Tipo 1 - Integrazione di Input e Output Simbolici con Reti Neurali. Questo rappresenta l'approccio standard del deep learning, in cui le reti neurali possono lavorare con input e output simbolici. Un esempio è la traduzione linguistica in cui il testo è trattato come simboli.

Tipo 2 - Sistemi Ibridi con Risolutori Simbolici, come ad esempio AlphaGo di DeepMind, in cui una rete neurale è collegata a un risolutore di problemi simbolici, la ricerca di albero di Monte Carlo, per affrontare compiti complessi.

Tipo 3 - Interazione tra Reti Neurali e Sistemi Simbolici Complementari. Questo tipo comprende sistemi in cui una rete neurale svolge un compito, ad esempio il rilevamento di oggetti, mentre un sistema simbolico specializzato risponde alle query basate sugli input e output della rete neurale.

Tipo 4 - Incorporazione di Conoscenza Simbolica nelle Reti Neurali. In questo caso, la conoscenza simbolica è integrata nell'addestramento delle reti neurali. Questo tipo include sistemi localistici, con il termine 'localistici' si intende sistemi in cui la conoscenza simbolica viene tradotta nei pesi iniziali della rete.

Tipo 5 - Vincoli Soft di Conoscenza Simbolica in Reti Neurali Distribuite. I sistemi di Tipo 5 sono distribuiti ma strettamente accoppiati. Qui, regole logiche simboliche vengono mappate su embedding che agiscono come vincoli soft sulla funzione di perdita delle reti neurali. Queste tecniche cercano di trasformare le rappresentazioni simboliche in spazi vettoriali in cui il ragionamento può avvenire attraverso calcoli matriciali su funzioni di distanza.

Tipo 6 - Ragionamento Simbolico all'interno di Reti Neurali. Il Tipo 6 rappresenta il livello più alto di integrazione, dove i sistemi sono in grado di eseguire il vero ragionamento simbolico all'interno di reti neurali. Questo è un obiettivo sfidante e in via di sviluppo.

Le tecniche e gli argomenti trattati sono in via di sviluppo, ci si sta spostando verso una collaborazione tra gli studiosi di AI simbolica e AI neurale al fine di garantire un modello ibrido e poter sfruttare le migliori capacità di entrambi. Migliori risultati potranno essere conseguiti negli anni a venire tramite ulteriore ricerca e studi in questo ambito: metodi completamente soddisfacenti per estrarre conoscenza simbolica da tali reti addestrate in termini di precisione, efficienza, comprensibilità delle regole e validità devono ancora essere trovati (Gibaut et al. 2023).

2.4 Ciclo Neuro-Simbolico

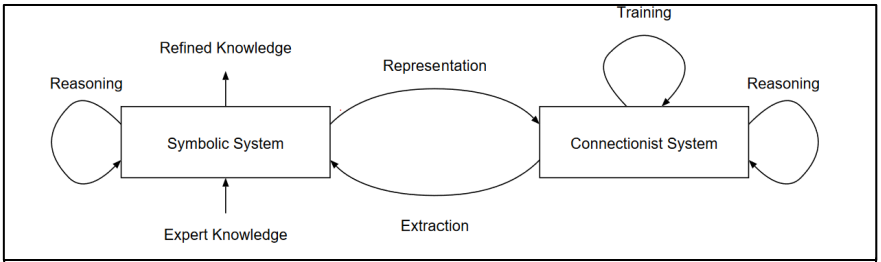

Fig. 7 – Ciclo Neuro-Simbolico. Fonte (Bader e Hitzler 2005).

È stato ora argomentato eloquentemente che la costruzione di un sistema AI completo, cioè un sistema AI semanticamente solido, spiegabile e in definitiva affidabile, richiederà uno strato di ragionamento solido combinato con l'apprendimento profondo (Garcez e Lamb 2023).

Per semplificare i concetti presenti nell'immagine (fig. 7), cerchiamo di stabilire un collegamento tra la parte destra dell'immagine, dove viene rappresentato un "sistema connessionista" basato su Machine Learning e Deep Learning, e la parte sinistra in cui viene mostrato un sistema basato su apprendimento e ragionamento basato su logiche simboliche. Nella parte destra il sistema a partire da grandi quantità di dati annodati utilizzando algoritmi di apprendimento, tra cui reti neurali, è in grado di effettuare previsioni accurate e migliorare in maniere autonoma l'accuratezza per svolgere attività e task specifiche. Dall'altro lato, troviamo i

"sistemi simbolici" i quali prevedono la rappresentazione e la manipolazione delle informazioni tramite simboli. L'elaborazione avviene su un piano logico, utilizzando simboli, connessioni astratte e relazioni logiche per creare nuove informazioni. In cui si prova a rendere il processo decisionale di un agente 'intelligente' spiegabile ed interpretabile da un essere umano, superando i limiti dei sistemi connessionisti.

In pratica un front-end (sistema simbolico) viene utilizzato per alimentare un sistema neurale o connessionista con conoscenze esperte simboliche (parziali), che può essere addestrato su dati grezzi. Le conoscenze acquisite attraverso il processo di apprendimento possono poi essere estratte e riportate al sistema simbolico (che ora funge anche da back-end), e rese disponibili per ulteriori elaborazioni in forma simbolica. In altri termini, i simboli che vengono assimilati, dedotti o persino creati possono funzionare come vincoli per la rete, contribuendo a potenziare le capacità di apprendimento in un ciclo che coinvolge apprendimento e ragionamento.

Nella figura 6 viene rappresentato una tipologia di ciclo neuro-simbolico, può essere interpretato secondo un sistema di tipo 3 (Tassonomia – 2.3), in cui una rete neurale svolge un compito, mentre un sistema simbolico specializzato risponde alle query basate sugli input e output della rete neurale. Parte della ricerca si concentra sulll' interpretazione di questo ciclo ponendo particolare attenzione sui procedimenti di *rappresentazione* ed *estrazione*, punti comunicanti tra i due sistemi al fine di permettere l'elaborazione delle informazioni a livello più basso (per percezione e riconoscimento di modelli efficienti) e la conoscenza astratta a livello più alto (per ragionamento, spiegabilità, estrapolazione e pianificazione) come risultato di una comunicazione diretta tra parte neurale e simbolica. Questo problema prende il nome di *integrazione*, esso può essere rappresentato come un sistema di tipo 6.

Il *ciclo di integrazione* neurale-simbolica include quindi (Besold et al. 2017):

- La traduzione della conoscenza simbolica nella rete;
- L'apprendimento di ulteriore conoscenza dagli esempi e la generalizzazione da parte della rete;
- L'esecuzione della rete (ragionamento);
- L'estrazione della conoscenza simbolica dalla rete. L'estrazione fornisce spiegazioni e facilita la manutenzione e l'apprendimento incrementale o il trasferimento.

Un'interessante proposta in merito alla questione aperta dell'integrazione di modelli neurali e simbolici è quella di Pober, Luck e Rodrigues (Pober, Luck, e Rodrigues 2022) in cui attraverso livelli di astrazione a partire da estrazione di feautures a livello sub-simbolico portano allo sviluppo di un linguaggio simbolico. Inoltre, nel lavoro ' *Dimensions of Neural-symbolic Integration — A Structured Survey'* viene presentato un framework in cui vengono fatti esempi tra modelli totalmente integrati e modelli ibridi e viene dimostrato che un modulo simbolico può essere integrato in modo pulito con un modulo neurale facilitando l'addestramento di quest'ultimo, ottenendo prestazioni empiriche che superano quelle dei lavori classici (Bader e Hitzler 2005).

2.5 XAI per implementare trasparenza nei sistemi a guida autonoma, prospettiva Neuro-Simbolica

Nonostante i recenti progressi, la guida autonoma è ancora lontana dall'essere pronta per soddisfare i requisiti di autonomia di livello 5. L'ostacolo principale è rappresentato dall'ambiente open-world in cui il veicolo autonomo deve operare. Navigare in questo ambiente richiede la capacità di prendere decisioni informate e agire in situazioni nuove e complesse.

2.5.1 XAI per auto a guida autonoma

Durante la fase di addestramento di una rete neurale artificiale, a partire dai dati grezzi, la stessa acquisisce competenze esperte nel dominio relativo a un problema specifico e apprende a generalizzare tali conoscenze, reagendo anche a situazioni mai prima incontrate, spesso in modo più efficace di quanto possa fare un essere umano. Tuttavia, le conoscenze assimilate dalla rete rimangono nascoste all'interno dell'architettura, essendo incapsulate nei pesi e nelle connessioni intrinseche alla rete stessa. Queste informazioni non sono direttamente accessibili per un'analisi esterna, il che rende oscure le decisioni prese dal sistema.

A supporto di tali problemi viene presentata l'Explainable AI, insieme di processi e metodi che si pongono come scopo quello di permettere agli utenti umani di comprendere e fidarsi dei risultati e delle uscite creati dagli algoritmi di apprendimento automatico. (Bader e Hitzler 2005)

Di seguito viene riportato un elenco delle principali decisioni prese da un sistema di auto a guida autonoma. Obiettivo della XAI in questo caso è dare una spiegazione dei punti sottoelencati (fig. 8):

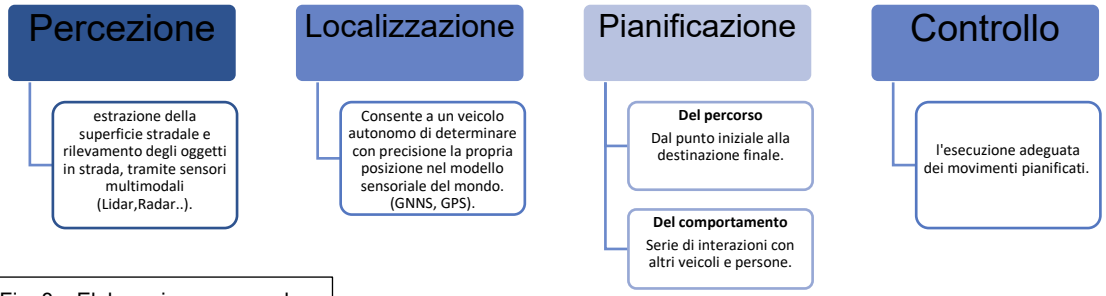

Fig. 8 – Elaborazione personale.

In merito a questi quattro aspetti, è importante tenere presente che l'IA svolge un ruolo fondamentale in compiti come la percezione dell'ambiente, il rilevamento degli oggetti e la pianificazione, portando a miglioramenti significativi in tali aree. Attualmente, obiettivo principale è concentrarci sulla spiegabilità di tali sistemi, al fine di ottenere l'approvazione sociale nei confronti di queste tecnologie. Invece di destare timori o preoccupazioni, è neceaasario che le persone si sentano a proprio agio con sistemi in grado di prendere decisioni cruciali per la loro salute. Le proprietà che un sistema XAI dovrebbe avere sono tuttora oggetto di discussione.

Atakishiyev ci presenta una tassonomia relativa agli 'stakeholders' coinvolti in tali sistemi, spiegando in che modo la XAI possa portare benefici alle diverse categorie di interessati (Atakishiyev et al. 2023). Nella figura 9 venongo riportate le proprietà di maggiore interesse. Inoltre, reputo essenziale, per la successiva analisi affrontata nel capitolo 3, fornire una visione complessiva delle diverse tipologie di spiegazioni, ciascuna delle quali varia in base alle specifiche esigenze dei destinatari (fig. 10).

Proprietà sistema per XAI	
Affidabilità	Capacità di dare sempre lo stesso risultato per lo stesso problema
Causalità	Capacità di correlare le variabili del problema in rapporti di causa-effetto.
Trasferibilità	Sapere esattamente cosa un modello rappresenta e come ragiona `e il primo passo per il riutilizzo di tale conoscenza in un altro frangente.
Informatività	Fornire informazioni sul problema gestito.
Equità	Fornire una garanzia di comportamento equo da un punto etico e morale.
Accessibilità	Implica la TRASPARENZA, primo passo verso dei sistemi realmente accessibili a tutti.
Interattività	La trasparenza garantirebbe un maggior grado di interattività con il sistema stesso da parte degli utilizzatori.
Privacy	L'opacità dei sistemi non permette infatti l'analisi delle informazioni che sono state effettivamente acquisite. Tramite XAI si prova a garantire la totale trasparenza su queste informazioni, eliminando ogni rischio per la privacy.

Fig. 9 – Proprietà XAI. Elaborazione personale.

Tipologie di spiegazioni	
Spiegazioni basate su filtri di causa	Basate sulla conoscenza disponibile. Le spiegazioni sono generate in base a filtri di causa come le domande "perché, "cosa succede se", etc. (ad esempio, "Perché l'auto ha preso la corsia sinistra invece di quella destra?").
Spiegazioni basate sul tipo di contenuto	Ad esempio: l'influenza dell'input, la sensibilità dell'input, le basi di caso e i fattori demografici (spiegazioni basate su quali variabili di input (cioè, caratteristiche di guida) contribuiscono di più alle azioni 'predictive').
Spiegazioni in base a un modello	Alcune operazioni di guida autonoma possono essere specifiche alle condizioni e alcune possono essere generali, indipendentemente dalle condizioni di guida (ad esempio, spiegare qualsiasi azione di guida autonoma, può essere specificata per questo gruppo di spiegazioni).
Spiegazioni basate sul tipo di sistema	Catturare le proprietà del sistema operativo. Basate sui dati (spiegando l'esito di un modello predittivo) o basate sugli obiettivi (spiegando i comportamenti di un agente basati sul raggiungimento del suo obiettivo in un ambiente predefinito).
Spiegazioni con interattività	Una volta fornita una spiegazione, un utente può porre ulteriori domande per comprendere meglio la spiegazione fornita.
Spiegazioni con uno scopo concreto	Fattibilità e gamma di spiegazioni che il sistema può generare, essendo locali o globali. Le spiegazioni locali sono limitate a spiegazioni su alcune o su un sottoinsieme di tutte le possibili azioni (cioè, spiegare una singola previsione in uno scenario di traffico specifico). Le spiegazioni globali, d'altra parte, sono in grado di spiegare tutte le decisioni di alto livello da un punto iniziale a destinazione.

Fig. 10 – Tipologie spiegazioni. Elaborazione personale.

In particolare, per quanto riguarda i veicoli a guida autonoma, la necessità di spiegazioni può essere sintetizzata attraverso tre prospettive distintive: la prospettiva psicologica, la prospettiva sociotecnica e la prospettiva filosofica. Considerando queste prospettive multidimensionali, la guida autonoma esplicabile può conferire i seguenti vantaggi ai soggetti interessati:

- **Progettazione orientata all'essere umano.** Interazioni tra l'essere umano e la macchina "user-friendly". Questo si traduce in un ambiente di guida autonoma più intuitivo e facilmente comprensibile per gli utenti umani.
- **Affidabilità.** La comprensione delle relazioni causali tra le azioni o le decisioni del sistema è una richiesta naturale da parte dell'essere umano.
- **Trasparenza e responsabilità.** Data l'impellente richiesta di spiegazioni derivante dagli standard normativi, come ad esempio il "diritto a una spiegazione" previsto dal GDPR, il concetto di responsabilità emerge come elemento cruciale. La responsabilità coniuga aspettative sociali e normative legislative nel panorama della guida autonoma, garantendo una maggiore trasparenza delle decisioni del sistema («Mercedes to Accept Legal Responsibility for Accidents Involving Self-Driving Cars» 2022).

2.5.2 XAI, Prospettiva Neuro-Simbolica

L'obiettivo nei prossimi decenni è quello di investigare e costruire sistemi di IA più avanzati che siano spiegabili, affidabili e basati su principi solidi che unifichino la capacità di imparare dall'esperienza e di ragionare su ciò che è stato appreso. In questo contesto, il Calcolo Neurosimbolico appare come un forte candidato per colmare queste lacune, integrando l'apprendimento dall'ambiente in un modo connessionista e il ragionamento su ciò che è stato appreso utilizzando l'elaborazione e la rappresentazione simbolica. (Gibaut et al. 2023)

Nei capitoli iniziali del nostro lavoro, ci siamo concentrati sulla comprensione dei sistemi neurali in relazione ai sistemi simbolici, analizzandone il funzionamento e le possibili applicazioni. Ora effettuiamo una breve osservazione su come la Computazione Neuro-Simbolica (CNS) possa portare vantaggi significativi al campo

dell'Explicable Artificial Intelligence (XAI). Sfruttando la caratteristica autoesplicativa dei sistemi simbolici, è possibile raggiungere un livello superiore di spiegabilità. Questa caratteristica è tipicamente assente nei sistemi neurali, che al momento costituiscono una parte fondamentale delle tecnologie impiegate nei veicoli a guida autonoma.

L'integrazione dei due livelli, come già anticipato nella sezione 2.4, potrebbe collegare l'elaborazione delle informazioni a basso livello, spesso coinvolta nella percezione e nel riconoscimento di modelli, con il ragionamento e la spiegazione a un livello superiore e più cognitivo. In particolare, l'integrazione tra questi due sistemi potrebbe contribuire a colmare alcune delle lacune presenti nei sistemi neurali, sfruttando comunque le potenti capacità di apprendimento e riconoscimento di modelli derivanti dai dati. Questa integrazione eleva il livello di ragionamento mediante l'uso dei simboli, aumentando la capacità di generalizzazione anche in situazioni "fuori dal comune", migliorando la trasparenza e la spiegabilità attraverso la progettazione di processi decisionali simbolici comprensibili per gli esseri umani. Questo approccio è ora oggetto di ricerca e in parte mira a mitigare l'effetto "scatola nera" associato ai sistemi neurali. (Confalonieri et al. 2021)

2.6 Riflessioni legali

Nei capitoli 2 e 3, è stata fornita una panoramica dei sistemi di Intelligenza Artificiale, con particolare attenzione alla materia di studio relativa all'Explicable AI (XAI) e agli approcci ad essa correlati, inclusi i sistemi neuro-simbolici. Ora, desideriamo concentrare la nostra attenzione su un nuovo aspetto relativo ai veicoli a guida autonoma, da un punto di vista principalmente legale. In questo contesto, intenderemo mettere in luce i punti critici dell'attuale sistema automobilistico, sottolineando i punti di forza che i sistemi autonomi potrebbero introdurre in questo settore. Presenteremo inoltre, al fine di avere una visione più estesa, le limitazioni e gli svantaggi di tali sistemi sia dal punto di vista tecnico che legale attraverso le regolamentazioni di principale importanza in contesto Europeo e statunitense, ponendo a confronto pro e contro delle corrispettive legislazioni.

2.6.1. Auto a guida autonoma come soluzione agli incidenti stradali

Nel corso del 2022, in Italia si sono verificati 3.159 decessi in incidenti stradali, registrando un aumento del 9,9% rispetto all'anno precedente. Inoltre, si contano 223.475 feriti, segnando un incremento del 9,2%, e 165.889 incidenti stradali, anch'essi in crescita del 9,2%. Le vittime in UE27 arrivano a contare 20669 persone, in aumento rispetto alle 19.932 del 2021 ma in diminuzione rispetto alle 22.761 del 2020. Negli Stati Uniti invece, la National Highway Traffic Safety Administration (NHTSA) ha pubblicato le sue ultime stime per i decessi dovuti a incidenti stradali nel 2022, stimando che 42.795 persone siano morte in incidenti stradali. («Drunk Driving | NHTSA» s.d.) Un altro importante aspetto sono le principali dinamiche attraverso le quali avvengono questi incidenti. In Italia tra i comportamenti errati alla guida si confermano come più frequenti la distrazione, il mancato rispetto della precedenza e la velocità troppo elevata. I tre gruppi costituiscono complessivamente il 38,1% dei casi (82.857). Da sottolineare come altri tipi di incidenti riguardino guida in stato di ebrezza o sotto effetto di stupefacenti, fattore vincolato totalmente alla presenza umana. (ISTAT, 2023).

Citando la “National Motor Vehicle Crash Causation Survey” effettuata nel 2008 dalla National Highway Traffic Safety Administration (agenzia governativa statunitense parte del Dipartimento dei Trasporti):

‘*L’errore umano è la causa determinante del 93% degli incidenti stradali’*

Ogni giorno, circa 37 persone negli Stati Uniti muoiono in incidenti causati da guidatori ubriachi, il che equivale a una persona ogni 39 minuti. Nel 2021 sono 13.384 le persone morte in incidenti stradali causati dall’ uso di alcol. Tutte queste morti potevano essere evitate.(«Drunk Driving | NHTSA» s.d.)

Sulla base dei dati sopra citati, occorre porre attenzione riguardo al problema delle vittime e dei feriti nelle dinamiche stradali al fine di ridurre tali eventi tragici. Diversi paesi stanno già lavorando per migliorare la sicurezza stradale, ad esempio attraverso l’obbligo di sistemi di guida più sicuri tramite sistemi ADAS («Cosa sono gli Adas e perché saranno obbligatori dal 2022» 2021) o ancora garantendo politiche stradali più sicure.

Ritengo che l'adozione su larga scala dei veicoli a guida autonoma potrebbe contribuire significativamente a ridurre gli incidenti stradali causati dall'errore umano, come la distrazione, il mancato rispetto delle regole e l'eccesso di velocità, rappresentando un potente strumento per la sicurezza stradale. Tuttavia, prima di valutare appieno i benefici, è necessario condurre analisi approfondite dei dati relativi agli incidenti causati dagli esseri umani rispetto a quelli generati da sistemi autonomi. Inoltre, è importante considerare l'impatto sull'occupazione nel settore automobilistico e definire linee guida internazionali per l'adozione di tali sistemi, oltre a rivedere l'infrastruttura stradale per adattarla alle esigenze dei veicoli autonomi.

2.6.2 ‘A.I. ACT’ e ‘Algorithmic Accountability Act’

‘Scopo del presente regolamento è promuovere l'adozione di un'intelligenza artificiale antropocentrica e affidabile e garantire un elevato livello di protezione della salute, della sicurezza, dei diritti fondamentali, della democrazia e dello Stato di diritto e dell'ambiente dagli effetti nocivi dei sistemi di intelligenza artificiale nell'Unione, sostenendo allo stesso tempo l'innovazione’ (Relazione sulla proposta di regolamento del Parlamento europeo e del Consiglio che stabilisce regole armonizzate sull’intelligenza artificiale (legge sull’intelligenza artificiale) e modifica alcuni atti legislativi dell’Unione 2023).

L'AI Act è una legislazione dell'Unione Europea (UE) che mira a regolare l'uso dell'intelligenza artificiale (IA) all'interno dell'UE (Commissione Europea 2021). Da un punto di vista giuridico, l’IA Act è stato definito il primo regolamento sull’IA al mondo e istituisce un quadro giuridico uniforme volto a regolare lo sviluppo, la commercializzazione e l’uso dei sistemi di IA, ovviamente, in conformità con i valori e i diritti dell’UE. Missione dell’AI act è di sviluppare una linea guida ed uno standard che possa essere applicato a livello globale alle AI, cercando un giusto compromesso tra la velocità della progressione tecnologica e la protezione dell’individualità umana e dei diritti fondamentali attraverso l’imposizione di vincoli e doveri nei confronti dei soggetti coinvolti nella creazione e distribuzione di sistemi AI. L’ AI act classifica le applicazioni dell'IA in quattro livelli di rischio: rischio inaccettabile, rischio alto, rischio limitato e rischio minimo o nullo. Tale classificazione è stata fatta in maniera tale che gli operatori che sviluppano o utilizzano sistemi di IA devono valutare con attenzione in quale categoria rientra il loro sistema per garantire il rispetto delle normative (Foti, Miriam 2023).

Vengono riportati i livelli di rischio di seguito:

- **inaccettabili.** Questa categoria comprende i sistemi di IA che presentano un rischio inaccettabile per i diritti, le libertà e la sicurezza delle persone. Ad esempio, i sistemi di IA progettati per manipolare le persone, discriminare in modo ingiusto o violare i principi fondamentali dei diritti umani rientrano in questa categoria. Sono severamente vietati nell'UE.
- **ad alto rischio.** L'alto rischio corrisponde a situazioni con un potenziale danno significativo alla salute, sicurezza o diritti fondamentali, derivante dalla gravità, probabilità, intensità e durata del danno e dalla sua portata. Sistemi ad alto rischio devono seguire rigorose normative, incluse valutazioni, documentazione, trasparenza, gestione dati, formazione e capacità di affrontare imprevisti. Gli obblighi includono valutazione e mitigazione dei rischi, informazioni chiare, documentazione dettagliata per valutazione da parte delle autorità e registrazione delle attività per tracciabilità.
- **Rischio limitato.** Questa categoria copre i sistemi di IA che comportano rischi minori e sono soggetti a requisiti meno stringenti rispetto a quelli ad alto rischio.
- **Rischi minimi.** Questa categoria riguarda i sistemi di IA che comportano rischi minimi o nessun rischio per i diritti e la sicurezza delle persone. Per questi sistemi, l'AI Act impone requisiti molto limitati o nessun requisito specifico.

La normativa dell'Unione Europea cerca di regolare l'uso dell'intelligenza artificiale attraverso un sistema chiamato "risk-based approach", che impone obblighi di conformità diversificati in base al livello di rischio associato alle applicazioni software e all'intelligenza artificiale. L'"AI Act" dell'UE (Regolamento (UE) 2022/022) non classifica direttamente i veicoli a guida autonoma secondo i livelli di automazione precedentemente menzionati. Tuttavia, se tali veicoli soddisfano i criteri di rischio specificati nell'AI Act, saranno considerati sistemi di intelligenza artificiale ad alto rischio. È importante notare che il livello di rischio dovrebbe essere valutato in base ai livelli di SAE precedentemente discussi. Inoltre, mentre l'AI Act europeo fornisce un quadro generale per la regolamentazione dell'intelligenza artificiale nell'UE, i dettagli specifici sulla regolamentazione

dei veicoli autonomi possono variare tra leggi nazionali e regolamenti regionali sul traffico stradale e la sicurezza stradale.

In America invece è sempre sempre mantenuto un “light touch” sulla regolamentazione dell’IA, aumentando gli investimenti in ricerca e sviluppo, tentando di assicurarne un uso sicuro ed etico. Con l’era Biden, vi è stato un maggior focus verso l’etica e l’affidabilità dei sistemi di IA. Ritengo utile ora portare le principali differenze tra l’‘A.I act’ di cui abbiamo appena parlato e l’‘Alghoritmic Accountability Act’, disegno di legge che affronta preoccupazioni pubbliche riguardo all'ampio utilizzo di sistemi decisionali automatizzati (ADS) e propone che le organizzazioni che implementano tali sistemi debbano intraprendere diverse misure concrete per identificare e mitigare i rischi sociali, etici e legali.

Riportiamo le più importanti differenze come indicate nel lavoro di Mökander (Mökander et al. 2022):

A. L'EU AIA è stato proposto dal ramo esecutivo dell'Unione Europea, evidenziando un forte sostegno istituzionale all'atto (anche se probabilmente subirà modifiche prima di diventare legge). Al contrario, l'US AAA deve ancora ottenere il sostegno al Senato o alla Camera.
B. L'EU AIA è un documento lungo, a volte opaco, che cerca di stabilire regole per l'uso degli ADS e fornisce dettagli su come questi debbano essere applicati. In confronto, l'US AAA adotta un approccio relativamente di alto livello.
C. I merito principale dell'US AAA è che è formulato in termini di ADS anziché usare il termine più popolarizzato "sistemi di intelligenza artificiale" preferito dalla Commissione Europea. I due termini vengono spesso usati in modo intercambiabile nella letteratura. Tuttavia, il termine "sistemi decisionali automatizzati" cattura meglio le caratteristiche tecniche di interesse, che possono fare affidamento su una miscela eterogenea di algoritmi di apprendimento automatico e framework di argomentazione codificati. Concentrandosi sulla regolamentazione dei "processi decisionali critici" invece che sui "sistemi di intelligenza artificiale ad alto rischio", l'US AAA evita la questione ontologica di cosa sia un sistema di intelligenza artificiale e tiene conto del fatto che il livello di automazione nei processi decisionali è meglio compreso come una differenza di grado su uno spettro.
D. L'US AAA si applica solo alle "grandi aziende" che hanno un fatturato annuo superiore a 50 milioni di dollari, un valore patrimoniale superiore a 250 milioni di dollari o che trattano le informazioni di oltre 1 milione di utenti. L'EU AIA offre un modello migliore, imponendo requisiti coerenti a tutti gli ADS ma offrendo un supporto mirato alle PMI per ridurre i costi per garantire e dimostrare la conformità (EU AIA, Articolo 55).

Tuttavia, va notato che l'impegno attuale degli Stati Uniti nel regolamentare gli ADS è piuttosto limitato, e questo è un punto di riflessione importante. Gli ADS non influenzano solo i consumatori, ma coinvolgono una vasta gamma di cittadini, il che sottolinea l'ampiezza del loro impatto. Inoltre, la forza di una politica è strettamente connessa alle istituzioni che la sostengono. Mentre l'EU AIA è parte di un piano a lungo termine e completo dell'UE per modellare l'ecosistema digitale nell'Unione e oltre, l'US AAA sembra rappresentare solo un tentativo frammentato. Dopo l'introduzione del GDPR nel 2016, abbiamo visto un cosiddetto "effetto Bruxelles" in cui le aziende multinazionali hanno scelto di armonizzare le proprie pratiche globali di gestione dei dati con le leggi dell'UE per ragioni pratiche. Ritengo ci sia in futuro la possibilità che l'EU AIA possa avere un effetto simile.

2.6.3 Livelli SAE e Responsabilità

Nel sotto-capitolo 2.5 abbiamo visto le principali azioni legate agli aspetti di AI nelle auto a guida autonoma (Percezione, Localizzazione, Pianificazione, Controllo), inoltre nel sotto capitolo 2.4.2 vengono riportati i principali sistemi Hardware, i quali forniscono fonti di dati multimodali che tramite l'elaborazione di Reti Neurali e sistemi intelligenti portano alla presa di decisioni autonome.

La visualizzazione dei "livelli SAE" per le auto a guida autonoma è un elemento importante per comprendere la distribuzione della responsabilità in base al livello di automazione. Questo schema (fig. 11) aiuta nell'interpretazione delle leggi applicabili ai casi concreti che coinvolgono auto a guida autonoma. A seconda del livello di automazione, la responsabilità può spostarsi dal conducente (livello 0) alla casa produttrice (livello 5) in caso di malfunzionamento o incidente. È un modo efficace per delineare chi è responsabile in diverse situazioni e per stabilire standard chiari per la sicurezza e la regolamentazione.

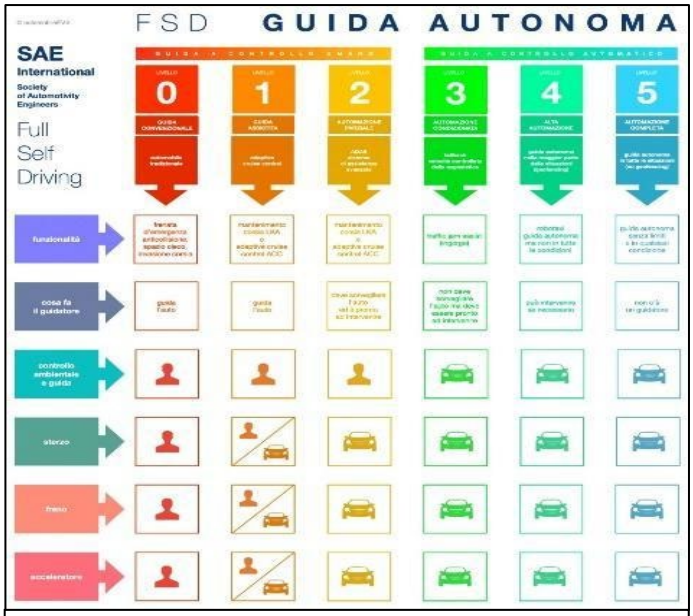

Fig. 11 - Livelli auto a guida autonoma SAE. Fonte («I livelli di guida autonoma di SAE International» 2021).

- Livello SAE 0. Nessuna automazione. Il conducente è responsabile di tutte le funzioni di guida.
- Livello SAE 1. Assistenza alla guida con sistemi che possono assumere temporaneamente il controllo dell'accelerazione, del freno o dello sterzo in determinate situazioni.
- Livello SAE 2. Guida semiautomatizzata con sistemi in grado di controllare accelerazione, freno e sterzo per brevi periodi, ma richiedono la supervisione continua del conducente.
- Livello SAE 3. Guida altamente automatizzata in cui il veicolo può gestire autonomamente l'accelerazione, il freno e lo sterzo in determinate condizioni, ma il conducente deve essere in grado di riprendere il controllo in caso di necessità.
- Livello SAE 4. Guida autonoma ad automazione completa in specifici scenari definiti, come autostrade, in cui il veicolo assume completamente il controllo e il conducente non deve supervisionare attivamente la guida.
- Livello SAE 5. Guida senza conducente, in cui il veicolo può guidare autonomamente in qualsiasi situazione e non richiede un conducente umano o controlli manuali. A questo livello si aprono scenari di nuove possibilità di trasporto per persone a mobilità ridotta, come pensionati, bambini o persone con diversi tipi di disabilità.

Responsabilità:

EU. La responsabilità legale nelle auto a guida autonoma è una questione complessa e controversa. A causa delle diverse modalità operative dei veicoli e della possibile necessità di intervento umano, stabilire chi è responsabile in caso di incidente è una sfida significativa. La responsabilità può variare in base al livello di

automazione del veicolo e alle circostanze dell'incidente. Ad esempio, il conducente potrebbe essere ritenuto responsabile se non risponde adeguatamente a un avviso del sistema autonomo, mentre il produttore potrebbe essere considerato responsabile in caso di malfunzionamento del sistema autonomo. La definizione chiara delle leggi e delle normative è essenziale per affrontare questa complessa questione.

L'introduzione dei veicoli autonomi complica la questione della responsabilità legale, creando una serie di possibili soggetti ritenuti responsabili in caso di incidenti, noto come il "problema filosofico delle molte mani"(Dastani e Yazdanpanah 2023). Le leggi e le normative variano da paese a paese, con alcune giurisdizioni che stanno elaborando leggi specifiche per affrontare questa sfida, mentre altre si affidano alle leggi esistenti. La responsabilità legale nelle auto a guida autonoma è una materia in evoluzione che richiede una revisione delle normative per adattarsi alle sfide uniche di questa tecnologia emergente. Ad esempio la Francia propone una prospettiva di responsabilità penale in capo a questi sistemi, tra le lacune da colmare vengono poste in evidenza:

- la definizione di ciò che si qualifica come un veicolo autonomo;
- la nozione di chi sia il “conducente” di un veicolo autonomo;
- questioni di responsabilità in caso di incidente che coinvolga un veicolo autonomo;
- l’uso di veicoli autonomi in attività di trasporto commerciale.

In questo caso, *Il conducente di un veicolo autonomo sarà esente da responsabilità penale per le infrazioni che si sono verificate quando le funzioni di guida sono state delegate al sistema automatizzato. A meno che il conducente: ‘stesse esercitando il controllo dinamico del veicolo’, ‘non abbia ripreso il controllo dinamico quando richiesto dal regolamento’; ‘non abbia rispettato i richiami della polizia’* (Mastromatteo et al. 2021). Dando un’interpretazione al primo punto osservato sopra, ritengo comprenda i veicoli SAE fino al tipo 4, in questo caso il conducente era momentaneamente il soggetto che compie l’azione sovrastando le scelte o consigli del sistema autonomo. I punti due e tre invece sono chiari.

Dal punto di vista civilistico, è interessante osservare come la responsabilità civile si adatta ai nuovi sistemi di auto a guida autonoma. Un emendamento all'articolo 8 della Convention on Road Traffic, un trattato internazionale, ha introdotto regole che consentono ai conducenti di veicoli autonomi di sollevare le mani dal volante, a condizione che possano riprendere il controllo del veicolo in qualsiasi momento. In particolare, l’articolo viene così modificato: ‘*Ogni guidatore deve, in ogni momento, poter controllare il suo veicolo’ diviene ‘Ogni guidatore deve essere sempre presente e abile a prendere il controllo del veicolo, i cui sistemi devono poter essere scavalcati o spenti in qualsiasi momento’* (De Palma, Valeria 2017). Queste sottili modifiche dimostrano come la legislazione deve evolversi per affrontare le sfide poste dalla tecnologia emergente e un'armonizzazione delle leggi a livello europeo può essere la chiave per far progredire questa tecnologia.

USA. Negli Stati Uniti bisogna prendere in considerazione che vi sono diverse Normative statali, il che significa che ci possono essere differenze significative tra gli stati. Ritengo interessante osservare come alcuni stati, come California e Texas, hanno adottato leggi specifiche per regolamentare i veicoli autonomi e affrontare il tema della responsabilità (Crosley, Tom 2020). Rispetto all’ Europa sono molto più all’avanguardia, basti pensare ad esempio all’introduzione di taxi a guida autonoma nella città di San Francisco (Mickle, Lu, e Isaac 2023), dove Waymo(«Waypoint - The official Waymo blog: Waymo’s next chapter in San Francisco» s.d.) e Cruise forniscono un servizio ai cittadini recentemente consolidato, anche se tra qualche contestazione.

Un caso emblematico che evidenzia le sfide legali delle auto a guida autonoma è l'incidente che coinvolse un veicolo Uber nel marzo 2018, causando la morte di Elaine Herzberg in Arizona (USA). In quell'incidente, il sistema Uber non aveva riconosciuto correttamente Herzberg come pedone, in parte anche a causa delle leggi locali che richiedono ai pedoni di dare precedenza alle macchine quando attraversano la strada al di fuori delle strisce pedonali contrassegnate: questo ha portato in parte ad una byas all' interno del sistema. Nonostante l'indagine successiva, nessuna delle parti coinvolte fu ritenuta penalmente responsabile dell'incidente(«Morte di Elaine Herzberg» 2023). Questo caso sottolinea l'importanza della sicurezza nell'adozione delle auto autonome e suggerisce che, anche se gli incidenti possono verificarsi, non dovrebbero scoraggiare l'adozione di questa tecnologia, poiché potrebbe contribuire in modo significativo a ridurre le vittime stradali. Difatti come abbiamo visto nella sezione 2.6.1, comportamenti errati condotti da umani alla guida sono la causa principale di incidenti e morti nelle strade di tutto il mondo. Le auto a guida autonoma hanno tra gli altri obiettivi la diminuzione di questi eventi, il singolo episodio non è nulla rispetto all' effetto che potranno avere a livello globale.

2.6.4 Privacy. 'GDPR' e 'Cloud Act'.

Negli Stati Uniti, la privacy dei dati è vista principalmente in un contesto commerciale, con un approccio orientato all'utilitarismo, in cui i dati sono considerati come beni da comprare e vendere. In contrasto, la legislazione europea sulla privacy riconosce la tutela dei dati personali come un diritto inalienabile degli individui. In Europa, il consenso degli utenti è richiesto persino per la semplice raccolta dei dati, mentre negli Stati Uniti la protezione dei dati è concentrata sull'uso dei dati e la raccolta può avvenire in modo più indiscriminato. Un esempio di questa differenza di approccio è il "Cloud Act" negli Stati Uniti («CLOUD Act» 2023), che permette alle forze dell'ordine americane di acquisire dati da servizi cloud, anche se i dati appartengono a utenti in altre nazioni, se il servizio opera nel mercato americano. Questo è visto come un conflitto con i diritti umani fondamentali e le normative europee sulla privacy, come il GDPR che stabilisce standard più rigidi per la protezione dei dati personali: difatti il Cloud Act non rispetta gli standard minimi della privacy violando l'articolo 48 del GDPR nel trasferimento dei dati personali verso paesi terzi (Regolamento (UE) 2016/679 2018).

Le principali differenze tra le regolamentazioni della privacy tra USA e EU sono:

GDPR vs. Leggi statali negli Stati Uniti. In Europa, il Regolamento generale sulla protezione dei dati (GDPR) è una normativa unificata che stabilisce standard rigorosi per la protezione dei dati personali. In contrasto, negli Stati Uniti, non esiste una legge federale sulla privacy unificata. La privacy è regolamentata principalmente a livello statale, il che significa che ci sono leggi sulla privacy diverse in vari stati e settori.

Consenso e Opt-in vs. Opt-out. Il GDPR impone una rigorosa politica di consenso, il che significa che le organizzazioni devono ottenere il consenso esplicito degli individui per raccogliere e utilizzare i loro dati personali. Negli Stati Uniti, il modello di "opt-out" è più comune, il che significa che le organizzazioni possono raccogliere i dati a meno che le persone non scelgano di rinunciarvi.

Diritto all'oblio e diritto all'accesso. Il GDPR conferisce agli individui il diritto di richiedere la cancellazione dei propri dati personali ("diritto all'oblio") e il diritto di accedere ai propri dati. Negli Stati Uniti, la legislazione sulla privacy non offre sempre tali diritti in modo così esteso o uniforme.

Inoltre interessante è osservare che nel GDPR è presente il 'diritto alla spiegazione'(Rubechini 2020), questo è in forte contrapposizione con le tecniche di DL che abbiamo visto finora.

Entità regolamentari. In Europa, l'Autorità europea per la protezione dei dati (EDPB) e le autorità nazionali per la protezione dei dati supervisionano e applicano il GDPR. Negli Stati Uniti, le autorità statali, come gli uffici del procuratore generale, hanno giurisdizione sulla privacy dei dati personali, e ci sono anche organizzazioni come la Federal Trade Commission (FTC) che affrontano questioni di privacy a livello federale.

Sanzioni e multe. Il GDPR prevede multe significative per le violazioni della privacy, che possono arrivare fino al 4% del fatturato annuo globale dell'azienda. Negli Stati Uniti, le sanzioni variano a livello statale e possono essere meno severe.

Le auto a guida autonoma comportano la raccolta e l'elaborazione di una vasta quantità di dati, inclusi dati sulla posizione, dati dei sensori del veicolo e dati dei passeggeri. Di seguito vengono riportate alcune delle questioni più importanti legate alla privacy nelle auto a guida autonoma:

Raccolta dei dati. I veicoli autonomi sono dotati di numerosi sensori, telecamere e sistemi di monitoraggio per rilevare l'ambiente circostante e guidare in modo sicuro. Questi sensori possono raccogliere dati sulla strada, sugli altri veicoli, sui pedoni, dati sui passeggeri e su altre informazioni ambientali. La questione principale è come questi dati vengono raccolti, utilizzati e conservati.

Trasmissione dei dati. In quanto i dati raccolti dai veicoli autonomi possono essere trasmessi a server remoti o a terze parti per scopi di analisi, manutenzione e miglioramento dei servizi.

Normative sulla privacy. Le normative sulla privacy, come il Regolamento generale sulla protezione dei dati (GDPR) in Europa o leggi sulla privacy simili in altre giurisdizioni, possono applicarsi alla raccolta e all'elaborazione dei dati da parte dei veicoli autonomi. Le aziende che sviluppano e operano veicoli autonomi devono essere conformi a queste leggi e garantire che i dati dei passeggeri e dei conducenti siano trattati in modo legale ed etico.

Consenso e trasparenza. È importante che i conducenti e i passeggeri siano informati in modo chiaro su come vengono utilizzati i loro dati e che abbiano la possibilità di dare il loro consenso. La trasparenza sulle pratiche di gestione dei dati è essenziale per garantire la fiducia del pubblico.

La competizione tra stati per la supremazia nell'industria della tecnologia non dovrebbe essere la priorità. È fondamentale invece promuovere la cooperazione internazionale per sviluppare tecnologie che rispettino i requisiti di privacy degli utenti. Gli approcci neuro-simbolici, visti nei capitoli precedenti, possono essere applicati con successo alla guida autonoma e all'intelligenza artificiale in generale, contribuendo a risolvere questioni come il diritto alla spiegazione ex. Articolo 22 GDPR, l'opacità degli algoritmi e i problemi di responsabilità e privacy. Ritengo che la creazione di sistemi esplicabili possa affrontare efficacemente queste sfide.

2.6.5 Etica, sicurezza nelle auto a guida autonoma considerando modelli neuro-simbolici

L'introduzione delle auto a guida autonoma ha sollevato importanti questioni etiche riguardo ai principi che dovrebbero guidare le scelte di queste macchine in situazioni critiche. Questi dilemmi includono la priorità tra la sicurezza dei passeggeri e quella dei pedoni, sollevando interrogativi cruciali di responsabilità, sicurezza e giustizia sociale. L'etica nell'Intelligenza Artificiale (AI) si basa su principi come la trasparenza, il rispetto dei valori umani, l'equità, la sicurezza, la responsabilità e la privacy. Il Regolamento AI dell'Unione Europea, approvato nel 2022, mira a creare un quadro giuridico per l'AI per promuovere la fiducia e mitigare i potenziali danni. Comprendere questi principi etici è cruciale mentre si sviluppa un'AI eticamente corretta, e ciò dovrebbe essere guidato dai seguenti principi fondamentali. L'introduzione delle auto a guida autonoma suscita entusiasmo per la sicurezza stradale ma pone complessi dilemmi etici: dovrebbero dare priorità alla sicurezza degli occupanti o minimizzare il danno totale, anche mettendo a rischio i passeggeri? Quali sono le scelte moralmente più corrette che un sistema autonomo dovrebbe prendere? Queste sfide richiedono un'analisi approfondita, poiché coinvolgono responsabilità, sicurezza e giustizia sociale. Gli attori dell'industria e i legislatori cercano di sviluppare normative e linee guida per affrontare queste questioni etiche cruciali.

L'etica nell'ambito dell'Intelligenza Artificiale (AI) si riferisce allo studio dei principi morali, delle normative, degli standard e delle leggi che si applicano all'AI, basandosi su principi fondamentali quali la trasparenza, il rispetto dei valori umani, l'equità, la sicurezza, la responsabilità e la privacy (Dhirani et al. 2023). Il Regolamento AI dell'Unione Europea, approvato nel 2022, ha l'obiettivo di creare un quadro giuridico per l'AI allo scopo di promuovere la fiducia e mitigare i potenziali danni che queste tecnologie potrebbero causare. Poiché il Regolamento AI è ancora in fase di sviluppo, è di fondamentale importanza comprendere i principi etici che lo sottendono e considerare come sviluppare un'AI eticamente corretta. Al fine di raggiungere questo obiettivo, è essenziale seguire i seguenti principi guida:

- **Trasparenza.**
  Svolgendo un ruolo cruciale nel monitoraggio dei risultati e garantendo la loro conformità con i principi morali umani, in modo che si possa comprendere, percepire e riconoscere inequivocabilmente il meccanismo decisionale del design.
- **Rispetto per i Valori Umani.**
  Le invenzioni basate sull'IA sono tenute a difendere i valori umani e a influire positivamente sul progresso delle persone e delle industrie, garantendo allo stesso tempo la sensibilità alle diversità culturali e alle credenze.
- **Equità.**
  Promuovere un ambiente inclusivo privo di discriminazione nei confronti dei dipendenti in base al genere, al colore, alla casta o alla religione è essenziale.
- **Sicurezza.**
  La sicurezza riguarda sia la protezione delle informazioni degli utenti che il benessere delle persone
- **Responsabilità.**
  I procedimenti decisionali dovrebbero essere verificabili, in particolare quando l'IA gestisce informazioni private o sensibili, come il diritto d'autore o l'identificazione di informazioni biometriche o record sanitari personali.
- **Privacy.**
  La protezione della privacy dell'utente durante l'utilizzo delle tecniche dell'IA deve essere mantenuta come massima priorità, ponendo attenzione ai dati sensibili dell'utente finale.

Secondo uno studio condotto da Audi («Myths about Autonomous Driving» s.d.), poiché il veicolo non è una entità razionale, prenderà decisioni basate sul codice programmato; quindi, le decisioni saranno basate sulla progettazione dell'azienda automobilistica. "Il veicolo potrà solamente adottare decisioni etiche e valori definiti dalle persone che lo hanno progettato - e li applicherà senza interpretazioni personali". Da queste affermazioni comprendiamo anche la necessità di creare una regolamentazione a livello Europe 'risk based' che si approcci 'ex ante' allo sviluppo di questi sistemi.

Per illustrare meglio dei tipici problemi in questo contesto è possibile fare riferimento al classico "Trolley Dilemma" applicato alla guida autonoma. Questo dilemma si basa su un vagone del tram che non può essere fermato e può essere controllato tramite una leva posta all'esterno del veicolo. Se il vagone continua sulla sua traiettoria, colpirà cinque persone, ma se si decide di cambiare binario ne colpirà solo una. Questo rappresenta un esperimento mentale classico in cui non esiste una soluzione intrinsecamente corretta. Tuttavia, è interessante notare che, dal punto di vista cognitivo, è un problema che un essere umano può risolvere "facilmente".

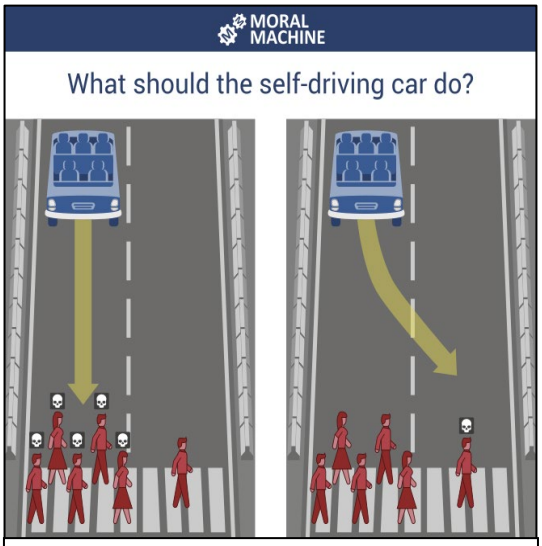

Fig. 12 - Decisioni morali in auto a guida autonoma. Fonte («Moral Machine» s.d.).

Nel dominio delle automobili autonome è possibile riprodurre il 'trolley dilemma' contestualizzato in ambiente stradale, come nell'esempio della figura 12.

Nel 2014, i ricercatori del MIT Media Lab hanno rielaborato il dilemma relativo alla guida autonoma: se un veicolo fosse costretto a fare una scelta tra investire delle persone, quale opzione sceglierebbe? Dovrebbe salvare cinque persone a scapito di una sola, oppure salvare la persona singola sacrificando il gruppo? Naturalmente, è essenziale anche proteggere la sicurezza delle persone all'interno del veicolo. Sul sito Moral Machine del MIT vengono presentate una serie di decisioni in cui l'auto si ritrova coinvolta in situazioni moralmente complesse. Interessante il test proposto nel sito, che crea una tabella valutativa finale delle scelte prese dall'utente in confronto con le scelte prese dalla totalità degli utenti: le scelte proposte sono fatte per mettere in difficoltà il soggetto sottoposto a prendere una decisione («Moral Machine» s.d.). Le entità automatizzate, in modo differente rispetto agli esseri umani, richiedono comunque una forma di 'coscienza' implementata nel codice, conformemente ai principi etici, al fine di poter prendere decisioni considerate adeguate dal punto di vista sociale. Si pone la questione di come fornire una risposta o una soluzione a questo interrogativo. È fattibile implementare un livello specifico di conoscenza all'interno dei sistemi di Intelligenza Artificiale per affrontare tali questioni e rendere una macchina in grado di affrontare sfide complesse comparabili a quelle umane?

La guida autonoma solleva importanti questioni etiche riguardo alla distribuzione del rischio tra conducenti, pedoni e ciclisti. È essenziale che i sistemi di guida autonoma ponderino le implicazioni etiche delle loro azioni, andando oltre il mero rispetto delle leggi stradali. In situazioni complesse, come un incidente imminente, i veicoli autonomi devono essere in grado di prendere decisioni che considerano i valori relativi alla sicurezza degli occupanti, dei pedoni e di altri utenti della strada. Ad esempio, un veicolo programmato per seguire alla lettera la legge rifiuterebbe di deviare attraverso la linea gialla continua, anche se ciò comportasse il rischio di investire un pedone ubriaco, e anche se l'altro lato della strada ha solo un'auto senza conducente che si sa

essere vuota. È evidentemente necessario rendere le auto a guida autonoma in grado di prender questo tipo di decisioni (Goodall 2016). L'incorporazione rigida delle leggi nel software potrebbe non essere la soluzione migliore, poiché i conducenti umani spesso considerano le leggi come flessibili in alcune situazioni. A differenza delle persone, che prendono queste decisioni istintivamente, i veicoli autonomi devono basare le loro decisioni su una strategia attentamente pianificata di gestione del rischio, considerando il valore degli oggetti e degli occupanti coinvolti in un incidente e la probabilità dell'evento temuto. Dai sistemi di auto a guida autonoma e la relativa presa di decisioni morali, il pubblico non si aspetta una saggezza sovrumana, ma piuttosto una giustificazione razionale per le azioni di un veicolo che tenga conto delle implicazioni etiche. Una soluzione non deve essere perfetta, ma dovrebbe essere ponderata e difendibile.

In ultimo, visti i problemi legati alla presa di decisioni talvolta cruciali da parte di questi sistemi, decisioni fortemente intrinseche alla sfera morale decisionale, ritengo che gli approcci discussi in questo lavoro possano migliorare la sicurezza delle decisioni delle auto a guida autonoma, affrontando le sfide legate alla presa di decisioni morali. Tuttavia, la capacità di prendere decisioni adeguate dipende anche dalla fase di addestramento del sistema, e non sarà mai possibile prevedere tutte le situazioni possibili sulla strada. In questo contesto, i modelli ibridi (che verranno esaminati in maniera approfondita nel capitolo 3) potrebbero offrire vantaggi significativi:

- In termini di ‘capacità di generalizzazione’, come visto in DRLSL for AD (sistema 5, situazione autostrada in senso di marcia opposto).
- Attraverso i ‘sistemi di tipo 2’, nella presa di decisioni attraverso ‘pensiero lento’, ovvero per quanto riguarda la presa di decisioni che intaccano la sfera di decisioni razionali e che richiedono un ragionamento diverso da quello classico. Attraverso agenti BDI.
- Sfruttando KG, come visto in KEP e CoSI, attraverso la creazione di vincoli relazionali per la presa di decisioni che richiedo un tipo di ragionamento più raffinato rispetto al semplice riconoscimento di pattern e stimoli da parte di reti neurali, tramite la comunicazione tra un sistema neurale e un sistema simbolico lo sfruttamento di ontologie adeguate e relazioni tra entità.

Portando un esempio concreto, nel lavoro di Swaminathan (Swaminathan s.d.), viene evidenziato come attraverso l’utilizzo di un sistema ibrido neuro-simbolico si sarebbe potuto evitare la morte di ‘Elaine Herzberg’ (sezione 6.3.2) nel caso contro UBER. Nell’ articolo viene ripetuto il concetto spiegato poco fa, ovvero che per quanto potenti siano le tecniche di DL, anche tali tecniche hanno dei limiti e vulnerabilità, ma ci sarà sempre una possibile singola situazione non conosciuta dall’algoritmo in cui non saprà come rispondere: questo può portare a errori anche fatali nel dominio di auto a guida autonoma. Ancora una volta vengono messe alla luce le potenzialità di un approccio ibrido per mitigare queste situazioni e problemi aumentando il livello di ragionamento e generalizzazione di questi modelli.

CAPITOLO 3

SISTEMI NEURO-SIMBOLICI PER AUTO A GUIDA AUTONOMA

Di seguito viene effettuata un'analisi di diversi sistemi sperimentali proposti negli articoli indicati nel titolo della sezione: le singole fonti di ogni sezione sono riportate nella tabella riassuntiva (fig. 20).

3.1 Sistema 1: 'KEP - Knowledge-based Entity Prediction for Improved Machine Perception in Autonomous System'

Ci troviamo all'interno della nostra nuova auto a guida autonoma, stiamo comodamente bevendo un caffè e leggendo un giornale mentre veniamo accompagnati in ufficio. Attraversiamo il quartiere residenziale, quando d'un tratto si intravede in lontananza un pallone da basket che rimbalza in prossimità della strada. La strada è vuota e il limite è di 50 km orari, limite che la macchina a guida autonoma sta rispettando. Un bambino spunta da dietro un palo della luce e corre in mezzo alla strada per recuperare un pallone. Noi in quanto umani siamo abituati a riconoscere questo tipo di situazioni e anzi, troviamo logico ricondurre la possibile presenza di un bambino o di un soggetto in prossimità di un pallone che si muove. Ma le macchine esattamente come possono arrivare a questo tipo di conclusione? Come possono evitare e fare previsione su possibili entità e situazione strettamente correlate con la scena presa in considerazione? Il problema sopra descritto comprende il modulo di percezione dei sistemi di auto a guida autonoma (sezione 2.5), che elabora i dati multimodali forniti da modelli (Lidar, Radar, GPS…), attraverso i quali vengono classificati gli oggetti e entità rilevati in scena. Le informazioni rilevate a livello percettivo verranno poi propagate agli strati successivi di pianificazione e controllo. Osserviamo come nel momento in cui vi è un errore a livello percettivo (fase 1), ad esempio il riconoscimento di un'entità non riconosciuta, questo errore verrà propagato ai moduli successivi e cambierà l'obiettivo desiderato. Nel caso della guida autonoma, ciò potrebbe portare a un incidente mortale.

Definiamo ora KEP (Wickramarachchi, Henson, e Sheth 2022) come 'compito di prevedere l'inclusione di entità potenzialmente non riconosciute in una scena', data la conoscenza attuale e di 'background' della scena rappresentata sotto forma di un grafo di conoscenza. Tale sistema mira a migliorare la percezione delle macchine autonome con lo scopo di completare la conoscenza in situazioni di non chiarezza e comprensione delle scene in input, ad esempio aiutando a prevedere situazioni in cui potrebbero esserci pedoni o oggetti non riconosciuti. Tramite l'utilizzo di un KG vengono percepite delle entità potenzialmente non riconosciute nella scena, vengono fornite rappresentazioni espressive e complete della conoscenza della scena con lo scopo di applicare l'apprendimento arricchito di conoscenza inducendo il completamento della conoscenza, prevedendo la conoscenza mancante nel grafo. Ad esempio, le percezioni ambientali delle scene (riconosciute dagli algoritmi di visione artificiale) (PCV) vengono rappresentate all'interno di un grafo di conoscenza (KG) (IBM s.d. c). Il processo di KEP prevede quindi un nuovo insieme di percezioni (PKEP) per le entità potenzialmente non riconosciute, ossia oggetti ed eventi non riconosciuti. Queste percezioni (PKEP) vengono utilizzate per arricchire il KG della scena aggiungendo nuovi nodi e relazioni. Si osservi che questo è un processo ciclico in cui il KG della scena viene aggiornato con nuovi fatti in modo continuo nel tempo.

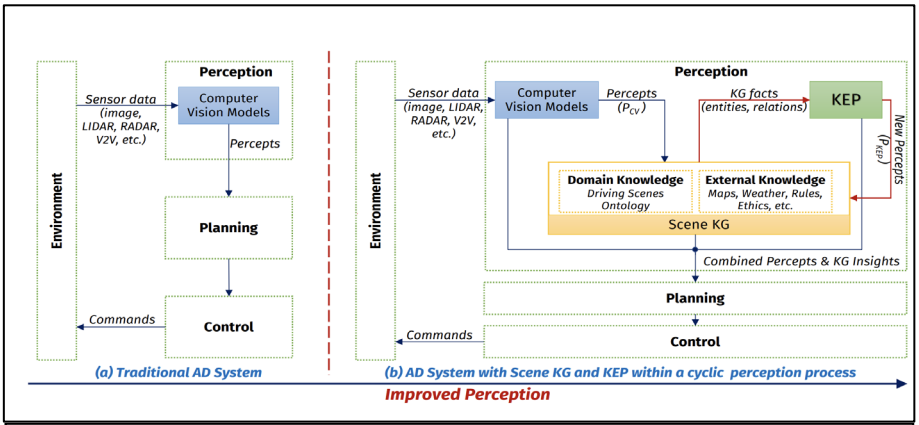

Fig. 13 – A) Traditional AD system. B) AD System with KG+KEP, cyclic perception process.

Tramite la combinazione di questi due modelli e integrando il tutto all' interno di un KG che fa da tramite tra la fase di percezione e pianificazione, *si tenta di instaurare nel sistema una capacità di ragionamento più elevate nella percezione e comprensione dell'ambiente e delle relative entità e relazione infra-entità presenti nella scena, mantenendo le potenti caratteristiche di riconoscimento di pattern dei modelli di PCV*. Questo si fa attraverso tre approcci, descritti nel testo citato: 'Knowledge Graph Embeddings' (KGE), che rappresenta le relazioni tra entità in un grafo; 'Association Rule Mining' (ARM), tecnica di data mining che trova correlazioni tra insiemi di elementi; 'Collective Classification', che cerca di assegnare etichette corrette a entità in un grafo basandosi sulle etichette di entità conosciute.

In questo modo i moduli di Planning e Controllo beneficeranno della nuova conoscenza appresa a livello di percezione. Tutto questo viene fatto in maniera ciclica. Il KG della scena viene aggiornato con nuovi fatti in modo continuo nel tempo e la comprensione del modulo di percezione migliora man mano che il KG viene completato.

Come risultato di questo processo:

- Il modulo di percezione viene arricchito con entità potenzialmente mancanti.
- I processi successivi (cioè pianificazione e controllo) beneficeranno dell'uso di un insieme più ricco di percezioni sull'ambiente. PCV + PKEP insieme ad altri indizi contestuali e di background dalla conoscenza relazionale nel KG della scena.

Tornando al caso in cui il modulo di percezione rileva una palla da basket a lato della strada e non riconosce che la presenza di un pallone può implicare la presenza di un'entità umana nelle vicinanze, perché magari in fase di addestramento nel modulo di percezione CV non si è mai presentata una situazione del genere, tramite le rappresentazioni della conoscenza di questo evento, ossia: (Pallone, Presenza_di, bambino), il KG della scena potrebbe fornire una intuizione o indicazione per gestire questo particolare caso tramite KEP.

Ritengo che ciò migliorerebbe significativamente la percezione da parte delle macchine nella guida autonoma (riducendo gli errori di classificazione errata, le omissioni e altre situazioni) e inoltre aumenterebbe la capacità di utilizzare la conoscenza relazionale e contestuale per generare spiegazioni delle previsioni, aumentando il grado di spiegabilità nel processo decisionale dell'auto a guida autonoma.

3.2 Sistema 2: ‘CoSI- A Knowledge Graph-Based Approach for Situation Comprehension in Driving Scenarios’

Un ulteriore problema nell'ambito delle strade riguarda l'interpretazione delle scene in presenza di numerose entità e/o veicoli, nonché la capacità di anticipare le azioni che tali entità potrebbero intraprendere nel breve termine all'interno delle situazioni di traffico stradale. La sicurezza stradale impone la necessità di acquisire una completa comprensione delle situazioni di guida, che comprende la percezione del traffico attuale, la comprensione del loro contesto e la capacità di anticipare gli sviluppi futuri. Spesso, tali competenze e la capacità di interpretare scenari di guida rimangono limitate per quanto riguarda i veicoli a guida autonoma di livello 5.

CoSI. (Intelligenza di Contesto e Situazione)

Questa è un approccio basato su un Grafo di Conoscenza (KG) che mira a integrare e strutturare informazioni eterogenee provenienti da diverse fonti e tipologie. In questo contesto, il KG assume il ruolo di uno strato di coerenza, rappresentando le informazioni attraverso entità e le loro relazioni, arricchite da assiomi semantici aggiuntivi. Un vantaggio evidente di questa modalità di rappresentazione dei dati è la possibilità di sfruttare il potere delle regole assiomatiche e delle capacità di ragionamento, consentendo l'inferenza di ulteriori conoscenze a partire da ciò che è già stato codificato. Componenti dedicati sfruttano queste informazioni semanticamente arricchite per svolgere compiti come la classificazione della situazione, la valutazione della difficoltà e la previsione della traiettoria. (Halilaj et al. 2021)

In questo a partire dal Dataset Sumo (Footnote4), in ambienti di guida simulata vengono effettuati dei test al fine di valutare tale sistema. Importante osservare come utilizzando tecniche di inserimento nel KG basate su un'architettura di Rete Neurale a Grafo (GNN) per un task di classificazione delle situazioni di guida otteniamo una precisione superiore al 95%, mentre gli approcci basati su vettori raggiungono solo il 75% di precisione per lo stesso task. Vediamo ora nello specifico come lavora questo sistema:

Comprensione della situazione. Viene descritto il processo di comprensione delle situazioni nel contesto delle auto intelligenti utilizzando un approccio basato su grafi di conoscenza. Viene riportata una panoramica riassuntiva dei principali punti trattati durante il processo di comprensione delle situazioni:

Osservazione Contestuale. Attraverso le sofisticate tecnologie sensoriali (Lidar, Radar, ecc.…) vengono raccolti dati sulle informazioni spaziali e temporali di ogni oggetto osservato.

Ingestione della conoscenza. La comprensione della situazione richiede l'integrazione e la strutturazione dell'abbondanza di informazioni provenienti da diverse fonti. I segnali grezzi inviati dai sensori vengono trasformati e arricchiti con ulteriori informazioni semantiche. Viene utilizzato un KG come componente centrale per l’aggregazione e organizzazione delle informazioni, catturando informazioni tra le entità presenti nell’ ambiente e le loro relazioni.

Formuliamo un KG come un insieme di triple G= H, R, T. Dove H è un insieme di entità (es. 'auto', 'strada', 'conducente'), T insieme di coppie di entità (entità o valori letterali), e R è l'insieme di relazioni che collegano le entità tramite vincoli relazionali. Una volta realizzato il processo di trasformazione, cioè la conversione dei dati di input da qualsiasi formato a triple, l'output viene memorizzato in un grafo di conoscenza. Le informazioni nel KG vengono aggregate e organizzate in modo intuitivo e gerarchico, rendendole facili da sfruttare e capire per gli esseri umani. Nel caso di CoSI le informazioni della scena vengono rappresentate nel KG (ora CKG) tramite istanze di concetti Ontologici, mirando a catturare le informazioni maggiormente rilevanti provenienti

dai sensori montati in un veicolo specifico. Codificando assiomi formali aggiuntivi il sistema ora è in grado di dedurre nuovi fatti da quelli dati tramite tecniche di ragionamento automatizzato. Per approfondire nell'articolo si può vedere come è strutturata l'ontologia utilizzata per CoSI.

Trasformazione ed Arricchimento. Il componente di Trasformazione ed Arricchimento converte i dati dei sensori in una rappresentazione semantica (RDF) utilizzando sia approcci dichiarativi che imperativi. Le mappe predefinite collegano i dati dei sensori ai concetti ontologici, mentre in situazioni complesse vengono eseguite query in tempo reale per arricchire i dati con nuove informazioni semantiche, migliorando la comprensione delle situazioni.

Estrazione della Conoscenza. Componente che consente di eseguire compiti successivi, come completare il grafo di conoscenza, prevedere collegamenti o classificare, utilizzando la conoscenza da diverse prospettive. Questo componente permette di eseguire query complesse e attraversare il grafo per recuperare informazioni pertinenti. Di conseguenza, le tecniche di incorporamento che operano a livello di grafo possono apprendere efficientemente la rappresentazione vettoriale della conoscenza simbolica dalla "nuova prospettiva".

Comprensione della situazione. Diversi componenti specializzati sono responsabili della comprensione della situazione, includendo la Classificazione della Situazione, la Valutazione della Difficoltà e la Previsione della Traiettoria. Questi componenti possono seguire due paradigmi principali: uno basato su tecniche che si avvalgono di regole dichiarative per classificare le situazioni in base alla logica, sfruttano la potenza espressiva della struttura del grafo di conoscenza che, in combinazione con tecniche di ragionamento, fornisce risultati interpretabili. L'altro metodo è basato sull'apprendimento che utilizza tecniche per apprendere modelli da un vasto numero di dati osservati, permettendo di effettuare previsioni o classificazioni senza la necessità di regole esplicite predefinite.

Ora, tramite il Dataset descritto a inizio paragrafo viene generato un grafo di conoscenza CoSI (CKG) con 915 milioni di triple. Queste istanze contengono informazioni sui punti di conflitto, la velocità del veicolo ego e del veicolo avversario e la direzione del movimento, rispettivamente. Viene fatta una *valutazione* sull'accuratezza dell'approccio basato su KG in un compito di classificazione della situazione, paragonando questo tipo di approccio rispetto a modelli basati su vettori.

In particolare, nel modello basato su KG viene utilizzata una rete convoluzionale grafica relazionale (R-GCN), un'estensione della rete convoluzionale grafica (GCN), per operare direttamente su un grafo e apprenderne le rappresentazioni. Estendendo la matrice delle rappresentazioni dei nodi H con le rappresentazioni delle caratteristiche per ciascun nodo formiamo una rete convoluzionale grafica relazionale multimodale (MRGCN). **MRGCN** è implementato con un 'hidden layer' con 40 nodiFootnote6, addestrato in modalità full batch con un ottimizzatore ADAM.
Questo modello viene confrontato con i seguenti modelli, a seguito dell'estrazione delle caratteristiche dal nostro KG per rappresentare le caratteristiche basate su vettori:

- **'Support Vector Machine (SVM)'**. Insieme di metodi di apprendimento supervisionato che utilizzano un sottoinsieme dei campioni di addestramento, chiamati vettori di supporto, nella funzione decisionale. È implementato con kernel a funzione di base radiale e DTC utilizza l'algoritmo 'Classification And Regression Trees'.
- **'Decision-Tree Classifier (DTC)'**. Applicazione ad albero decisionale adatta al nostro compito di classificazione

- ‘**Multi-Layer Perceptron (MLP**)’. Viene utilizzata una rete neurale a percettroni multipli addestrata tramite con ‘back-propagation’ per il compito di classificazione. Strutturato da 2 strati nascosti con 40 neuroni nascosti per strato, addestrati in modalità full batch con un ottimizzatore ADAM.

I *risultati complessivamente migliori sono ottenuti dal metodo MRGCN utilizzando le 5 caratteristiche più importanti. Si possono fare le seguenti osservazioni in base ai risultati:*

	MRGCN	SVM	DTC	MLP
Single features				
Vehicle signal	0.468	0.510	**0.511**	**0.511**
Longitudinal lane position	0.501	0.566	0.507	**0.569**
Steering angle	0.607	0.524	**0.641**	0.510
Distance between vehicles	**0.673**	0.529	0.491	0.542
Lane ID	0.883	0.520	**0.890**	0.514
Combined features				
ALL 36 features	**0.938**	0.708	0.750	0.733
5 most important features	**0.953**	0.648	0.746	0.742

Fig. 14 – Risultati MRGCN vs SVM, DTC, MLP.

- Gli esperimenti mostrano che il classificatore basato su KG ottiene risultati notevolmente migliori rispetto a tutti i classificatori basati su vettori, nel caso in cui vengano utilizzate le 5 caratteristiche più importanti o tutte le caratteristiche, rispettivamente. *Ciò suggerisce che la rappresentazione basata su grafo fornisce informazioni più discriminanti rispetto alla rappresentazione classica basata su vettore di caratteristiche.*
- Gli esperimenti di classificazione con singole caratteristiche mostrano che i metodi basati su vettori sono superiori al classificatore basato su KG per tutte le caratteristiche tranne che per la "Distanza tra veicoli". *Ciò suggerisce che i metodi basati su KG non hanno un vantaggio evidente nelle situazioni semplici, ma superano i metodi basati su vettori nelle situazioni complesse in cui esistono molte relazioni e interazioni tra i partecipanti.*

Quindi riassumendo abbiamo presentato CoSi e abbiamo visto come l’ontologia CoSI posta come scheletro di un KG fornisca una rappresentazione semantica dei concetti principali in un contesto di guida. Procedendo con l’integrazione di dati da fonti multimodali eterogenee in un KG, comunque, si è dimostrato la possibilità dell’applicabilità di tale approccio. Ritengo importante osservare come attraverso il CKG sia possibile rappresentare le informazioni più complesse del dominio di guida, e quando ciò è combinato a reti neurali basate su grafi, porta a prestazioni superiori ottenendo fino al 95 % di precisione. Inoltre, ritengo che, come visto per il lavoro precedente, (KEP) sia possibile sfruttare la struttura composta da relazioni semantiche tra entità per avere un processo decisionale più trasparente rispetto agli altri modelli basati su vettori e strati nascosti di reti neurali. Ritengo anche che attraverso la integrazione di conoscenza esterna all’ interno dei grafi di conoscenza si possa avere un vantaggio nell’ integrare concetti come di ragionamento comune, sociali, etici, morali e legali. Ritengo che tutto questo in relazione con l’ontologia relazione CoSi e lo sfruttamento dell’estensione delle reti neurali convoluzionali possa essere un ottimo punto di partenza per incrementare l’affidabilità e il funzionamento di tali sistemi, aumentando dall’ altro lato la fiducia nell’ utilizzo da parte dell’utente finale.

3.3 Sistema 3: ‘ML-MAS: a Hybrid AI Framework for Self-Driving Vehicles’

Prendendo in considerazione i problemi precedentemente affrontati nei primi due lavori, analizziamo ora una soluzione potenziale di Intelligenza Artificiale ibrida presentata nel seguente studio. Il framework proposto è coerente con le esigenze precedentemente identificate, relative alle sfide connesse all'utilizzo esclusivo del componente sub-simbolico (neurale). Pertanto, questo modello è stato concepito con l'obiettivo di migliorare il comportamento autonomo del veicolo attraverso lo sfruttamento della componente logica simbolica in forma

di agente razionale BDI, questa parte sarà attivata quando si verificheranno problematiche o situazioni di difficile risoluzione mediante le tecniche di Machine Learning convenzionali (Shukairi e Cardoso 2023).

Nel lavoro viene esposto come secondo lo psicologo Daniel Kahneman le decisioni umane sono prese attraverso una cooperazione tra "Sistema 1: Pensiero veloce" e "Sistema 2: Pensiero lento". Il Sistema 1 viene utilizzato per decisioni intuitive, approssimative, rapide e inconsce. In contrasto, il Sistema 2 viene utilizzato in situazioni più complesse che richiedono un pensiero logico e razionale. Kahneman stima, inoltre, che circa il 95% del nostro pensiero si basi sul Sistema 1 per prendere decisioni. Applicando questo concetto all'AI ibrida, il Sistema 1 rappresenta il componente ML (decisioni inconsce apprese dall'esperienza), mentre il Sistema 2 rappresenta il componente AI simbolica (decisioni razionali). Traducendo tale riflessione in un contesto stradale, il sistema 1 prende parte delle decisioni che tipicamente dobbiamo prendere in situazioni di guida, magari tornando a casa e attraversando le solite strade, quartieri e situazioni. Nel momento in cui decidiamo di prendere una nuova strada o ci vengono poste delle situazioni non note alla nostra esperienza o quotidianità, è lì che vengono prese decisioni secondo la seconda tipologia di sistema. Nel nostro caso la componente logica in forma di agente razionale BDI sarà attivata quando si verificheranno problematiche o situazioni di difficile risoluzione mediante le tecniche di Machine Learning convenzionali. Il Sistema 1 è un modello ML pre-addestrato (LAV e TransFuser), mentre il Sistema 2 è un agente razionale basato sul modello Belief-Desire-Intention (BDI) programmato nel sistema multi-agente Jason.

Vengono di seguito presentati i componenti del sistema: BDI BRIDGE, Symbolic AI (BDI) MAS agent, Orchestrator, Pre-Trained ML-Model.

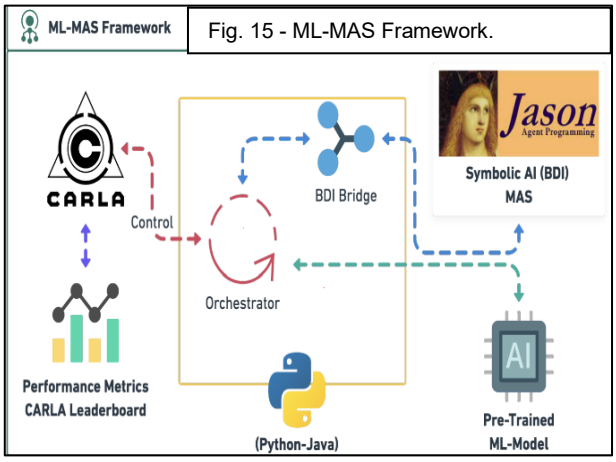

Fig. 15 - ML-MAS Framework.

BDI bridge. Il ponte BDI costituisce un meccanismo di comunicazione fluida tra l'agente BDI e l'API di CARLA. L'architettura di questo ponte, si basa sulla comunicazione tramite socket ed è composta da un client (implementato in Java-Jason) e un server (in Python). Per ottimizzare l'efficienza, sia il client che il server operano su tre 'thread' principali: uno per l'invio dei messaggi, uno per la ricezione dei messaggi e uno per la gestione dei messaggi. I messaggi sono definiti nel noto formato JSON, e sia il client che il server utilizzano un interprete e un parser JSON per identificare il tipo di messaggio e determinare come trattare ciascun messaggio.

Orchestratore. Questo componente svolge il ruolo di coordinatore delle decisioni tra i due sistemi. Le sue funzioni includono il caricamento del modello ML, della sua configurazione e dei pesi pre-addestrati, oltre a instradare i dati dai sensori richiesti dal modello. Garantisce altresì la disponibilità dei dati dei sensori necessari per l'agente BDI, eseguendo la preelaborazione dei dati e inviando solo ciò di cui l'agente ha bisogno.

La comunicazione con l'agente BDI avviene attraverso il ponte BDI, e l'orchestratore attende una risposta sia dal modello ML che dall'agente BDI. Per ottimizzare il processo, se in un determinato frame di simulazione non sono necessari dati per l'agente, si basa unicamente sull'azione del modello ML.

Algorithm 1: Main function to run a scenario.

```
1  Function main()
2      Route_list ← getRoutesAndScenarios()
3      FPS ← 20
4      foreach Route ∈ Route_list do
5          Sensors_list ← getMLSensors() · getBDISensors()
6          loadWorldMap(Route, Sensors_list)
7          ML_Model ← prepareMLWeights()
8          BDI_Agent ← connectBDIBridge()
9          Game_time ← 0
10         System_start_time ← getCurrentTimeInSec()
11         while Route_status ≠ finished do
12             Game_time ← Game_time + 1
13             for k ← 1 to FPS do
14                 Sensors_data ← collectData(Sensors_list)
15                 Control ← runStep(Sensors_data)
16                 Metrics ← Metrics · eval(Route, Control)
17                 Route_status ← status(Route, Metrics)
18                 if Route_status = finished then
19                     break
20         System_time ←
             getCurrentTimeInSec() − System_start_time
21         store(Metrics, Game_time, System_time)
```

Algorithm 2: Orchestrator function to run a step.

```
1  Function runStep (Sensors_data)
2      Last_control ← getLastControl()
3      Repeat_counter ← getRepeatCounter()
4      Frame_number ← Frame_number + 1
5      if Repeat_counter ≥ 1 then
6          Repeat_counter ← Repeat_counter − 1
7          return Last_control        // Previous BDI control
8      ML_control ← getMLControl(Sensors_data)
9      Preprocessed_data ← preprocess(Sensors_data)
10     if Preprocessed_data ∉ BDI_triggers() then
11         Repeat_counter ← 0
12         return ML_control
13     Last_control, Repeat_counter ←
         getBDIControl(Preprocessed_data, ML_control)
14     if Last_control = noaction then
15         Repeat_counter ← 0
16         return ML_control
17     addBDIMetrics(Last_control, Repeat_counter)
18     return Last_control                // BDI control
```

Fig. 16 – Algoritmi componente Orchestratore.

Per comprendere al meglio il ruolo dell'orchestratore ritengo utile visionare i seguenti algoritmi:

L'Algoritmo 1 fornisce un'anteprima della funzione principale che gestisce il caricamento delle rotte e degli scenari configurati. Itera attraverso ogni percorso, assicurandosi che sia preparato e configurato correttamente l'ambiente di simulazione (comprensivo del caricamento della mappa urbana, posizionamento degli attori, condizioni meteorologiche, ecc.) e i sistemi necessari (ad esempio, caricamento dei pesi nel modello ML, connessione con l'agente BDI tramite il ponte BDI, ecc.), come indicato nelle righe 5-10. Per ciascun frame temporale di gioco (for k← 1 to FPS), l'orchestratore raccoglie dati dai sensori (do), provenienti dall'ambiente simulato, e invoca la funzione **runStep**. Quest'ultima si interfaccia con il modello ML e l'agente BDI, restituendo un messaggio di controllo da uno dei due. Questo messaggio viene quindi trasmesso all'ambiente per l'esecuzione attraverso la funzione eval().

L' algoritmo due descrive la funzione 'RunStep' richiamata a sua volta dall' algoritmo 1.

Le righe 5-7 verificano la presenza di azioni da eseguire in modo ripetuto. Questo meccanismo consente all'orchestratore di eseguire la stessa azione per un numero specificato di frame senza dover richiedere l'intervento dell'agente BDI o del modello ML fino a quando la sequenza di ripetizioni non è terminata. Questa funzionalità può risultare utile in situazioni di guida, ad esempio quando il sistema deve eseguire frenate multiple in rapida successione in risposta a specifici scenari stradali.

Nel caso in cui non vi siano azioni BDI da ripetere, l'orchestratore inoltra i dati dei sensori al modello ML. Se i dati non innescano alcun piano BDI, l'orchestratore si basa sul controllo fornito dal modello ML per l'esecuzione delle azioni.

L' orchestratore rappresenta parte centrale del processo e, sebbene non sia un esempio perfetto di sistema neuro-simbolico come definita nel capitolo 2.4, è comunque un modello interessante che merita ulteriori analisi. Attraverso la creazione di questo framework e le attività svolte dall'orchestratore e dal ponte BDI, emergono

opportunità di comunicazione tra sistemi simbolici e sub-simbolici. Sebbene i sistemi rimangano separati nel processo decisionale a livello globale, siamo in grado di sviluppare un sistema con la capacità di integrare le conoscenze simboliche per affrontare le sfide emerse nei modelli di Machine Learning nell'ambito dell'auto a guida autonoma. Questo approccio rappresenta un passo significativo verso l'ottimizzazione delle prestazioni e della sicurezza nei contesti di guida autonomi.

Agente BDI. *Nel* contesto del lavoro, sono state individuate sette situazioni critiche che richiedono l'intervento del Sistema 2. Queste situazioni si basano su infrazioni e collisioni osservate nei modelli di Machine Learning più avanzati durante le simulazioni sul Leaderboard. Per affrontare ciascuna di queste situazioni, è stato implementato un piano nell'agente BDI Jason, il cui obiettivo principale è prevenire collisioni e gestire il traffico in modo efficiente, consentendo al veicolo di navigare in modo sicuro e autonomo. Nel lavoro vengono riportati 7 metodi: *e*vitare collisioni in incroci ravvicinati, evitare collisioni in lontananza all'incrocio, evitare collisioni frontali, evitare collisioni posteriori, uscita dall'incrocio, navigazione in caso di ingorgo stradale.

Ad esempio il metodo per evitare collisioni in incroci ravvicinati *(fig.17):* questo piano viene attivato quando l'agente rileva un ostacolo vicino che rappresenta un rischio di collisione, ad esempio dove un ciclista sta per attraversare la strada

```
+! frame(F): f (F, X, _, _, MinY) & MinY < 2.0 & X < 4.5 &
 info (F, Sp) & Sp > 0.5
<- control (2, 0.0, 0.0, 1.0, false, false, (Sp*3)).
```

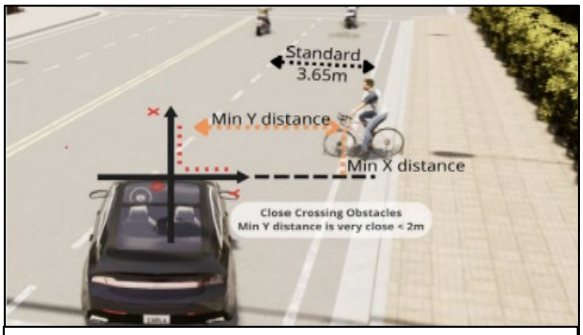

Fig. 17 – Agente BDI per evitare collisioni.

+! frame(F) è un evento che si attiva quando vengono soddisfatte determinate condizioni
Le condizioni per attivare questo piano sono definite nel contesto (tra: e <-), tra cui vincoli sulla distanza minima in Y (MinY), distanza massima in X (X) e velocità minima (Sp).
Le azioni da intraprendere quando il piano viene attivato e il contesto è valido sono specificate dopo <-. *In questo caso, ordina al veicolo di applicare un "freno deciso" (control (2, 0.0, 0.0, 1.0, false, false, (Sp*3))).* Viene eseguita un'azione di "*hard break*" (frenata intensa) *per una durata di 2 secondi*. La velocità del veicolo è influenzata dalla velocità corrente, moltiplicata per 3 (Sp3), il che contribuisce a calcolare quanto frenare.

Questa serie di operazioni possono essere utili nell' affrontare scenari più complessi che quotidianamente si verificano in contesti stradali. *Sfruttando gli agenti BDI e il ponte BDI attraverso la coesione ramificata dall' orchestratore è possibile aumentare la comprensione e la corrispettiva correttezza nel reagire a tali situazioni.* Attraverso gli agenti BDI siamo in grado di fornire al framework la capacità di raggiungere obiettivi complessi in modo razionale descrivendo piani per conseguirli (ad esempio, evitare collisioni frontali, posteriori e incrociate) utilizzando "stati mentali" come credenze, desideri e intenzioni.

La 'Leaderboard' di CARLA fornisce metriche progettate per misurare diversi aspetti della guida. La metrica finale del punteggio di guida è una moltiplicazione tra due metriche di aggregazione: completamento del percorso e penalità per infrazione. Ci sono cinque sotto-metriche che influenzano negativamente il

completamento del percorso: guida fuori strada, deviazione dal percorso, blocco dell'agente (il veicolo è bloccato per un certo periodo di tempo), time-out della simulazione e time-out del percorso. Applicando il framework ML – MAS al modello LAV di ML, attraverso diverse metriche di valutazione fornite da CARLA e diverse sotto-metriche aggiunte nella valutazione del modello, si manifestano i seguenti risultati:

i. ML-MAS ha ridotto le collisioni per chilometro di oltre la metà. La riduzione più significativa è stata nelle collisioni con i veicoli, che sono diminuite da 0,247/km a soli 0,044/km. Il numero totale di collisioni per km per il modello LAV era di 0,394/km rispetto a solo 0,12/km per ML-MAS.

ii. Mostra un lieve miglioramento in due infrazioni, il semaforo rosso e il segnale di stop. ML-MAS ha commesso più infrazioni fuori strada rispetto al modello LAV, 0,01/km in più per la precisione. Ciò accade perché il modello BDI considera azioni più complesse come retromarcia e inversione opposta a un ostacolo, mentre i modelli ML spesso utilizzano solo azioni di accelerazione, sterzata e frenata. Eseguendo azioni più complesse, l'agente BDI potrebbe scegliere di commettere una piccola infrazione per evitare una più grave.

Nel lavoro viene evidenziato come l'interferenza totale dell'agente razionale nei percorsi valutati è stata dell'8,1%, con il restante 91,9% delle azioni inviate dal modello ML. Ciò conferma l'aspettativa che il Sistema 1 (LAV) dovrebbe agire per la maggior parte del tempo, mentre il Sistema 2 (agente BDI) dovrebbe interferire solo in una piccola percentuale del tempo per fornire decisioni razionali, andando a confermare la teoria di Kahneman.

Per testare la generalità del framework, vengono condotti anche esperimenti con un secondo modello di apprendimento automatico, il modello TransFuser. ML-MAS è riuscito a migliorare il punteggio di guida del 3,6%, passando dal 44,8% (‘TransFuser’) al 48,4% (ML-MAS). Anche il completamento del percorso è stato nuovamente migliorato notevolmente, passando dall'85,9% per ‘TransFuser’ al 98,2% per ML-MAS. Sfortunatamente, ML-MAS ha commesso più infrazioni, ottenendo un punteggio di penalità per infrazione peggiore, 48,9% rispetto al 51,0% di TransFuser.

In conclusione, è evidente come l'integrazione o, più semplicemente, l'interazione tra questi due tipi di sistemi possa rappresentare una potente risorsa per affrontare varie problematiche riscontrate nei contesti stradali. Nel caso in esame, l'aggiunta di sette piani nell'agente BDI, attraverso i quali può intervenire e assumere il controllo in situazioni specifiche, si rivela un ottimo punto di partenza per sviluppare sistemi ibridi efficaci. Avvicinandosi ad una prospettiva neuro-simbolica, è possibile individuare nei risultati dell'articolo una risposta al problema dell'integrazione discusso nel capitolo 2.4 della tesi, con richiamo al lavoro precedentemente analizzato; infatti, come si può leggere nel lavoro analizzato:

> *Da un punto di vista più teorico, desideriamo esplorare la comunicazione tra il modello di ML e l'agente BDI, in modo che il processo di selezione di un'azione sia il risultato di una deliberazione diretta tra entrambi i componenti.*

Questo romperebbe il limite di avere due sistemi separati che attraverso ponti BDI, orchestratori e comunicazioni Client-Server. Invece ci si pone come obiettivo quello di far comunicare direttamente i sistemi per entrambi al fine di deliberare una decisione finale.

Inoltre, ritengo che l'interpretazione e la spiegabilità delle azioni eseguite dal Sistema 2, rappresentato dall'agente BDI razionale, risultino più agevoli nei sette scenari precedentemente elencati. Sono dell'opinione che attraverso la creazione di ulteriori scenari e con la speranza di sviluppare un sistema in cui le azioni siano il risultato di una deliberazione congiunta dei due componenti, si possa conseguire un sistema di supporto ai sistemi di Machine Learning per la guida autonoma più efficiente e in grado di gestire in maniera più efficace le tipiche situazioni degli "ambienti open world". Tramite l'impiego di agenti relazionali Jason, il sistema sarà in grado di tenere traccia dei processi decisionali sottostanti alle situazioni presentate al sistema, aumentando così il livello di trasparenza nel processo decisionale. Infine, portando alla luce anche i lati negativi di questo sistema, ritengo che prima che tali modelli possano essere implementati in sistemi reali e non ambienti simulati debba essere sicuro che le decisioni prese non vadano a recare ulteriori rischi nell' ambente circostante. Forse potrebbe essere utile a questo scopo un sistema di DRL come vedremo nel sistema 5 che vada direttamente a filtrare tra le possibili azioni che può prendere un sistema prendendo in considerazione solo le azioni ritenute sicure.

3.4 Sistema 4: 'Exploiting T-norms for Deep Learning in Autonomous Driving'

In questo articolo viene presentato un nuovo approccio per l'implementazione di una perdita basata su t-norma in reti neurali, con l'obiettivo di incorporare vincoli logici di conoscenza di base nei modelli di deep learning. Gli autori hanno affrontato il problema della gestione efficiente della memoria quando si utilizzano t-norme in scenari di elaborazione intensiva, come la rilevazione di eventi per la guida autonoma (Stoian, Giunchiglia, e Lukasiewicz 2023). A partire dalla premessa proposta nell' articolo ritengo sia interessante analizzare il contesto di applicazione di tale studio spiegando le operazioni coinvolte in un problema di rilevamento degli eventi.

Un problema di rilevamento degli eventi $\mathcal{P}$ è una coppia $(\mathcal{A}, \mathcal{X})$, dove $\mathcal{A}$ è un insieme finito di etichette e $\mathcal{X}$ è un insieme di coppie $(X, \mathcal{Y})$ in cui:

- $X \in \mathbb{R}3\times W \times H$ è il tensore associato a ciascun frame nel video.
- Wrappresenta la larghezza e H l'altezza di ciascun frame, mentre 3 è il numero di canali utilizzati nella codifica RGB.
- $\mathcal{Y}$ è il ground truth di X e comprende un insieme di coppie $(b, \mathcal{L})$, in cui:
  $b \in \mathbb{R}4$ rappresenta le coordinate di un 'bounding box', ovvero un rettangolo che segna la posizione di un agente nel frame.$\mathcal{L}$ rappresenta l'insieme di etichette associate a b.

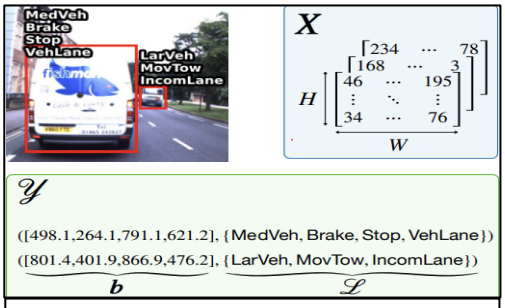

Fig. 18 - Problema di rilevamento.

In questa applicazione, il compito principale è identificare e categorizzare gli eventi o le azioni che si verificano nell'ambiente circostante mentre un veicolo autonomo si muove su strada. Questi eventi possono includere situazioni come il cambio di corsia, il superamento di un semaforo rosso o il riconoscimento di pedoni attraversanti. È fondamentale che il sistema sia in grado di comprendere e rispondere in modo adeguato agli eventi rilevati. Pertanto, è necessario attribuire a ciascuna situazione rilevata un insieme di etichette che descrivono l'evento e le condizioni circostanti. Queste etichette possono includere informazioni sull'azione in corso (ad esempio, "cambio di corsia"), sulla posizione degli oggetti (ad esempio, "pedoni sulla strada") e

sull'entità coinvolta (ad esempio, "veicolo"). L'articolo si concentra sulla creazione di modelli di deep learning in grado di eseguire questa rilevazione di eventi in modo efficiente ed efficace. In particolare, *gli autori cercano di incorporare vincoli logici di conoscenza di base all'interno di questi modelli, al fine di migliorare la comprensione e la categorizzazione degli eventi.* Questi vincoli logici possono includere relazioni tra le etichette, ad esempio "Se è presente l'etichetta 'semaforo rosso', allora l'etichetta 'fermo' deve essere presente".

Nel capitolo successivo viene mostrato un modello classico di previsione di queste tipologie di eventi e inoltre viene proposto un nuovo modello che lavora invece su vincoli logici proposizionali, in cui viene affrontato il problema di rilevamento degli eventi con vincoli logici proposizionali ($\mathcal{P}$, Π), il quale consiste in un problema di rilevamento degli eventi $\mathcal{P}$ e in un insieme finito di vincoli Π, espressi sull'insieme $\mathcal{A}$ di etichette in $\mathcal{P}$. Viene poi spiegato come implementare le funzioni di perdita basate su t-norma, prima nel modo standard e poi in modo efficiente in termini di memoria utilizzando tensori sparsi.

> Viene riportato un esempio concreto su come lavora questo sistema dato un problema di rilevamento degli eventi con vincoli logici proposizionali ($\mathcal{P}$, Π). Attraverso tale sistema, date P e C (matrice delle previsioni, matrici dei vincoli positivi e negativi), l'obiettivo è calcolare il grado di soddisfazione di ciascun vincolo per ciascuna previsione, il che può essere espressa in modo compatto come una matrice G di dimensioni D × |Π|. Attraverso le t-norme viene calcolato il grado di soddisfazione delle restrizioni logiche (G). La funzione di regolarizzazione logica basata su t-norme calcola quanto il modello rispetti queste restrizioni logiche durante l'addestramento. Viene poi messo in evidenza come questo sistema sia inadeguato in quanto richiede di lavorare con matrici dense tridimensionali, il che comporta un grande sovraccarico a livello computazionale e rende il metodo impraticabile, specialmente per domini applicativi come la guida autonoma.

Quindi viene proposto il secondo approccio risolutivo tramite rappresentazione a matrice sparsa:

> Questa soluzione si basa principalmente sull'intuizione che nella pratica la maggior parte dei vincoli sono scritti su un sottoinsieme delle etichette disponibili in $\mathcal{A}$, e che questo sottoinsieme è di solito molto più piccolo di $\mathcal{A}$. Come si può osservare nell' analisi sperimentale, infatti si può osservare come pur essendoci 41 etichette disponibili in ROAD-R, il vincolo più lungo è scritto su un sottoinsieme molto più piccolo. Solo 15 etichette. Di conseguenza, $C+$ e $C-$ () contengono principalmente zeri. Viene quindi ideato un metodo per catturare la perdita basata sulla logica che sfrutta questa proprietà di sparsità ed evita in definitiva i costi computazionali elevati indotti dalle matrici 3D, operando solo su matrici 2D.

Quindi hanno testato la perdita basata su t-norm nel compito del rilevamento degli eventi per la guida autonoma, dove l'obiettivo è assegnare a ciascun 'bounding box' ('scatola' che contiene percezioni visive mappate) rilevato in ciascun video un sottoinsieme di etichette, inclusa un'etichetta di agente e un sottoinsieme di etichette di azione e posizione. Viene utilizzato il dataset recentemente introdotto per la guida autonoma, ROAD-R, estensione di ROAD, con 243 vincoli annotati manualmente. Viene utilizzato il detector *3D-RetinaNet* con una struttura di base *ResNet50* combinata con un Random *Connectivity Gated Recurrent Unit* (RCGRU) per l'apprendimento delle caratteristiche temporali, nel lavoro officiale vengono riportate le fonti in cui si possono trovare informazioni aggiuntive.

Viene eseguito un test del metodo con tre diverse perdite basate su t-norm (cioè Gödel, Łukasiewicz e Product) utilizzando il 10%, il 20%, il 50%, il 75% e il 100% dei dati annotati disponibili in ROAD-R, addestrando per rispettivamente 110, 70, 45, 30 e 30 epoche. Vengono poi calcolato le perdite basate su t-norm rispetto a tutti i 243 vincoli di ROAD-R. Come risultato si ottiene che l'integrazione della conoscenza di base (tramite la perdita basata su t-norm) nei modelli neurali aiuta di più quando ci sono pochi dati disponibili. Infatti, i modelli presentati hanno ottenuto un miglioramento fino al 1,85% e al 3,95% utilizzando rispettivamente il 10% e il 20% dei dati di addestramento etichettati.

Ho voluto riportare questo metodo per portare alla luce un altro di tipo di interazione tra modelli neurali e simbolici (tipo 5), inoltre si può osservare come a partire dal dataset ROAD-R, testando la perdita basata su t-norma su diverse quantità di dati etichettati, viene dimostrato che le t-norme contribuiscono effettivamente a migliorare le prestazioni dei modelli all'avanguardia in campo di auto a guida autonoma.

3.5 Sistema 5: 'DRLSL - Towards safe autonomous driving policies using a neuro-symbolic Deep Reinforcement Learning approach'

In questo articolo viene presentato un nuovo approccio chiamato DRL con Logiche Simboliche (DRLSL) per affrontare le sfide della guida autonoma in ambienti reali. Questo approccio combina l'apprendimento profondo per rinforzo (Deep Reinforcement Learning - DRL) con le logiche simboliche del primo ordine (FOLs) per migliorare la sicurezza dei sistemi di DRL. Il DRL richiede che il modello "sbagli" per apprendere dagli errori, il che può essere pericoloso in scenari di guida reali. *Per affrontare questo problema, il DRLSL introduce il ragionamento basato sulla conoscenza umana utilizzando regole logiche simboliche per guidare il processo di addestramento del DRL*. Questo permette al DRL di apprendere in modo sicuro e di rispettare le regole di guida fin dall'inizio, riducendo il rischio di azioni non sicure.

Sfruttando le logiche simboliche rappresentiamo la conoscenza di base umana sull'ambiente stradale e suggeriamo azioni sicure in ciascuno stato. Attraverso questo processo vengono respinte direttamente le azioni non sicure e il DRL è costretto a compiere solo azioni sicure tra quelle suggerite dalle logiche simboliche. Questo approccio permette al DRL di apprendere comportamenti sicuri in modo rapido e di conformarsi alle regole stradali definite. Inoltre, l'uso delle logiche simboliche fornisce una rappresentazione trasparente e interpretabile della conoscenza, aiutando a comprendere e convalidare le azioni dell'agente. Vengono poi date delle nozioni base sul funzionamento generale dei momenti di DRL, approfondendo il funzionamento di una specifica rete neurale (DQN) come algoritmo di apprendimento con rinforzo e infine viene introdotto il funzionamento della logica di primo ordine attraverso il quale è possibile rappresentare conoscenza attraverso fatti e regole con lo scopo di trarre inferenze logiche (Sharifi, Yildirim, e Fallah 2023). Per capire come funziona il sistema ritengo importante fare una breve introduzione con gli aspetti fondamentali dei punti presi in considerazione.

DRL. In generale, un problema di apprendimento profondo per rinforzo (DRL) può essere descritto come un processo decisionale di Markov (MDP) che comprende uno spazio di stati (S), un insieme di azioni (A), una funzione di ricompensa (R), transizioni di stato (P), e un fattore di sconto (γ). In ogni passo temporale, l'agente osserva lo stato attuale (s), sceglie un'azione (a) dall'insieme delle azioni disponibili (A), e riceve una ricompensa (R) in base a questa azione e alle transizioni di stato successive (s'). L'obiettivo dell'agente è apprendere una politica (π) che mappa gli stati (s) alle azioni (a) per massimizzare la ricompensa cumulativa attesa nel tempo. Viene inoltre presentata la rete neurale profonda per il Q-learning (DQN).

FOLs**.** Si tratta di un formalismo utilizzato per la rappresentazione della conoscenza e il ragionamento (Knowledge Representation and Reasoning). In FOL, si utilizzano fatti e regole per rappresentare la conoscenza e derivare inferenze logiche. Una regola è composta da una testa (head) e un corpo (body) ed è definita nel formato seguente: head: - body.

- La parte "head" della regola rappresenta un predicato di output che esprime una relazione tra oggetti o concetti.
- La parte "body" specifica le condizioni sotto le quali il predicato nella testa è vero.

Ogni predicato è costituito da un funtore e da argomenti, scritti come functor(arg1, arg2, ..., argn), in cui gli argomenti possono essere costanti o variabili. Inoltre, le regole in FOL sono spesso formulate come clausole di Horn, che sono implicazioni logiche con una sola proposizione positiva nella testa e zero o più proposizioni nel corpo. Le clausole di Horn seguono il formato:

H: - B1, B2, ..., Bn.

Dove con i che varia da 1 a n, rappresentano le premesse nel corpo della regola, mentre H, la testa della regola, rappresenta una conclusione. Questa regola implica che se la congiunzione di B1 a Bn è vera, allora anche H è vera; in caso contrario, H è falso.

Metodo proposto DRLSL. A questo punto viene introdotto un programma logico simbolico (SLP), in pratica tramite questo sistema abbiamo la possibilità, utilizzando le FOLs, di manipolare le espressioni simboliche create per rappresentare e ragionare sulla conoscenza e sulle relazioni logiche. In particolare, tramite l'SLP vengono creati vincoli logici al fine eliminare uno spazio di azioni non sicure, proponendo al sistema solo lo spazio Atsafe contenente le azioni sicure filtrate. Una volta che l'SLP determina At(safe) ad ogni passo temporale, l'agente DRL utilizza il metodo ϵ-greedy per selezionare un'azione sicura at(safe) da At(safe). Nell' algoritmo 1 del lavoro possiamo osservare lo pseudo-codice per la rete neurale Q con logica simbolica (DQNSL) per garantire all' agente di selezionare solo azioni dall' insieme At(safe).

Auto a guida autonoma e SLP. In un sistema AD un SLP può essere utilizzato per garantire la sicurezza codificando regole basate su FOL e vincoli relativi al comportamento di guida. Ad esempio, il sistema può essere progettato per seguire le regole del traffico, mantenere una distanza di sicurezza o evitare collisioni. Inoltre, un SLP può essere integrato con le tecniche DRL incorporando vincoli logici di guida nel processo di ottimizzazione della politica dell'agente DRL. Incorporando regole di sicurezza come vincoli, il sistema può essere addestrato a ottimizzare il suo comportamento pur rispettando comunque le regole. Ciò viene fatto seguendo i passaggi:

- Viene usato 'Prolog' linguaggio di programmazione logico simbolico per definire le regole di sicurezza desiderate in un ambiente autostrada (sezione 2.2.2).
- Successivamente, viene utilizzata la conoscenza umana (BK) per proteggere le azioni non sicure da A. In questo caso rappresentiamo regole come clausole sulla situazione spaziale, temporale dei Vecoli intorno all' AD, per fare comprendere al veicolo quando potersi ad esempio spostare di corsia in sicurezza.
- Infine, attraverso FOLs un insieme di regole logiche viene applicato per determinare At(safe).
    - Ad esempio, la regola "se non c'è alcun veicolo nelle sezioni a sinistra, mantieni la tua corsia con una velocità sicura o vai nella corsia a sinistra" o "se le sezioni a destra e a sinistra sono

occupate, mantieni la corsia" possono essere utilizzate per estrarre At(safe) per i cambi di corsia.

DRLSL for AD. In DRLSL, dopo avere determinato l'insieme di azioni sicure (At(safe)) vengono utilizzate le regole basate su FOL nel nucleo SLP, viene inoltre integrato questo insieme all'agente DRL utilizzando la libreria PySwip, che funge da collegamento tra Prolog e Python. Quindi, l'insieme At(safe) viene passato come azioni disponibili alla rete DQN per lo stato corrente. L'agente DQN, quando utilizza la logica simbolica (DQNSL), seleziona l'azione migliore da At(safe) utilizzando il metodo ϵ-greedy, come mostrato nell'Algoritmo 1.

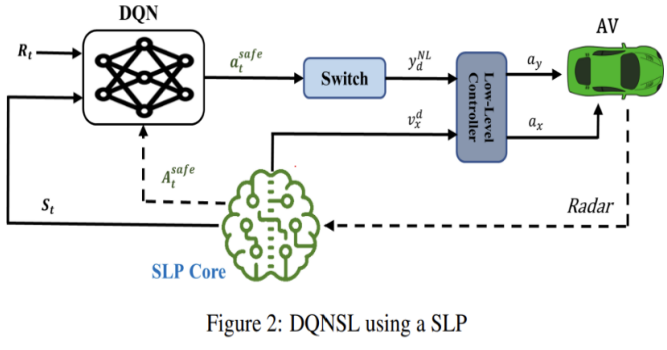

Figure 2: DQNSL using a SLP

Fig. 19 – DQNSL tramite SLP.

Di fondamentale importanza è osservare come questa limitazione dell'insieme delle azioni ad At(safe) assicura che la rete DQN apprenda solo azioni sicure e rispetti le regole di sicurezza. L'integrazione dell'SLP nel DRL porta all'approccio DRLSL, che tiene conto dei vincoli di sicurezza oltre alla massimizzazione delle ricompense. Ciò porta a vantaggi come la capacità di selezionare azioni sicure, riducendo il rischio di incidenti e aumentando l'affidabilità dei sistemi di guida autonoma (AD). Inoltre, un SLP fornisce una rappresentazione trasparente e interpretabile per la ragionevolezza sulla sicurezza, semplificando la verifica e la certificazione del sistema.

Infine, viene valutato il modello sia in fase di training che in fase di test per valutare il sistema utilizzando il dataset HighD. La rete era composta da tre strati completamente connessi, con i primi due strati composti da 256 nodi e l'ultimo strato composto da tre nodi per rappresentare lo spazio delle azioni. Tramite la libreria PyTorch vengono attuati i calcoli della rete neurale. L' algoritmo ADAM è stato utilizzato per ottimizzare le reti. Vengono valutate le prestazioni del metodo proposto confrontando i risultati del DQNSL con quelli del DQN.

Training. Viene fatto per 1.500 episodi in entrambi i metodi. Possiamo osservare che:

-Viene rilevata instabilità per l'agente DQN in quanto seleziona azioni in modo casuale dall'intero spazio delle azioni, mentre l'agente DQNSL seleziona azioni solo dall'insieme di azioni sicure.

-L'agente DQNSL riceve costantemente ricompense più elevate, indicando che evita azioni non sicure come collisioni, violazioni di corsia e manovre pericolose. D'altra parte, durante le fasi iniziali dell'addestramento, l'agente DQN subisce notevoli ricompense negative a causa delle frequenti violazioni di corsia e collisioni.

-Un'altra differenza significativa riguarda la velocità dell'addestramento. Come osservato in Figura 4, la ricompensa per l'agente DQNSL converge dopo 500 episodi, mentre la ricompensa dell'agente DQN converge in modo instabile dopo 970 episodi. L'eliminazione delle azioni non sicure gioca un ruolo cruciale in questa disparità. Limitando lo spazio delle azioni ed eliminando azioni.

In sintesi, l'approccio DQNSL migliora significativamente la sicurezza durante la fase di esplorazione. L'agente DQNSL supera l'agente DQN tradizionale evitando azioni non sicure e convergendo più rapidamente.

Invece, nella fase di test, ciò che ritengo importante evidenziare dai risultati in fase di test è il seguente punto. Viene posta una situazione particolarmente complicata, ovvero si pone il veicolo nel senso opposto di marcia

al fine di valutare le capacità dei vincoli definiti. L'agente DQN esce frequentemente dall' autostrada, evidenziando difficoltà nel gestire situazioni 'fuori dal comune'. In contrasto, l'agente DQNSL è rimasto costantemente all'interno dei confini dell'autostrada, evidenziando l'efficacia delle regole simboliche definite nel core di SLP. Inoltre, abbiamo valutato le prestazioni di entrambi i modelli quando il veicolo autonomo viaggiava in direzione destra-verso-sinistra, come mostrato in Figura 9, al fine di valutare la capacità dei vincoli definiti sull'autostrada di generalizzare la guida in diverse direzioni. Evidenziando ancora una volta come la natura simbolica di queste regole consente una facile generalizzazione a nuovi ambienti, consentendo il trasferimento e l'utilizzo delle regole definite in ambienti simili, garantendo una maggiore applicabilità e una migliorata capacità a generalizzare.

È stato presentato un nuovo tipo di modello che lavora su un diverso tipo di apprendimento rispetto ai modelli precedenti. ponendo in evidenza le problematiche e le difficoltà ed è stato dimostrato ancora una volta come le caratteristiche simboliche integrate a modelli neurali possano portare vantaggi a questi modelli. Nel dominio di auto a guida autonoma sono tante le sfide, la salute e l'incolumità dei passeggeri va messa al primo posto. Innescare meccanismi di ragionamento più elevato all' interno di questi modelli è fondamentale, inoltre attraverso definizione di regole e meccanismi di inferenza viene elevato il grado di interpretabilità dei dati nelle decisioni e scelte di questi sistemi. Ritengo che l'unione e la coesistenza di questi sistemi abbia grosse potenzialità e che rendendo questi sistemi più sicuri e interpretabili possa giovare anche nell' applicazioni delle leggi in ambito legale, soprattutto nel comprendere di chi sia la responsabilità in una determinata scelta. Senza lasciare nulla al caso.

3.6 Tabella comparativa e riflessioni

MODELLO	PARTE SIMBOLICA LOGICA	PARTE SUB SIMBOLICA NEURALE	TIPO TASSONOMIA	XAI	DATASET E AMBIENTE	FONTI
KEP per percezione	Attraverso KG, sfrutto relazioni tra entità per comprensione della scena del tipo <t,r,h>	percezioni ambientali riconosciute da algoritmi di visione artificiale (PCV) e rappresentati in un KG	Tipo 3. Interazione tra Reti Neurali e Sistemi Simbolici Complementari	Causabilità Equità Accessibilità Trasferibilità	PandaSet/ Nuscenes. CARLA	(Wickramarachchi, Henson, e Sheth 2022)
CoSI per comprensione scene. MRGCN	Classificare situazioni tramite regole dichiarative e potenza espressiva del KG	Previsioni o classificazioni tramite metodo di apprendimento in assenza di regole esplicite	Tipo 3. Interazione tra Reti Neurali e Sistemi Simbolici Complementari	Causalità Trasferibilità Equità Spiegazioni basate su filtri di causa	Simulations of Urban Mobility (SUMO)Footnote 4	(Halilaj et al. 2021)
ML- MAS	Sistema 2. Agenti razionali BDI interrogati per risolvere situazioni più complesse sui dati	Sistema 1. Modelli di ML pre addestrati per generazione dati pre elaborati. [LAV],[TransFuser]	Tipo 2 - Sistemi Ibridi con Risolutori Simbolici	Agenti BDI	Benchmark Longest6. CARLA	(Shukairi e Cardoso 2023)
T-norms	Conoscenza integrata nelle reti neurali come vincoli logici indotti da T-norme	3D-RetinaNet architettura per 'processamento di video online'. Random Connectivity Gated Recurrent Unit (RCGRU) attraverso LSTM	Tipo 5. Vincoli Soft di Conoscenza Simbolica in Reti Neurali Distribuite	Regole e valori Fuzzy in parte riflettono le decisioni prese dal sistema	DataSet ROAD-R	(Stoian, Giunchiglia, e Lukasiewicz, 2023)
DRLSL	Logiche simboliche del primo ordine FOLs (ragionamento su conoscenza) per imporre vincoli e limitare lo spazio delle azioni	Tecniche di Deep Reinforcment Learning. DQN Network	Tipo 3. Interazione tra Reti Neurali e Sistemi Simbolici Complementari	Interpretabilità attraverso SLP e FOLs	Dataset highD. Ambiente simulato basato su Pygame	(Sharifi, Yildirim, e Fallah 2023)

Fig. 20 – Elaborazione personale.

La valutazione è fatta in base a quanto visto nelle sezioni del capitolo 2:

i. Per quanto concerne l'analisi dei *compiti simbolici e sub simbolici*, ci riferiamo alle sezioni 2.2 e 2.4, in cui sono state delineate le differenze tra i due approcci e approfondita la tematica del ciclo neuro simbolico e dell'interazione tra i due sistemi. Mettendo in evidenza la tipologia di tali sistemi e il loro modo di interagire, miriamo a evidenziare i principi e gli studi condotti nella sezione 2.4 riguardo al ciclo neuro simbolico.
ii. La *tassonomia* segue quanto esposto nel lavoro "3rd Wave", conformemente alla classificazione dei tipi elaborata da Kautz. Si procede dalla categoria dei modelli ibridi più tradizionali al livello 1, per arrivare ai modelli completamente integrati di tipo 6.
iii. Esaminando quanto esposto nella sezione 2.5 e partendo dalla visualizzazione dei tipici compiti di guida (percezione, localizzazione, pianificazione, controllo), riteniamo di notevole interesse analizzare le principali sfide e i principali contributi che tali sistemi possono apportare al campo dell'intelligenza artificiale spiegabile (*XAI*). Questo contributo potrebbe contribuire a consolidare la fiducia degli utenti in questa area in via di sviluppo.
iv. Infine, considerando la complessità dello sviluppo di tali sistemi, va notato che condurre sperimentazioni su strade effettive spesso risulta impraticabile. Questi sistemi, se privi di sicurezza o se non sottoposti a valutazioni adeguate, possono costituire un rischio significativo e avere un impatto determinante sulla vita di altre persone. Pertanto, riteniamo opportuno menzionare anche l'uso dell'ambiente *di guida simulata* e dei diversi *set di dati* utilizzati durante le fasi di addestramento, validazione e test dei modelli trattati.

Ritengo che i sistemi analizzati in questo capitolo abbiano il potenziale per affrontare alcuni dei principali problemi legati all'opacità dei sistemi di apprendimento profondo (DL) e alla capacità di generalizzazione a nuove situazioni non precedentemente viste durante la fase di addestramento del modello. Questo è di fondamentale importanza in un ambiente "open world" come la guida autonoma.

Tra tutti i sistemi considerati, ritengo che gli studi presentati in KEP e CoSI, mediante l'uso di grafi di conoscenza (KG), possano essere particolarmente utili. Le rappresentazioni delle entità basate su connessioni relazionali e la creazione di un'ontologia possono contribuire a creare un sistema con una comprensione più avanzata dell'ambiente circostante, il che rende più significative le decisioni prese da questi sistemi automatizzati. Nel contesto della guida, ritengo essenziale che le decisioni non si basino esclusivamente sul riconoscimento di pattern tramite modelli di DL, ma che si sfrutti la potenza dei sistemi simbolici per simulare la capacità di guida umana all'interno dei veicoli. Le situazioni che possono verificarsi sulle strade sono estremamente varie, ed è chiaro che ha più senso dotare questi sistemi di capacità di ragionamento ad alto livello riguardo alla situazione e all'ambiente, piuttosto che cercare di presentare in fase di addestramento tutte le possibili situazioni, che sono praticamente infinite, verificabili sulla strada.

CAPITOLO 4
CONCLUSIONI E LAVORI FUTURI

Conclusioni

Il presente lavoro di tesi, attraverso un'analisi delle caratteristiche intrinseche dei modelli di Intelligenza Artificiale, ha voluto esplicitare le loro potenzialità e le diverse aree di applicazione, evidenziando i punti deboli dei modelli di Deep Learning nel contesto complesso dell'automazione della guida autonoma e proponendo come soluzione un approccio neuro-simbolico. Si è evidenziato come tale approccio, basato sull'integrazione di metodi simbolici in combinazione con modelli di Deep Learning, sia in grado di elevare il livello di capacità di ragionamento, riducendo in parte l'opacità intrinseca dei sistemi e migliorandone la spiegabilità, seguendo anche il concetto di Explicable Artificial Intelligence (XAI).

Gli attuali sistemi di auto a guida autonoma hanno il potenziale per ridurre incidenti, morti e il traffico, ma presentano anche rischi significativi. Elementi chiave come la presa di decisioni morali, la capacità di generalizzazione, la robustezza, la spiegabilità delle scelte del sistema e il livello di ragionamento sono fondamentali. Lo sviluppo di sistemi ibridi in questo settore può affrontare con successo queste sfide, contribuendo a costruire la fiducia del pubblico verso queste tecnologie. L'implementazione dei principi di spiegabilità (XAI) in tali modelli potrebbe trasformarli in strumenti potenti per potenziare la sicurezza stradale. La disciplina dell'Explainable AI inoltre dovrebbe lavorare in linea con i principi di trasparenza presenti all' interno dell''A.I act', al fine di risolvere problematiche legate al traffico e incidenti stradali dissipando così le preoccupazioni degli utenti e consentendo, di conseguenza, la loro diffusione e utilizzo su scala globale.

Ritenendo un'analisi puramente tecnica di questi problemi insufficiente, si è evidenziato il quadro legislativo di riferimento in Europa e negli Stati Uniti, con un confronto tra i diversi approcci nei confronti dell'intelligenza artificiale in generale e, in particolare, dei sistemi di guida autonoma, ponendo una particolare attenzione a questioni cruciali come la responsabilità, la privacy e le scelte etiche e morali connesse a tali modelli.

L'" A.I. Act" sembra poter fornire linee guida globali per la regolamentazione dell'uso dei sistemi di Intelligenza Artificiale, perché nonostante gli Stati Uniti siano in una posizione avanzata nello sviluppo di tali sistemi, l'Europa dimostra una maggiore attenzione per questioni legate alla responsabilità, alla protezione dei dati personali e allo sviluppo di sistemi etici ed equi.

Il nucleo della ricerca è stata l'analisi dell'approccio neuro-simbolico. Sono stati esaminati cinque modelli sperimentali e innovativi, con l'obiettivo di esplorarne in dettaglio la struttura interna, e si è cercato di categorizzarli, evidenziando le contribuzioni sia della componente neurale (Deep Learning) sia della componente simbolica. Attraverso questa analisi, sono stati identificati i vantaggi chiave derivanti dall'implementazione di un sistema simbolico in interazione con un modello neurale per quanto riguarda le azioni e i comportamenti delle auto, tenendo conto delle sfide principali relative alle possibili azioni intraprese dalle auto a guida autonoma e dei problemi connessi ai vari moduli di percezione dell'ambiente.

Si è potuto dedurre che nel contesto della guida è essenziale che le decisioni non si basino esclusivamente sul riconoscimento di pattern tramite modelli di DL, ma che utilizzino anche sistemi simbolici in grado di aumentare la capacità di ragionamento e la correttezza delle decisioni prese per simulare la capacità di guida umana all'interno dei veicoli.

Futuri lavori

È risultato evidente nei lavori sull'argomento presi in considerazione, che il campo dell'IA e della guida autonoma è in costante evoluzione. I risultati e le conclusioni presentati in questo lavoro possono essere considerati solo come un punto di partenza per ulteriori ricerche. Nella sezione seguente, indicheremo alcune delle direzioni possibili per futuri lavori di ricerca. Questi punti non solo rappresentano le sfide non risolte e le opportunità emergenti, ma riflettono anche l'importanza di un approccio interdisciplinare, che abbraccia gli aspetti tecnici, etici, legali ed economici dell'auto a guida autonoma e dell'IA in generale.

Responsabilità nelle auto a guida autonoma. In questo contesto ritengo importante la definizione di un chiaro quadro legislativo, soprattutto a livello Europeo, con una veloce recepimento da parte dei singoli stati delle normative già emanate. Potrebbe anche essere utile creare leggi apposite per ogni livello di automazione in conformità con i livelli SAE delle auto a guida autonoma: ciò permetterebbe di definire in modo chiaro se la responsabilità sia da imputare al conducente o al produttore. In questo contesto le tecniche di XAI e i metodi neuro-simboli sono utili per implementare 'by design' sistemi che riescano a ricostruire lo scenario e a rendere il processo decisionale dell'auto autonoma spiegabile o per lo meno interpretabile.

Sviluppo di un'etica nelle auto a guida autonoma. Aspetto legato in parte con il tema della responsabilità: in questo caso occorre concentrarsi su come possano venire rappresentate e come possano essere date le giuste istruzioni ad un sistema autonomo nel caso sia posto in scenari in cui entra in gioco la componente etica e morale. Ritengo utile osservare come questi temi possano essere oggetto di studio nei modelli KEP e CoSI, tramite KG e la creazione di ontologie. La questione è se e come tramite vincoli relazionali il sistema possa essere influenzato nelle scelte e se attraverso questi meccanismi si possa creare un agente moralmente corretto. Un'altra possibilità, invece di creare vincoli relazionali come in KEP, potrebbe essere quella di implementare un'etica attraverso il sistema DRSL: ipotizzo che con questo sistema, tramite SLP e utilizzo di FOLs, si possano eliminare dalle azioni possibili eseguibili quelle non corrette a livello etico.

A.I act sistemi critici. Uno sviluppo ulteriore è la definizione delle auto a guida autonoma all' interno dell'AI act come sistema a rischio elevato. Ora come ora la categoria non è ancora stata definita. Inoltre, secondo l'atto in questione, è necessaria una certificazione per i sistemi critici di Intelligenza Artificiale. Sarebbe interessante che tali certificazioni tenessero conto dello sviluppo di sistemi secondo metodologie di XAI, valutando l'efficacia delle attuali norme applicabili a queste tecnologie, fondendo così dominio tecnico e legale.

Integrazione dei sistemi neuro-simbolici. Il sistema di tipo 6 viene definito da Kautz come *'il livello più alto di integrazione, dove i sistemi sono in grado di eseguire il vero ragionamento simbolico all'interno di reti neurali'*(Kautz 2022). Questo è tuttora un ambito di ricerca aperto. Nel momento in cui verranno sviluppati questo tipo di sistemi, sarà necessario valutarne l'efficacia secondo le attuali metriche di valutazione rispetto a modelli classici, osservando il contributo che potrebbero dare in questo dominio. In particolare, dovrebbero essere oggetto di analisi i miglioramenti che questo approccio può apportare in situazioni di guida particolarmente complesse, come osservato già nei sistemi considerati nel mio lavoro di tesi. In conclusione, l'approccio neuro-simbolico, sebbene molto promettente, presenta anche alcune sfide in termini di complessità computazionale e di addestramento del modello: la totale integrazione dei sistemi è tuttora un argomento in via di sviluppo e che necessita di essere approfondita con ulteriori studi.

BIBLIOGRAFIA

Atakishiyev, Shahin, Mohammad Salameh, Hengshuai Yao, e Randy Goebel. 2023. «Explainable Artificial Intelligence for Autonomous Driving: A Comprehensive Overview and Field Guide for Future Research Directions». arXiv. http://arxiv.org/abs/2112.11561.

Bader, Sebastian, e Pascal Hitzler. 2005. «Dimensions of Neural-symbolic Integration - A Structured Survey». arXiv. http://arxiv.org/abs/cs/0511042.

Bazzocchi, Luciano. 1988. «Intelligenza artificiale e sistemi esperti». *Nuova civiltà delle macchine* 6 (1/2): 113–23.

Besold, Tarek R., Artur d'Avila Garcez, Sebastian Bader, Howard Bowman, Pedro Domingos, Pascal Hitzler, Kai-Uwe Kuehnberger, et al. 2017. «Neural-Symbolic Learning and Reasoning: A Survey and Interpretation». arXiv. http://arxiv.org/abs/1711.03902.

«CLOUD Act». 2023. In *Wikipedia*. Consultato 29 settembre 2023 https://en.wikipedia.org/w/index.php?title=CLOUD_Act&oldid=1168872635#External_links.

Commissione Europea, Commissione. 2021. «Proposta di regolamento del parlamento europeo e del consiglio che stabilisce regole armonizzate sull'intelligenza artificiale». COM2021/206 final, Bruxelles.

Confalonieri, Roberto, Ludovik Coba, Benedikt Wagner, e Tarek R. Besold. 2021. «A Historical Perspective of Explainable Artificial Intelligence». *WIREs Data Mining and Knowledge Discovery* 11 (1): e1391. https://doi.org/10.1002/widm.1391.

«Cosa sono gli Adas e perché saranno obbligatori dal 2022». 2021. YouGo. Consultato 30 settembre 2023 https://www.you-go.it/consigli/cosa-sono-gli-adas-e-perche-saranno-obbligatori-dal-2022/.

Crosley, Tom. 2020. «Are Self-Driving Cars Legal in Texas? (Are They Safe?)». Crosley Law. Consultato 29 settembre 2023 https://crosleylaw.com/blog/are-self-driving-cars-legal-in-texas/.

De Palma, Valeria. 2017. «Le Auto a Guida Autonoma e la responsabilità nella circolazione stradale». Consultato 27 settembre 2023 https://www.diritto.it/le-auto-guida-autonoma-la-responsabilita-civile-nella-circolazione-stradale/.

Dhirani, Lubna Luxmi, Noorain Mukhtiar, Bhawani Shankar Chowdhry, e Thomas Newe. 2023. «Ethical Dilemmas and Privacy Issues in Emerging Technologies: A Review». *Sensors* 23 (3): 1151. https://doi.org/10.3390/s23031151.

«Drunk Driving | NHTSA». s.d. Consultato 29 settembre 2023. https://www.nhtsa.gov/risky-driving/drunk-driving.

Foti, Miriam. 2023. «AI Act: con il voto del Parlamento l'UE traccia il futuro dell'Intelligenza Artificiale». Consultato 25 settembre 2023 *Altalex*. https://www.altalex.com/documents/news/2023/06/23/ai-act-ue-traccia-futuro-intelligenza-artificiale.

Garcez, Artur d'Avila, Marco Gori, Luis C. Lamb, Luciano Serafini, Michael Spranger, e Son N. Tran. 2019. «Neural-Symbolic Computing: An Effective Methodology for Principled Integration of Machine Learning and Reasoning». arXiv. http://arxiv.org/abs/1905.06088.

Garcez, Artur d'Avila, e Luís C. Lamb. 2023. «Neurosymbolic AI: The 3rd Wave». *Artificial Intelligence Review*, marzo. https://doi.org/10.1007/s10462-023-10448-w.

Gibaut, Wandemberg, Leonardo Pereira, Fabio Grassiotto, Alexandre Osorio, Eder Gadioli, Amparo Munoz, Sildolfo Gomes, e Claudio dos Santos. 2023. «Neurosymbolic AI and its Taxonomy: a survey». https://doi.org/10.48550/arXiv.2305.08876.

Goodall, Noah J. 2016. «Can you program ethics into a self-driving car?» *IEEE Spectrum* 53 (6): 28–58. https://doi.org/10.1109/MSPEC.2016.7473149.

Halilaj, Lavdim, Ishan Dindorkar, Jürgen Lüttin, e Susanne Rothermel. 2021. «A Knowledge Graph-Based Approach for Situation Comprehension in Driving Scenarios». In *The Semantic Web*, a cura di Ruben Verborgh, Katja Hose, Heiko Paulheim, Pierre-Antoine Champin, Maria Maleshkova, Oscar Corcho, Petar Ristoski, e Mehwish Alam, 699–716. Lecture Notes in Computer Science. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-77385-4_42.

IBM a «What Is Artificial Intelligence (AI) ? | IBM». s.d. Consultato 30 settembre 2023. https://www.ibm.com/topics/artificial-intelligence.

IBM b«What Is Machine Learning? | IBM». s.d. Consultato 30 settembre 2023. https://www.ibm.com/topics/machine-learning.

IBM c «What is a Knowledge Graph? | IBM». s.d. Consultato 29 settembre 2023. https://www.ibm.com/it-it/topics/knowledge-graph.

Kautz, Henry A. 2022. «The Third AI Summer: AAAI Robert S. Engelmore Memorial Lecture». *AI Magazine* 43 (1): 105–25. https://doi.org/10.1002/aaai.12036.

Mastromatteo, Alessandro. 2021. «Auto a guida autonoma, nuovi obblighi nel regolamento europeo su AI». Agenda Digitale. 23 aprile 2021. https://www.agendadigitale.eu/cultura-digitale/auto-a-guida-autonoma-regole-e-responsabilita-nel-regolamento-europeo-su-ai/.

«Mercedes to Accept Legal Responsibility for Accidents Involving Self-Driving Cars». 2022. *Driving.Co.Uk from The Sunday Times* (blog). Consultato 29 settembre 2023

https://www.driving.co.uk/news/technology/mercedes-to-accept-legal-responsibility-for-accidents-involving-self-driving-cars/.

Mickle, Tripp, Yiwen Lu, e Mike Isaac. 2023. «'This Experience May Feel Futuristic': Three Rides in Waymo Robot Taxis». *The New York Times*, 21 agosto 2023, sez. Technology. https://www.nytimes.com/2023/08/21/technology/waymo-driverless-cars-san-francisco.html.

Minsky, Marvin L. 1991. «Logical Versus Analogical or Symbolic Versus Connectionist or Neat Versus Scruffy». *AI Magazine* 12 (2): 34–34. https://doi.org/10.1609/aimag.v12i2.894.

Mökander, Jakob, Prathm Juneja, David S. Watson, e Luciano Floridi. 2022. «The US Algorithmic Accountability Act of 2022 vs. The EU Artificial Intelligence Act: What Can They Learn from Each Other?» *Minds and Machines* 32 (4): 751–58. https://doi.org/10.1007/s11023-022-09612-y.

«Moral Machine». s.d. Moral Machine. Consultato 11 settembre 2023. http://moralmachine.mit.edu.

«Morte di Elaine Herzberg». 2023. In *Wikipedia*. Consultato 29 settembre 2023 https://it.wikipedia.org/w/index.php?title=Morte_di_Elaine_Herzberg&oldid=133772045.

«Myths about Autonomous Driving». s.d. Audi MediaCenter. Consultato 29 settembre 2023. https://www.audi-mediacenter.com:443/en/press-releases/myths-about-autonomous-driving-14729.

Pober, Joseph, Michael Luck, e Odinaldo Rodrigues. s.d. «From Subsymbolic to Symbolic: A Blueprint for Investigation». Consultato 29 settembre 2023. https://ceur-ws.org/Vol-3212/paper6.pdf.

«Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (regolamento generale sulla protezione dei dati), Arricchito con riferimenti ai Considerando, Aggiornato alle rettifiche pubblicate sulla Gazzetta Ufficiale dell'Unione europea 127 del 23 maggio 2018 » Consultato 29 settembre 2023. https://www.garanteprivacy.it/documents/10160/0/Regolamento+UE+2016+679.+Arricchito+con+riferimenti+ai+Considerando+Aggiornato+alle+rettifiche+pubblicate+sulla+Gazzetta+Ufficiale++dell%27Unione+europea+127+del+23+maggio+2018.pdf/1bd9bde0-d074-4ca8-b37d-82a3478fd5d3?version=1.9

«Relazione sulla proposta di regolamento del Parlamento europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale) e modifica alcuni atti legislativi dell'Unione | A9-0188/2023 | Parlamento Europeo». Consultato 29 settembre 2023. https://www.europarl.europa.eu/doceo/document/A-9-2023-0188_IT.html.

Rubechini, Patrizio. 2020. «Il "diritto alla spiegazione" ex art. 22 GDPR e il Brexit case». *IRPA* (blog). 23 novembre 2020. https://www.irpa.eu/il-diritto-alla-spiegazione-ex-art-22-gdpr-e-il-brexit-case/.

Sharifi, Iman, Mustafa Yildirim, e Saber Fallah. 2023. «Towards Safe Autonomous Driving Policies using a Neuro-Symbolic Deep Reinforcement Learning Approach». arXiv. http://arxiv.org/abs/2307.01316.

Shukairi, Hilal Al, e Rafael C. Cardoso. 2023. «ML-MAS: A Hybrid AI Framework for Self-Driving Vehicles». In *AAMAS '23: Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems*, 1191–99. International Foundation for Autonomous Agents and Multiagent Systems (IFAAMAS). https://abdn.elsevierpure.com/en/publications/ml-mas-a-hybrid-ai-framework-for-self-driving-vehicles.

Stoian, Mihaela C, Eleonora Giunchiglia, e Thomas Lukasiewicz. s.d. «Exploiting T-Norms for Deep Learning in Autonomous Driving».

Swaminathan, Nandhini «"Neuro-Symbolic" AI. Where deep learning meets traditional… ». The Research Nest. s.d. Consultato 2 ottobre 2023. https://medium.com/the-research-nest/neuro-symbolic-ai-2fc77b544126.

«Waypoint - The official Waymo blog: Waymo's next chapter in San Francisco». s.d. Waypoint – The official Waymo blog. Consultato 3 ottobre 2023. https://waymo.com/blog/2023/08/waymos-next-chapter-in-san-francisco.html.

Wickramarachchi, Ruwan, Cory Henson, e Amit Sheth. 2022. «Knowledge-Based Entity Prediction for Improved Machine Perception in Autonomous Systems». *IEEE Intelligent Systems* 37 (5): 42–49. https://doi.org/10.1109/MIS.2022.3181015.

Yalçın, Orhan G. 2021. «Symbolic vs. Subsymbolic AI Paradigms for AI Explainability». 21 giugno 2021. Consultato 20 settembre 2023 https://towardsdatascience.com/symbolic-vs-subsymbolic-ai-paradigms-for-ai-explainability-6e3982c6948a.

Zorzi, Marco. 2022. «Intelligenza artificiale: Introduzione». Università degli studi di Padova. Consultato 29 settembre 2023 https://psico.elearning.unipd.it/enrol/index.php?id=4535.