EXPLAINABLE ARTIFICIAL INTELLIGENCE
PROFILI CIVILISTICI PER L'AUTOMOTIVE DI ULTIMA GENERAZIONE

Nell'ultimo decennio stiamo assistendo ad un notevole crescendo di applicazioni basate sull'utilizzo dell'Intelligenza Artificiale (IA). Dal campo medico al campo industriale, automotive, al settore consumer, l'IA sta significativamente migliorando i risultati e le metodologie di risoluzione delle problematiche presenti nei menzionati settori, contribuendo pertanto ad un complessivo miglioramento della società odierna. Ciò nonostante, nella comunità scientifica si è riscontrato un certo scetticismo in relazione all'utilizzo dell'IA per la soluzione dei problemi nei settori chiave sopra citati e ciò per la semplice ragione, che spesso i modelli di IA venivano concepiti come dei "Black-Box" incomprensibili all'utente umano e poco trasparenti in riferimento alle logiche interne di funzionamento. Per ovviare a tale problematica, si è sviluppato nel tempo il concetto di "Explainable Artificial Intelligence" (EAI) ossia uno sviluppo sostenibile dei modelli di IA che preveda l'utilizzo di metodologie "illustrative" delle elaborazioni compiute dall'algoritmo, consentendo pertanto un uso "consapevole" di queste tecnologie. Mediante la disamina di un caso-studio realmente implementato, nel presente contributo si evidenzieranno i vantaggi ed i profili giuridici e sociali che l'EAI offre nel delicato oltre che cruciale settore dell'Automotive.

di **Francesco RUNDO**, ingegnere informatico. Ha un Dottorato di Ricerca in Matematica Applicata conseguito presso l'Università di Catania. Svolge l'attività di Ingegnere-ricercatore presso la divisione di Ricerca e Sviluppo ADG della STMicroelectronics, occupandosi specificamente di Deep Learning e Modellistica Matematica avanzata per applicazioni nel settore industriale.

1. Introduzione

I principali car-makers sono da anni impegnati nell'integrazione delle nuove tecnologie all'interno delle c.d. *Next-Generation Cars* (auto del futuro), veicoli di ultima generazione che cambieranno per sempre il concetto di mobilità (si parla appunto di *mobilità innovativa*). Si tratta di innovazioni in ambito *automotive* estremamente avanzate che si distribuiscono in vari livelli di automazione e che vedranno l'introduzione di autovetture sempre più intelligenti e sicure. Si parla, appunto, di sistemi di assistenza alla guida, sistemi di comunicazione tra autovetture, monitoraggio della dinamica di guida sino ad arrivare ai livelli di innovazione più avanzati che includono la parziale o completa autonomia dell'autoveicolo. Si auspica che in un prossimo futuro le menzionate tecnologie diventino lo standard *de-facto* essendo in dotazione, di serie, su tutti i veicoli di nuova generazione. Le moderne tecnologie di Intelligenza Artificiale (IA) fungono da acceleratore a questi processi di innovazione in ambito automotive, rendendo possibile l'implementazione di soluzioni un tempo irrealizzabili.

Per tale ragione, la maggior parte delle aziende automobilistiche è impegnata in forti investimenti in Ricerca e Sviluppo nonostante, ad oggi, mancano ancora i presupposti legati alle infrastrutture e alle normative di legge a supporto di tali processi innovativi. Per far sì che le auto intelligenti a guida autonoma possano finalmente diffondersi, saranno necessari ingenti investimenti per **modernizzare le infrastrutture stradali,** supportare le comunicazioni sicure (blockchain), supportare l'IoT in ambito automotive ovvero definire assetti normativi che disciplinano correttamente le valutazioni eseguite dai sistemi di Intelligenza Artificiale.

Il limite intrinseco che ad oggi impedisce l'integrazione completa (strutturale e giuridica) nel settore *automotive* delle soluzioni basate su Intelligenza Artificiale è spesso legato all'incapacita di decifrare i meccanismi che i sistemi di IA utilizzano per processare i dati. L'IA è spesso vista come una "Black Box" che produce un risultato in funzione di un certo dato di ingresso. Nel presente contributo si presenteranno i vantaggi insiti nella moderna ***Explainable Artificial Intelligence*** (EAI) cioè algoritmi di intelligenza artificiale "interpretabili e spiegabili" (per l'uomo/utente) che pertanto possono fornire utili informazioni per migliorare l'integrazione strutturale e giuridica di queste nuove teconolgie non solo nel settore automotive ma in generale nel settore industriale, medicale , etc..

Attraverso un esempio applicativo si mostrerà come l'EAI può miglioare significativamente la sicurezza stradale ed in generale la sicurezza del guidatore fornendo nel contempo valide interpretazioni dei dati processati utilizzabili per fini giuridici ed assicurativi ovvero per interagire con le infrastrutture esterne realizzate a supporto dei processi innovativi collegati alle Next-Generation Cars.

Infatti sia in ambito automotive che in qualsiasi altro campo applicativo, la normativa impone che le regole che disciplinano il settore devono essere chiare, univoche , possibili e comprensibili tra le parti. Giusto per citare un esempio applicativo semplice che chiarisca il concetto, in ambito contrattuale il codice civile richiede che in relazione al contratto, la redazione di quest'ultimo deve soddisfare precisi requisiti tra cui determinabilità, possibilità, liceità a pena di nullità ex art. 1346 c.c. Recentemente, ad esempio, si parla molto dei c.d. *smart-contract* intelligenti ovvero dell'intervento dell'IA nella stesura, analisi e chiusura di contratti in qualsiasi settore. Ebbene, l'algoritmo di IA che sarà sviluppato per la redazione degli smart-contracts (anche in ambito automotive), dovrà soddisfare i requisiti imposti dalla normativa vigente, e dunque, ad una verifica di un organo esterno, per esempio di un organo giudicante o di un organo di controllo, dovranno essere chiari, trasparenti e comprensibili i passi che l'algoritmo di IA ha compiuto per la negoziazione dello smart-contract al fine di pervenire ad una corretta valutazione del suo operato. Analogamente, è facile estendere le ridette valutazioni in campo assicurativo, amministrativo, etc.. Il seguente paragrafo introduce pertanto il concetto insito nelle moderne tecnologie di EAI fornendo dei brevi richiami al riguardo.

2. Explainable Artificial Intelligence (EAI): Brevi richiami

Negli ultimi anni, l'Intelligenza Artificiale (AI) ha avuto un notevole slancio contribuendo a fornire nuove ed interessanti aspettative su molti settori applicativi tra cui, nel caso che qui interessa, vale la pena annoverare il settore *Automotive*. Tuttavia, affinchè le soluzioni di AI siano largamente utilizzabili è necessa-

rio renderli "spiegabili" o "interpretabili" ossia bisogna fornire dei paradigmi comprensibili all'utente che illustrano i "processi interni" compiuti dall'algoritmo di IA per produrre un determinato risultato.
In tale ambito, nella communita scientifica dei ricercatori di IA, sono stati coniati dei termini specifici che illustrano molto chiaramente il concetto di *Explainable Artificial Intelligence (EAI)*. In special modo, introdurremo i concetti di *Understandability (Logica comprensibilità), Comprehensibility (comprensibilità), Interpretability (interpretabilità), Explainability (spiegabilità), Transparency (Trasparenza)* [1].
Nello specifico, un algoritmo o sistema basato su IA è ***logico-comprensibile (o equivalentemente, l'intelligibile)*** se il modello che implementa risulta logico e comprensibile ad un essere umano senza necessita di spiegarne in profondità la sua struttura interna o i passi algoritmici usati dal modello per elaborare i dati internamente. Un sistema di IA è dotato della caratteristica di ***comprensibilità*** se l'algoritmo implicito di apprendimento può essere rappresentato in modo comprensibile ad un essere umano. Questa nozione di comprensibilità deriva dai postulati di *Michalski* [2], che affermava che "*i risultati dell'induzione del computer dovrebbero essere descrizioni simboliche di determinate entità, semanticamente e strutturalmente simili a quelle che un esperto umano potrebbe produrre osservando le stesse entità. Le componenti di queste descrizioni dovrebbero essere comprensibili come singole "porzioni" di informazione, direttamente interpretabili in linguaggio naturale, e dovrebbero mettere in relazione concetti quantitativi e qualitativi in modo integrato*". Un sistema di AI deve altresi' essere ***interpretabile*** ossia deve essere in grado di fornire il significato semantico dei passi algoritmici implementati in termini comprensibili ad utente. Similmente il sistema di IA deve essere ***spiegabile*** ossia deve essere dotato di una interfaccia "user-friendly" per l'essere umano cosi da rendere chiaro e replicabile l'algoritmo sottostante. Infine un sistema di AI sarà ***trasparente*** se è in grado di rendere evidente in modo chiaro e logico le valutazioni algoritmiche compiute per produrre un determinato risultato. Poichè i livelli di trasparenza dei modelli di IA possono differire in entità, questi sono divisi in tre categorie: *modelli simulabili*, *modelli scomponibili* e *modelli trasparenti algoritmicamente*. Pertanto si comprende che la moderna ricerca in ambito IA è prevalentemente impegnata non solamente nello sviluppo di soluzioni efficienti, altamente performanti e sostenibili , quanto piuttosto, nella ricerca di soluzioni che pur avendo le caratteristiche appena citate, siano altresi' utilizzabili nel contesto sociale che richiede appunto che tali algoritmi siano comprensibili, interpretabili, spiegabili e trasparenti ad una verifica di un utente esterno. Nella figura che segue è illustrata la menzionata differenza tra il precedente paradigma di IA basato sul modello Black-Box che indirizzava le problematiche senza però fornire un modello algoritmico comprensibile all'utilizzatore, dalle moderne ed avanzate tecniche di EAI.

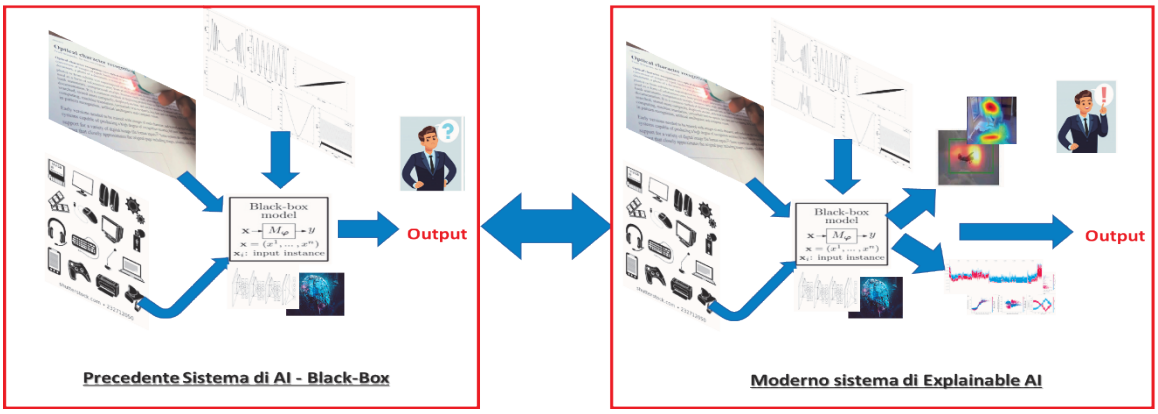

Figura 1 - Raffronto tra il precedente modello di IA (Black-Box) vs moderni paradigmi di Explainable AI.

Esistono diverse metodologie per rendere interpretabile e comprensibile un algoritmo di IA. Tra queste, analizzeremo nel presente contributo la metodologia Grad-CAM [3].
La metodologia *Gradient-Weighted Class Activation Mapping* (Grad-CAM), utilizza i gradienti dell'errore generati durante la procedura

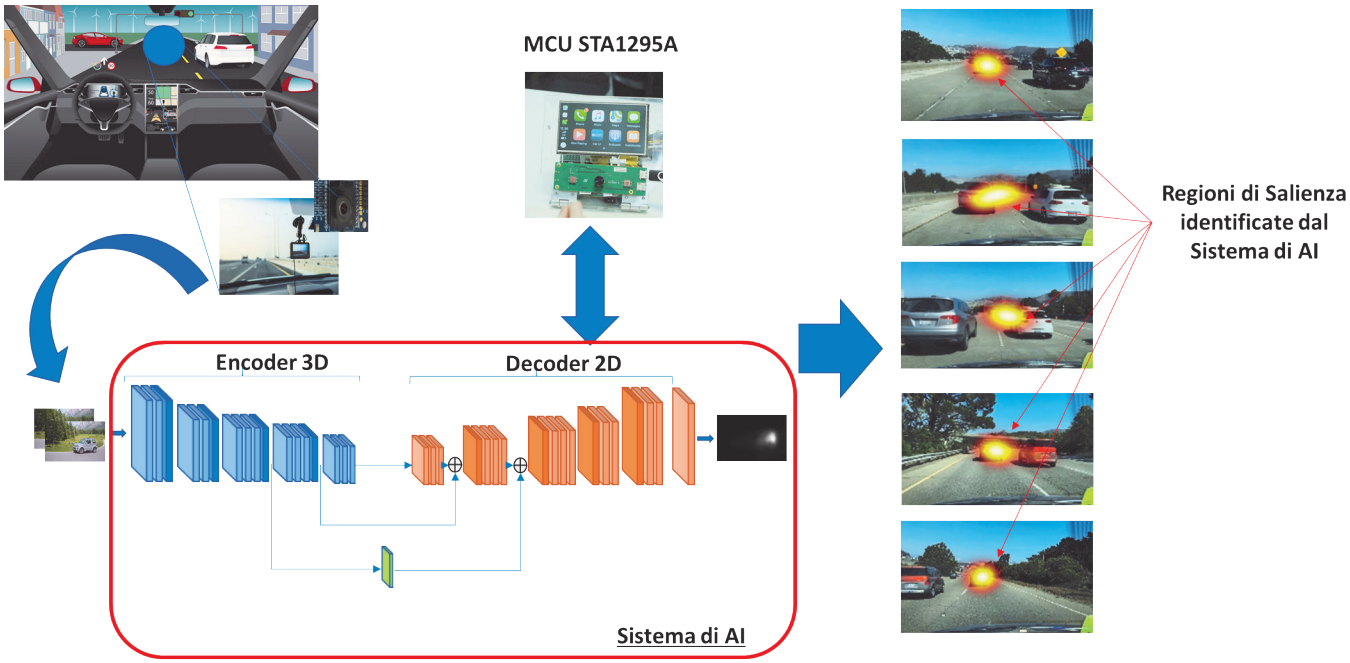

Figura 2 - Sistema di guida autonoma mediante algoritmo di IA: Il sistema identifica gli oggetti salienti nella scena di guida (regioni in giallo nei frames catturati dalla camera esterna e rappresentativi della scena di guida) e sulla base di questi dati decide le azioni da compiere.

di apprendimento del sistema di IA, al fine di produrre una mappa di salienza che evidenzia le regioni delle *features di apprendimento* (e dunque dei dati di ingresso) che maggiormente contribuiscono al risultato finale, durante la computazione algoritmica. Pertanto, attraverso la tecnica Grad-CAM sarà possibile evidenziare le parti "salienti" dei dati di ingresso che maggiormente contribuiscono al risultato algoritmico rendendo pertanto comprensibile e trasparente l'approccio usato dal metodo IA nella definizione della strategia risolutiva.
Nel seguente caso-studio sarà mostrato un esempio applicativo in campo automotive nel quale l'utilizzo di Grad-CAM renderà comprensibile ed interpretabile l'algoritmo di IA implementato.

3. Explainable IA in Automotive: Caso-Studio

Dalla fine degli anni novanta ricercatori ed esperi hanno lavorato senza sosta sull'implementazione di sistemi di guida assistita c.d. ADAS (*Adaptive Driving Assistance Systems*). Nel corso del decennio successivo, l'industria automobilistica ha dato ulteriore impulso a tale ricerca gettando le basi per la definizione dei sistemi successivi ai modelli ADAS ossia i sistemi di guida autonoma. A tal fine è stato redatto uno standard che definisce i 6 livelli di automazione per il settore automotive [4]: Si parte dal livello o "nessuna autonomia" (l'autista guida completamente senza l'aiuto del sistema di guida assistita) per poi progredire nei livelli di automazione e passare ai livelli 1 e 2 di "automazione parziale" ,dunque, ai livelli 3 "automazione condizionata"e 4 "alta automazione" sino ad arrivare al livello 5, quando il veicolo si muove in completa autonomia ovvero senza alcuna intervento da parte del guidatore. Appare evidente che nei processi di automazione degli autoveicoli, l'IA rivesta un ruolo strategico in quanto consente di ottenere "l'intelligenza" necessaria a ciascuno dei livelli di automazione alla guida. Ebbene, nell'ambito della guida autonoma o assistita risulta estremamente importante comprendere la logica sottostante dei modelli di IA adottati al fine di stabilire, in caso di eventi imprevisti (incidenti, azioni errate compiute dal veicolo autonomo, etc..), il grado di responsabilità del sistema di IA ovvero le parti della logica sottostante che hanno contribuito o meno all'evento imprevisto. Appare evidente senza necessita che si approfondisca ulteriormente, il perchè di tale esigenza quanto meno ai fini giuridici, assicurativi ed in ottica di valutazione della responsabilità civile, amministrativa e se occorre anche penale. Il seguente caso studio mostra un classico sistema che è stato realizzato sia

per assistenza alla guida (Livello di Automazione "2") che per la guida autonoma (Livello di Automazione "4" o "5"). Nella figura che segue viene riportato il modello di IA (Deep Learning) che realizza questo sistema:

Il sistema riportato in Fig. 2 è stato realizzato in un framework hardware basato su piattaforme STA1295A [5] dotato di accelerazione grafica adatta ad applicazioni di IA che siano *automotive-grade*. La rete di Deep learning, cuore del modello IA riportato in Fig.2, è composta da una Fully Convolutional Neural Network (FCNN) [6] composta da un Encoder 3D che processa e codifica i video frames che vengono catturati dalla camera esterna all'auto che campiona la scena di guida e da un Decoder 2D che farà un analisi semantica dei frames cercando di identificare gli oggetti salienti (auto, uomini, ostacoli, etc..) nella scena di guida cosi da intraprendere le azioni appropriate (assistenza al guidatore ovvero guida autonoma). Ciò è particolarmente importante, come si evidenzia dai frames riportati in Fig. 2, durante le fasi più rischiose della guida quali esempio, un sorpasso o un cambio corsia.

Pertanto, il sistema di AI realizzato mediante l'architettura illustrata in Fig. 2 sarà in grado non solamente di fornire il livello di assistenza alla guida o automazione richiesto ma fornirà, altresi', elementi di facile interpretazione e comprensione (oggetti salienti evidenziati nei frames della scena di guida), che consentiranno di identificare i processamenti logico-matematici sulla base dei quali il modello di AI intraprenderà determinate azioni. In caso di incidenti o eventi imprevisti, dall'analisi delle mappe di salienza dei frames rappresentativi della scena di guida, sarà possibile comprendere le informazioni che il sistema di AI ha identificato e processato individuando, nel caso ad esempio di assistenza alla guida, le precise responsabilità del sistema di IA e quelle del guidatore in relazione alle azioni che il modello algoritmico aveva suggerito in raffronto con quanto il conducente ha invece concretamente applicato. Una imparziale e puntuale analisi di questi dati espletata dagli organi di controllo offrirà una sorta di "scatola nera" dell'autoveicolo contribuendo alla precisa e puntuale ricostruzione dell'accaduto ovvero alla precisa definizione delle responsabilità civili, assicurative, amministrative e se necessario anche penali delle parti coinvolte nell'evento imprevisto.

4. Conclusioni

Dal caso studio sopra riportato, si delineano gli indubbi vantaggi dei moderni sistemi di EAI in paragone ai precedenti sistemi di IA considerati di fatto dei "Black-Box" incapaci di fornire elementi di "prova" a supporto delle elaborazioni algoritmiche compiute. Dall'ambito automotive all'ambito industriale, medicale, consumer e via dicendo, i sistemi di EAI offrono notevoli vantaggi in ordine al loro utilizzo giuridicamente sostenibile oltre che alla loro efficienza in termini di performance. Si auspica che in futuro una normativa più sensibile a questi temi ed una infrastruttura tecnologica idonea a sostenere tali sistemi, possa contribuire maggiormente alla diffusione di una metodologia di sviluppo dell'Intelligenza Artificiale che potrebbe realmente migliorare significativamente non solamente il settore automotive ma quanto, piuttosto, ogni area scientifica in cui possono trovare applicazioni i moderni sistemi di "intelligenza comprensibile". ©

REFERENCES

[1] Alejandro Barredo Arrieta, et al, Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI, Elsevier Information Fusion Journal, Volume 58, June 2020, Pages 82-115

[2] D.M. West The future of work: robots, AI, and automation, Brookings Institution Press (2018)

[3] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, 2017, pp. 618-626, doi: 10.1109/ICCV.2017.74.

[4] https://web.archive.org/web/20170903105244/https://www.sae.org/misc/pdfs/automated_driving.pdf

[5] https://www.st.com/en/automotive-infotainment-and-telematics/sta1295.html

[6] E. Shelhamer, J. Long and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 640-651, 1 April 2017, doi: 10.1109/TPAMI.2016.2572683.