

Intelligenza Artificiale Responsabile e Affidabile presso l'Università degli Studi di Catania

Salvatore Amato¹, Alberto Andronico¹, Elsa Bivona¹, Sebastiano Battiato²,
Daniela Giordano³, Alfio Grasso¹, Antonio Las Casas¹, Luca Guarnera², Gabriella Nicosia¹,
Alessandro Ortis², Simone Palazzo³, Giovanni Raiti¹, David Roccaro¹, Ugo Salanitro¹, Biagio
Spampinato¹, Concetto Spampinato³

¹Dipartimento di Giurisprudenza

²Dipartimento di Matematica e Informatica

³Dipartimento di Ingegneria Elettrica, Elettronica e Informatica

Università degli Studi di Catania

{nome.cognome}@unict.it

Abstract

Le attività di ricerca svolte presso l'Università degli Studi di Catania nell'ambito dell'affidabilità e responsabilità nelle applicazioni di intelligenza artificiale (AI) affrontano queste tematiche da molteplici prospettive, sia tecnologiche che etico-giuridiche. I risultati ottenuti negli ultimi anni, oltre ad essere oggetto di pubblicazioni scientifiche nelle principali riviste del settore, hanno posto le basi per il coinvolgimento dei partecipanti in diversi progetti. L'Università di Catania inoltre partecipa al Dottorato nazionale in Intelligenza Artificiale (Salute e scienze della vita) con la tematica "Explainable AI" (XAI).

1 Affidabilità e Interpretabilità

Le attività di ricerca del gruppo del Pattern Recognition and Computer Vision Laboratory (PeRCeVe Lab) presso il Dipartimento di Ingegneria Elettrica, Elettronica e Informatica riguardano lo studio del legame tra affidabilità e interpretabilità di modelli decisionali basati su modelli neurali. Difatti, oggi si parla sempre più spesso dell'impiego dell'intelligenza artificiale in molteplici ambiti, ma per una completa adozione, soprattutto in settore critici (come può essere quello sanitario) è fondamentale garantire affidabilità delle scelte dei modelli AI e conseguentemente anche spiegare agli stakeholder le motivazioni dietro quelle scelte. Proprio in quest'ottica, il gruppo del PeRCeVe Lab si occupa di studiare le relazioni che esistono tra questi due aspetti investigando come l'adversarial learning può supportare e implicare l'interpretabilità e la sua generalizzazione in diversi scenari, che vanno dall'analisi di immagini mediche per la profilazione di tumori del polmone all'automotive:

- In [Pino *et al.*, 2021] è proposto un approccio per la caratterizzazione genomica del tumore al polmone tramite analisi delle CT, promuovendo l'interpretabilità alle-

nando il modello in modo da essere robusto ad attacchi "white-box";

- Sempre in ambito sanitario, di recente è stato anche studiato l'utilizzo delle architetture a Transformer in ambito visuale, sia come strumento affidabile per la classificazione di dati visuali che negli aspetti legati all'interpretabilità. In particolare, l'uso combinato di architetture ibride convolutive-transformer ha permesso di creare modelli caratterizzati da bias induttivo, interpretabilità e robustezza in grado di affrontare sfide complesse, come la predizione dell'espressione genica in tumori polmonari partendo da dati TC o la classificazione del grado di severità di lesioni al pancreas [Proietto Salanitro *et al.*, 2022].
- Allo stesso modo, in ambito automotive, è stato studiato l'effetto della tecnica gradient reversal sulle capacità di generalizzazione di sistemi di rilevamento dei punti salienti osservatori da guidatori [Bellitto *et al.*, 2021; Rundo *et al.*, 2021].

1.1 Progetti

Le attività di ricerca oggetto di questa tematica sono state svolte grazie al supporto dei seguenti progetti:

- RehaStart (PO FESR 2014-2020, Azione 1.1.5, CUP: G79J18000610007)
- iHOSP - Assistenza medica domiciliare a distanza attraverso una infrastruttura interattiva di smart medicine (PO FESR 2014-2020, Azione 1.1.5, CUP: G59J18000510007)
- Progetto "Adaptive Brain-Derived Artificial Intelligence Methods for Event Detection" – BrAIn - Bando Piano di incentivi per la ricerca di Ateneo 2020/2022 (Pia.ce.ri.) – STARTING GRANT 2020, UNICT

1.2 Principali partecipanti

Concetto Spampinato, Daniela Giordano, Simone Palazzo, Giovanni Bellitto, Federica Proietto Salanitri, Matteo Pennisi, Raffaele Mineo.

2 AI contro la manipolazione dei contenuti multimediali: DeepFake

Nell'ambito delle attività di ricerca su Multimedia Forensics and Security (<https://iplab.dmi.unict.it/mfs/>) attivo presso il dipartimento di Matematica e Informatica, sono in corso di sviluppo tecniche e metodologie per l'analisi di contenuti multimediali manipolati, con particolare focus ai DeepFakes, che rappresenta uno degli asset scientifici della sede, con numerose pubblicazioni e diverse attività progettuali, organizzative (conferenze, workshop, ecc.) e di formazione. Si segnalano inoltre i vari studi rivolti all'individuazione di "best practice" nei vari ambiti della Digital Forensics anche alla luce del sempre più pervasivo utilizzo delle tecniche di Intelligenza Artificiale, per il trattamento e l'analisi dei file di natura multimediale [Battiatto *et al.*, 2021a], [Battiatto *et al.*, 2021b], [Costantini *et al.*, 2021], [Giudice *et al.*, 2021], [Guarnera *et al.*, 2020], [Guarnera *et al.*, 2021].

Sono al momento attive diverse collaborazioni scientifiche di formazione e di ricerca formalizzate con enti e istituzioni pubbliche (Racis Carabinieri, Polo Tecnico Presidenza del Consiglio dei Ministri, Polizia Scientifica, Scuola Nazionale Magistratura, ecc.) e privati (Amped, Cy4GATE, IISFA, ecc.) nonché attività di consulenza scientifica a servizio di FF.OO. e della magistratura inquirente. Presso UniCT è inoltre attivo lo spinoff iCTLAB leader nazionale nel settore della Digital Forensics e della prototipazione avanzata di soluzioni di Computer Vision (settori: intelligence e retail).

2.1 Principali partecipanti

I partecipanti alle attività di ricerca sono Sebastiano Battiatto, Luca Guarnera, Alessandro Ortis, Oliver Giudice.

3 Aspetti etico-giuridici in AI

Il gruppo di ricerca presso il Dipartimento di Giurisprudenza, composto da filosofi del diritto, privatisti, pubblicisti, laboristi, ha già elaborato un progetto condiviso con l'acronimo SMART, che ha avviato ricerche intorno ai temi attinenti la rilevanza dei profili etico-giuridico connessi con le diverse applicazioni dell'intelligenza artificiale. Il gruppo di ricerca ha organizzato un convegno di rilevanza internazionale nel periodo 30 settembre 2021–2 ottobre 2021, coinvolgendo docenti universitari, anche stranieri (G. Teubner), magistrati, tecnici, professionisti legali, affrontando sia le questioni generali etiche e giuridiche nella gestione degli algoritmi, sia questioni tematiche, attinenti all'uso dell'intelligenza artificiale in medicina, nelle decisioni algoritmiche, con particolare riferimento alla giustizia predittiva, nonché in relazione agli smart contract e allo smart working.

Il diritto di fronte alla IV Rivoluzione industriale

L'ormai imminente quarta rivoluzione industriale promette di abbattere i confini tra il mondo fisico, digitale e biologico.

Ma "rivoluzione" non è di per sé sinonimo di miglioramento. Compito del diritto è promuovere e assicurare i valori umani anche nell'era dell'evoluzione digitale. A partire dal problema delle barriere tra organico e inorganico e tra materiale e immateriale, la linea generale della ricerca ha sviluppato i risultati scientifici e le riflessioni filosofico-giuridiche più recenti sul rapporto tra intelligenza sintetica e intelligenza biologica [Amato, 2020], [Amato, 2021].

La responsabilità da robot

Robot capaci di assumere decisioni autonome, conseguenti a un processo di adattamento self learning, possono determinare un danno difficile da risarcire. Costruire la soggettività degli agenti software o modificare la disciplina della responsabilità civile adattando le norme in tema di sicurezza e di responsabilità del produttore. Robotica avanzata e danno medico: quale modello di responsabilità è più idoneo a rispondere ai problemi sollevati dall'avvento di sistemi di IA ad alto rischio [Salanitro, 2020].

Le discriminazioni algoritmiche

Verifica del rischio discriminatorio insito nei sistemi decisionali automatizzati di machine learning. I modelli ordinali alla prova dei pericoli posti dai bias e dal fenomeno dell'interoperabilità dei sistemi di IA. Risk management approach vs accountability approach: approcci diversi nel panorama legislativo europeo alla prova del principio di effettività delle tutele [Grasso, 2022].

Limiti alle decisioni algoritmiche e giustizia predittiva

La decisione algoritmica. L'utilizzo dell'intelligenza artificiale nell'amministrazione pubblica. Legittimità e trasparenza del procedimento amministrativo nella data-driven society. Giudice-robot e informatica giuridica: fino a quando i canoni del giusto processo potranno tollerare, senza smarrirsi, l'impiego dell'intelligenza artificiale. Dinamiche del contraddittorio e autonomia del magistrato vs speditezza e prevedibilità del giudizio [Andronico e Casadei, 2021], [Andronico, 2021], [Spampinato, 2021], [Roccaro, 2020], [Roccaro, 2021].

Dati personali, lavoro e contratto nell'era digitale

Nell'ambito dell'uso dei dati personali e del lavoro, il gruppo di ricerca del Dipartimento di Giurisprudenza ha affrontato le seguenti tematiche: 1) Il tema della rilevanza del consenso nella cessione dei dati personali: rilevanza dell'autonomia privata e degli istituti contrattuali; 2) La disciplina dello smart working; 3) Algoritmi e diritto del lavoro e 4) Smart contract e applicazione della disciplina del contratto. [Bivona, 2021], [Nicosia, 2020], [Nicosia e Saracini, 2021], [Nicosia, 2021].

3.1 Progetti

Le attività di ricerca oggetto di questa tematica sono state svolte grazie al supporto dei seguenti progetti:

- "SMART Soggetti e Macchine tra Autonomia e responsabilità quali tutele. Le discipline sostanziali e processuali per la società della IV rivoluzione, dopo la pandemia" — Linea 2 del Piano di incentivi per la ricerca del-

l'Ateneo di Catania 2020/2022 (principal investigator: Ugo Salanitro)

3.2 Principali partecipanti

I partecipanti alle attività di ricerca sono: Salvatore Amato, Alberto Andronico, Elsa Bivona, Alfio Guido Grasso, Antonio Las Casas, Gabriella Nicosia, Giovanni Raiti, David Roccaro, Ugo Salanitro, Biagio Spampinato.

Riferimenti bibliografici

- [Amato, 2020] S. Amato. Biodiritto 4.0. intelligenza artificiale e nuove tecnologie. *Torino: Giappichelli*, 2020.
- [Amato, 2021] S. Amato. Artificial intelligence and constitutional values. *Encyclopedia of Contemporary Constitutionalism*, 2021.
- [Andronico e Casadei, 2021] Alberto Andronico e T. Casadei. Introduzione a algoritmi ed esperienza giuridica. *Ars Interpretandi*, 2021.
- [Andronico, 2021] Alberto Andronico. Giustizia digitale e forme di vita. *Teoria e Critica della Regolazione Sociale/Theory and Criticism of Social Regulation*, 2021.
- [Battiato et al., 2021a] Sebastiano Battiato, Oliver Giudice, Francesco Guarnera, e Giovanni Puglisi. Estimating previous quantization factors on multiple jpeg compressed images. *EURASIP Journal on Information Security*, 2021(1):1–11, 2021.
- [Battiato et al., 2021b] Sebastiano Battiato, Oliver Giudice, Francesco Guarnera, e Giovanni Puglisi. First quantization estimation by a robust data exploitation strategy of dct coefficients. *IEEE Access*, 9:73110–73120, 2021.
- [Bellitto et al., 2021] G Bellitto, F Proietto Salanitri, S Palazzo, F Rundo, D Giordano, e C Spampinato. Hierarchical domain-adapted feature learning for video saliency prediction. *International Journal of Computer Vision*, 129(12):3216–3232, 2021.
- [Bivona, 2021] E. Bivona. Smart contracts e interferenze con la disciplina sui contratti: il sistema dei rimedi alla prova degli algoritmi. *Persona e mercato*, 2021.
- [Costantini et al., 2021] Federico Costantini, Fausto Galvan, Marco Alvisè De Stefani, e Sebastiano Battiato. Assessing information quality in iot forensics: Theoretical framework and model implementation. *Journal of Applied Logics*, 2021.
- [Giudice et al., 2021] Oliver Giudice, Luca Guarnera, e Sebastiano Battiato. Fighting deepfakes by detecting gan dct anomalies. *Journal of Imaging*, 7(8), 2021.
- [Grasso, 2022] A. G. Grasso. The bad algorithm. Automated Discriminatory Decisions in the European General Data Protection Regulation. *Working paper*, 2022.
- [Guarnera et al., 2020] Luca Guarnera, Oliver Giudice, e Sebastiano Battiato. Fighting deepfake by exposing the convolutional traces on images. *IEEE Access*, 8:165085–165098, 2020.
- [Guarnera et al., 2021] Francesco Guarnera, Oliver Giudice, Dario Allegra, Filippo Stanco, Sebastiano Battiato, Salvatore Livatino, Vito Matranga, e Angelo Salici. A robust document identification framework through f-bp fingerprint. *Journal of Imaging*, 7(8):126, 2021.
- [Nicosia e Saracini, 2021] Gabriella Nicosia e Paola Saracini. Bienestar organizativo, igualdad de oportunidades y la prohibición de la discriminación en la administración italiana. In *Los derechos fundamentales en el empleo público: Sistemas jurídicos italiano y español: Cuestiones actuales y puntos críticos*, pages 33–50. Comares, 2021.
- [Nicosia, 2020] G Nicosia. Pandemia, nuovi lavori (e servizi) essenziali e nuove libertà, 2020.
- [Nicosia, 2021] G. Nicosia. Potere datoriale pubblico, lezione dello statuto e management umanistico nello smart working post-emergenziale. *Quaderni della Rivista DLM*, 2021.
- [Pino et al., 2021] Carmelo Pino, Simone Palazzo, Francesca Trenta, et al. Interpretable deep model for predicting gene-addicted non-small-cell lung cancer in ct scans. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 891–894. IEEE, 2021.
- [Proietto Salanitri et al, 2022] Federica Proietto Salanitri e et al. Neural transformers for intraductal papillary mucosal neoplasms (ipmn) classification. In *Submitted to EMBC 2022*. IEEE, 2022.
- [Roccaro, 2020] David Roccaro. Tecnologie normative: verso un diritto avvolto dal digitale. *Jura Gentium: Rivista di filosofia del diritto internazionale e della politica globale*, 17(2):114–123, 2020.
- [Roccaro, 2021] David Roccaro. La chiave di accesso alla “società algoritmica”. *Ars interpretandi*, 26(1):69–82, 2021.
- [Rundo et al., 2021] Francesco Rundo, Roberto Leotta, Sebastiano Battiato, Concetto Spampinato, e Sabrina Conoci. Gradient reversal domain adaptation pipeline in advanced driver assistance systems. In *2021 AEIT International Conference on Electrical and Electronic Technologies for Automotive (AEIT AUTOMOTIVE)*, pages 1–6. IEEE, 2021.
- [Salanitro, 2020] Ugo Salanitro. Intelligenza artificiale e responsabilità: la strategia della commissione europea. *Rivista di diritto civile*, 66(6):1246–1276, 2020.
- [Spampinato, 2021] B. Spampinato. Per un primo inquadramento teorico della decisione amministrativa algoritmica. *Rivista Giuridica AmbienteDiritto.it*, 2021.