I quaderni di Agenda Digitale.eu

Maggio-Agosto 2021

n. 0008

Agendadigitale.eu è una testata scientifica e giornalistica registrata al Tribunale di Milano
Dati di riferimento

Iscrizione ROC n. 16446
ISSN 2421-4167
Numero registrazione 1927, Tribunale di Milano
Editore: Digital360
Focus e ambito:
La rivista scientifica, i Quaderni di Agendadigitale.eu, pubblica fascicoli quadrimestrali in open access.
Lo scopo è creare un luogo per accompagnare i passi dell'Italia verso la necessaria rivoluzione digitale, con approfondimenti multidisciplinari a firma di esperti delle materie afferenti all'Agenda Digitale italiana ed europea

Submission e norme editoriali
Per effettuare una submission è necessario concordare prima un argomento e le misure precise contattando info@agendadigitale.eu.
Inviare un abstract di circa 500 caratteri alla testata, presentando l'articolo.
Le misure del testo finale saranno comprese tra 6mila e 20mila caratteri, salvo accordi per misure superiori.
I riferimenti bibliografici dovranno essere preparati in conformità alle regole dell'APA style, 6a edizione (si vedano le linee guida e il tutorial).
Gli autori sono invitati a tener conto degli articoli già pubblicati nella rivista e di citarli nel loro contributo qualora siano ritenuti di interesse per il tema trattato.

Comitato Scientifico e editoriale

Francesco Sacco, Università Bocconi di Milano
Donatella Sciuto, Politecnico di Milano
Nicola Strizzolo, Università di Udine
Luca Gastaldi: eGov, sanità, telecomunicazioni, procurement pubblico, design thinking, Smart Working, Politecnico di Milano
Maurizio Gentile, professore associato, Università di Roma LUMSA
Antonio Ghezzi: strategia, business model, startups, mobile, Politecnico di Milano
Nicola La Sala, registro degli operatori della comunicazione, fattura elettronica, industria4.0, editoria, cittadinanza digitale; Agcom
Emanuele Lettieri, sanità Politecnico di Milano
Maria Beatrice Ligorio, psicologia, università di Bari
Marika Macchi, economia, Unifi
Riccardo Mangiaracina: fatturazione elettronica, eCommerce, logistica e trasporti, export, Politecnico di Milano
Mirco Marchetti, Sicurezza informatica, unimore
Chiara Marzocchi, economia, Università di Manchester
Cristina Masella, Sanità, Politecnico di Milano
Davide Mula, sanità digitale, cyber security, privacy; Agcom
Simone Mulargia, internet and social media studies; Lumsa
Comitato di referaggio
Coordinatore: Luca Gastaldi, Polimi
Mauro Andreolini, sicurezza informatica, Unimore
Luca Baccaro, concorrenza, diritto comunicazioni elettroniche e dei media; studio legale Lipani Catricalà & Partner
Raffaello Balocco, IT e innovazione, Politecnico di Milano
Francesco Capparelli, privacy, cyber security, ecommerce, data management, identità digitale; studio legale ICT Legal Consulting
Ida Cortoni, media education e digital literacy; Dipartimento di Comunicazione e Ricerca Sociale, Sapienza Università di Roma
Giuseppe D'Acquisto, Autorità garante privacy, sicurezza e privacy
Daniela Di Donato, Docente di lettere, Dottoranda di ricerca presso Sapienza Università di Roma-Dipartimento di Psicologia dei processi di sviluppo e socializzazione, Collaboratrice del Crespi
Francesco Di Giorgi, diritto dell'informazione e della comunicazione, tutela dei consumatori, diritto delle comunicazioni elettroniche; Agcom
Leonella Di Mauro, data management, e-commerce, tutela del consumatore, diritto delle comunicazioni elettroniche; Agcom)
Gabriele Ferri, comunicazione e digitale, università Milano Bicocca
Francesco Paoletti, docente di organizzazione aziendale e gestione delle risorse umane, Università degli Studi di Milano-Bicocca
Franco Pizzetti, diritto, privacy, università di Torino
Barbara Quacquarelli, scienze umane e formazione, università Milano Bicocca
Filippo Renga: turismo digitale, smart agrifood, finance and banking, mobile, Politecnico di Milano
Angelo Rovatti, tutela del diritto d'autore, diritti connessi, Diritto dei media; Agcom
Christian Ruggiero, sociologia del giornalismo e comunicazione politica; Dipartimento di Comunicazione e Ricerca Sociale, Sapienza Università di Roma
Franco Torcellan, Laboratorio RED del CISRE – Centro Internazionale di Studi sulla Ricerca Educativa Università Ca' Foscari Venezia
Angela Tumino: Internet of Things, logistica e trasporti, smart city, Politecnico di Milano

Simone Vannuccini, economia, SPRU

Indice del Fascicolo

Intelligenza artificiale, verso una terza dimensione ibrida "uomo-macchina": la sfida (morale) da vincere

Gli sviluppi dell'intelligenza artificiale suscitano attenzione e ottimismo, ma anche dubbi e incertezze. Il futuro che ci attende sarà probabilmente caratterizzato da una **terza dimensione ibrida** in cui uomo e macchina saranno in simbiosi in un territorio in cui tutto è sospeso. Un po' come previsto da Orwell e in Matrix

Di **Fabio De Felice**, Università degli Studi di Cassino e del Lazio Meridionale e **Antonella Petrillo**, Università degli Studi di Napoli "Parthenope"

Nel 2014 il fisico Stephen Hawking in un suo articolo pubblicato sul quotidiano britannico The Independent lanciò l'allarme sui pericoli legati al rapido progresso dell'intelligenza artificiale. Ma già nel 1945 lo scrittore George Orwell sosteneva che sebbene la società di allora non era ostile all'idea di innovazione in quanto tale rigettava le innovazioni nocive per la natura e l'umanità. Pertanto, l'uomo poteva usare con discernimento i prodotti della tecnologia e dell'industria innovativa applicando a tutti lo stesso criterio: ciò mi rende più umano o meno umano? Oggi, la straordinaria potenza dei computer sommata ai progressi sull'intelligenza artificiale e ad una maggiore comprensione del cervello umano stanno suscitando molta attenzione ed ottimismo. È pertanto lecito immaginare una società delle "mangrovie", ibrida, in cui i livelli fisico biologico e digitale trovano un equilibrio, un trade off possibile?

Orwell, profezia e distopia

George Orwell viene universalmente riconosciuto come uno dei più importanti autori della letteratura **distopica**, in cui alcune tendenze del tempo contemporaneo, percepite come causa di conseguenze negative, lasciano presagire un futuro a tinte fosche. Dove prevalgono esperienze del vivere quotidiano opprimente, talvolta spaventoso, frutto di condizioni ambientali fuori controllo, tali da compromettere la **sopravvivenza biologica** di numerose specie, inclusa la specie uomo. E soprattutto di **scelte tecnologiche** che, superando la soglia della dismisura, si avviano agli effetti più funesti. **George Orwell**, oltre a lasciarci romanzi come "1984" e "La fattoria degli animali", ha svolto una consistente attività di giornalista e saggista. Ed è in questo contesto che gli scenari profetizzati nei suoi libri si dispiegano in una versione problematica in cui le domande ricevono embrioni di risposta, individuano prospettive suggestive per i giorni nostri, giungendo al cuore delle questioni aperte di oggi. Una per tutte, la "rivoluzione digitale" al nostro cospetto, vista come arma a doppio taglio, perché conferisce una poderosa spinta alla transizione verso il futuro (industriale, economica, sociale e politica), ma al tempo stesso rende possibile livelli pervasivi di sorveglianza e controllo dei cittadini fin dentro la dimensione più privata della sfera intima. Gettando le premesse anche per una **ibridizzazione delle persone**, sempre più coinvolte da funzioni e servizi ad alto contenuto di software digitale tale da supporre un influsso della realtà

artificiale sull'apparato percettivo e cognitivo dell'uomo: il sistema tecnologico che finisce col tradursi in leva per un **salto** persino nella evoluzione della specie umana, inteso quasi come **cambio genetico** del suo cromosoma ricombinante.

Più umano, meno umano

E tuttavia si deve proprio ad Orwell un posizionamento favorevole nei riguardi del **concetto** di **innovazione**. Nella sua produzione saggistica è infatti rintracciabile un recupero **della centralità (e responsabilità) dell'uomo** al quale è restituito il compito di discernere quanto di nocivo, per la natura e per l'umanità, emerge dal progresso tecnologico. Suo è l'interrogativo categorico, valido come cartina al tornasole dell'etica della scienza. La seguente domanda fatale: ciò (che sperimento e che concepisco come risultante di un procedimento scientifico) … mi rende più umano o meno umano? Vale a dire, questa o quella soluzione tecnologica consente all'uomo di dispiegare le sue facoltà o le riduce e le comprime? **Domanda fondamentale**, che precede e pone un limite a quella posta alla base del moderno nel mondo occidentale: soddisfa o non soddisfa i miei bisogni? Che cosa direbbe mai, se potesse, degli interrogativi dei nostri giorni, dinanzi alla prospettiva, sempre più incalzante, di sostituire l'uomo con un **assistente virtuale**, una forma di **intelligenza artificiale** forse ancora primitiva, o di consentirgli di vivere in una dimensione virtuale mediante l'estensione delle sue facoltà cognitive ed emotive in un avatar? Fino al punto di immaginare un punto di non ritorno, ossia il momento in cui il primato delle macchine sull'uomo diviene possibile.

Matrix: là dove l'uomo diviene una neuro-simulazione interattiva

Cinquant'anni dopo Orwell, è il cinema che si incarica di raccogliere il testimone delle visioni di futuro e delle profezie escatologiche dalle mani del romanzo distopico. **Matrix**, ad esempio, è un film del 1999 insignito di 4 Oscar in cui, declassato perché incapace di sviluppare, come ogni altra specie di mammiferi fa "per istinto", una felice omeostasi con l'ambiente circostante, l'**uomo** rappresenta il **vero pericolo** per una intelligenza artificiale in grado di ricavare giudizi per inferenza, capace di cogliere e valutare le conseguenze, per la vita sulla Terra. Scritto e diretto dalle sorelle Lana e Lilly Wachowski, il titolo richiama la "**matrice di numeri**", elemento che deriva da strutture matematiche utilizzato in informatica per associare dati, o sistemi di dati, tra loro. Ed ecco le parole del monologo dell'agente Smith, esponente del sistema Matrix, espresse quando interroga Morpheus, ufficiale della resistenza di Sion, dopo averlo catturato.

"… Improvvisamente ho capito che voi non siete dei veri mammiferi: tutti i mammiferi di questo pianeta d'istinto sviluppano un naturale equilibrio con l'ambiente circostante, cosa che voi umani non fate. Vi insediate in una zona e vi moltiplicate, vi moltiplicate finché ogni risorsa naturale non si esaurisce. E l'unico modo in cui sapete sopravvivere è quello di spostarvi in un'altra zona ricca. C'è un altro organismo su questo pianeta che adotta lo stesso comportamento, e sai qual è? Il virus. …siete una piaga".

Degradato ad agente infetto, pertanto totalmente indesiderabile, per l'uomo non resta altra possibilità di redenzione che **farsi guidare da un algoritmo**.

Intelligenza artificiale o primordiale?

Forse il **mix di intelligenza artificiale** e **pensiero cognitivo** umano offre la soluzione a un problema che l'umanità non riesce ad affrontare e risolvere. Il salto può essere fornito dalla **ibridazione** tra realtà umano e realtà artificiale? Di "macchine pensanti" si parla in effetti fin dalla prima metà del secolo scorso, ossia i tempi di **Alan Turing**, matematico e crittografo britannico considerato uno dei padri dell'informatica. Nel frattempo, le **neuroscienze** hanno seguito il proprio percorso di ricerca, giungendo a definire l'intelligenza come un insieme di facoltà differenti, che concorrono in varia misura a permettere di "leggere dentro" le cose e le persone, vale a dire **capire-percepire-conoscere**.

Secondo la **teoria monofattoriale**, l'intelligenza sarebbe genericamente "il risultato finale della collaborazione tra diverse funzioni cerebrali". La teoria **multifattoriale** invece sostiene che l'intelligenza è composta da capacità differenti e indipendenti, ognuna destinata a compiti diversi. Si parla infatti, in questo ambito, più opportunamente di "intelligenze" che diversamente sviluppate in ogni individuo segnano le caratteristiche distintive tra le persone.

Alcuni esponenti delle **neuroscienze** sono giunti al convincimento che la capacità di "intelligere" è proporzionata alla capacità di elaborare pensieri, ragionamenti e contenuti concettuali a partire dalle emozioni. Ed è qui che le macchine segnano il passo. Infatti, esse sono capaci di **elaborare le informazioni** in quantità e velocità senza precedenti, realizzando associazioni osservate dalle differenti forme, colori, volumi, proporzioni degli oggetti (basti pensare al riconoscimento facciale o alle soluzioni del deep learning, ossia apprendimento profondo dell'intelligenza artificiale che si basa su diversi livelli di rappresentazione delle reti neurali. Ma sono meno attrezzate per "inferire" una serie di attributi di un elemento anche se nel momento dato non si manifestano e non potendoli percepire direttamente. Ma ancora più lontane sono le macchine dalla capacità di svolgere un **ragionamento controfattuale**, del tipo "se fosse successo A, allora sarebbe accaduto B". L'essere umano fa ragionamenti del genere in continuazione, non potendo usare sempre la logica delle implicazioni fattuali, le quali ci dicono che se accade una cosa, allora ne accade un'altra (se è vero l'antecedente, è vero anche il conseguente).

Conclusioni

Non conosciamo il futuro, non sappiamo come andrà a finire. Sappiamo però che senza ombra di dubbio stiamo vivendo un periodo in cui le rivoluzioni nelle biotecnologiche e nelle tecnologie informatiche stanno mettendo l'umanità di fronte a **prove difficili** forse le più difficili in cui si sia mai imbattuta. Assistiamo ad un **cambio di paradigma** epocale nell'interazione tra uomo e macchina. Tanto più riusciremo a comprendere i meccanismi biochimici che controllano le dinamiche emotive umane tanto più i computer potranno diventare abili nell'analizzare il comportamento umano. La domanda orwelliana, pertanto, ritorna e torna l'eco del suo scetticismo. Lo sviluppo più probabile dell'impatto dell'intelligenza artificiale sulle attività antropiche (o, meglio: dell'intelligenza umana su quella artificiale) sarà probabilmente una **intersezione** tra i due insiemi: **uomo & macchina**. Vale a dire una **terza dimensione ibrida** in cui i livelli fisico-biologico e digitale saranno in simbiosi in un territorio che più che un limbo dantesco in cui tutto è sospeso e in attesa somiglierà all'acqua salmastra, che si forma nel punto in cui i fiumi incontrano il mare, dove pullulano le **mangrovie**, come afferma con eccellente metafora **Luciano Floridi**. Ma esistono "le foreste a mangrovia" che popolano il tratto di territorio condiviso tra terra e mare, comportandosi in maniera differente rispetto all'influsso delle maree. Analogamente l'uomo dinanzi alle ondate di intelligenza artificiale. Le ondate di AI stanno rimodellando il mondo in cui viviamo, che diverrà radicalmente diverso da quello che conosciamo.

È un bene? È un male? Dipende da noi, l'ultima sfida non è tecnologica, ma **morale**. Ancora una volta, torna la domanda cruciale: "Ciò (che sto facendo, che stiamo producendo) mi rende più umano o meno umano?".

Bibliografia

Fabio De Felice, Antonella Petrillo. Effetto digitale. Visioni d'impresa e Industria 5.0. McGraw-Hill Education, 2021.

Daniel Goleman. Lavorare con intelligenza emotiva. Come inventare un nuovo rapporto con il lavoro. BUR Biblioteca Univ. Rizzoli, 2020.

Stefano Quintarelli. Intelligenza artificiale. Cos'è davvero, come funziona, che effetti avrà. Bollati Boringhieri Editore, 2020.

Luciano Floridi. La quarta rivoluzione. Come l'infosfera sta trasformando il mondo. Editore Cortina Raffaello, 2017.

George Orwell. 1984. Mondadori, 2016.

Michael Paradiso, Mark F. Bear, Barry W. Connors. Neuroscienze. Esplorando il cervello. Edra Editore, 2016.

Intelligenza artificiale, obiettivo regole privacy per renderla "umana"

Arriva il Regolamento Ue sull'intelligenza artificiale. Il focus è sulla protezione dei dati, che può garantire la creazione di macchine che somiglino sempre di più all'uomo, che si avvicinino al suo modo di ragionare, senza coup de théâtre non comprensibili dalla ragione umana. Le sfide

Giuseppe D'Acquisto, Funzionario del Garante per la protezione dei dati personali - Titolare dell'insegnamento di intelligenza artificiale presso il Dipartimento di Giurisprudenza dell'Università LUISS Guido Carli

È attesa a ore la pubblicazione da parte della Commissione Europea del draft di Regolamento sull'**intelligenza artificiale**, che è stato preceduto da un *corpus* di lavori preparatori nell'ultimo biennio sui temi dell'etica[1], della responsabilità (liability)[2] e della sicurezza[3].

Dalla lettura di questa documentazione possiamo comprendere tutte le **aspettative** e le **preoccupazioni** che hanno alimentato questo dibattito, e provare a intuire, almeno nelle linee generali, gli obiettivi regolatori che stanno ispirando le scelte della Commissione.

Le aspettative si comprendono tutte se osserviamo le **previsioni economiche**, ad esempio quelle fatte dal Parlamento UE nel 2020[4], che stimano per il 2025 un impatto dell'intelligenza artificiale sulla produttività globale intorno a 10.000 miliardi di euro (circa un decimo del PIL mondiale). Si tratta dunque di un'area di grande importanza strategica e di un fattore chiave per la crescita economica, anche in vista di un rilancio post-Covid. Le preoccupazioni invece riguardano gli impatti sul tipo **di relazioni tra uomini e macchine**, e tra uomini e uomini, in una società in cui macchine sempre più intelligenti avranno un ruolo crescente nei processi decisionali ad ogni livello.

È dunque importante costruire un **sistema di regole** in grado di promuovere lo sviluppo, preservare la centralità della persona e allo stesso tempo creare fiducia sull'impiego dell'intelligenza artificiale.

Il concetto di autonomia decisionale della macchina

In questo difficile esercizio il concetto di **autonomia decisionale della macchina**, che è il tratto distintivo dell'intelligenza artificiale rilevato dalla Commissione Europea[5], è cruciale. Se abbiamo l'ambizione di regolamentare l'intelligenza artificiale, dobbiamo infatti comprendere bene in cosa si esplica **l'autonomia** della macchina, il suo campo di applicazione, e il tipo di regole che possono essere immaginate per regolamentarne il funzionamento. Individuare regole che non confliggano con il modo di funzionare di macchine autonome è un passaggio fondamentale se si vuole garantire effettività delle tutele.

È bene osservare che l'autonomia decisionale della macchina non è legata al concetto di **necessità** della macchina per il conseguimento di un risultato, e neppure alla complessità del risultato stesso. A ben riflettere, ci sono molte situazioni (e, in vero, ve ne sono da diversi decenni ormai) in

cui l'impiego di uno strumento di calcolo automatico è necessario per raggiungere un risultato, eppure l'azione dello strumento non può essere qualificata come autonoma. Pensiamo a progetti molto articolati, con calcoli molto complessi. In tutti questi progetti c'è un momento a partire dal quale la capacità computazionale dell'uomo si arresta e lascia il passo all'intervento della macchina che esegue tutti i calcoli al posto dell'uomo. Però non dobbiamo confondere questa necessità legata alla complessità computazionale con il concetto di autonomia decisionale. Se il risultato è corretto grazie ai calcoli eseguiti dalla macchina non posso dire che questo è il risultato dell'autonomia della macchina. E, per contro, se il risultato è sbagliato per un errore di calcolo non posso affermare che l'uomo non sia responsabile dell'errore perché i calcoli sono stati eseguiti da una macchina.

L'uomo è sempre pienamente responsabile, anche se è la macchina ad arrivare al risultato, corretto o sbagliato che sia, in quanto è l'uomo che sceglie i criteri di progetto e i modelli matematici, ovvero gli algoritmi, che descrivono il problema che si vuole risolvere. Fintanto che l'uomo è in grado di spiegare con una teoria il funzionamento di un algoritmo non siamo in presenza di una decisione autonoma, e possiamo ricondurre il risultato ottenuto dalla macchina interamente alla responsabilità dell'uomo. Dunque, se c'è una teoria sviluppata dall'uomo non c'è autonomia decisionale da parte della macchina. La presenza di una teoria è una forte tutela, dal momento che rende possibile una piena **trasparenza** sul processo decisionale e una piena accountability sui risultati. Non è di questo che parliamo quando ci riferiamo all'autonomia decisionale dell'intelligenza artificiale.

I conflitti tra regolamentazione e funzionamento delle tecnologie

I conflitti tra regolamentazione e funzionamento delle tecnologie affiorano quando l'azione della macchina diventa realmente autonoma, ossia quando **l'intervento della macchina non è sorretto da una teoria** che spieghi i fenomeni. Questo è il campo di applicazione dell'intelligenza artificiale, già oggi molto ampio e certamente destinato a crescere in futuro: l'impiego di strumenti di calcolo in tutte quelle situazioni in cui manca del tutto, o manca ancora, la spiegazione del fenomeno, ma vogliamo comunque conseguire il risultato.

Questo obiettivo con l'intelligenza artificiale è in effetti raggiungibile: possiamo ottenere risultati in situazioni nelle quali non abbiamo ancora una scienza consolidata, o nelle quali il livello di complessità dei fenomeni è tale da rendere praticamente impossibile avere una teoria sullo sfondo degli algoritmi.

Questo non significa che con l'intelligenza artificiale entriamo in una sfera di decisioni illogiche e ci consegniamo all'arbitrio della macchina. Però dobbiamo essere consapevoli che con l'intelligenza artificiale assistiamo a **un cambiamento profondo del paradigma conoscitivo**. Da una parte abbiamo macchine non autonome che funzionano sulla base di una teoria sviluppata dall'uomo, dall'altra abbiamo macchine autonome che funzionano unicamente sulla base dei dati che ricevono in ingresso e di un solo paradigma conoscitivo di tipo quantitativo: la ricerca di correlazioni tra dati. Sono due modi razionali di procedere, ma incommensurabili: uno a misura d'uomo, l'altro a misura di macchina.

Autonomia decisionale e correlazione tra dati

Questa è la principale novità introdotta dal concetto di autonomia decisionale applicato alle macchine, ed esistono moltissime situazioni in cui i dati sono presenti in abbondanza ma non disponiamo affatto, o non disponiamo ancora, di una teoria che spieghi i fenomeni che hanno generato quei dati. La medicina digitale e la **diagnostica per immagini**, per esempio, sono interamente basate sulle correlazioni tra dati, così come l'interpretazione della voce alla base del funzionamento dei tanti **assistenti domotici** che usiamo ogni giorno, o ancora il riconoscimento facciale impiegato nella **videosorveglianza**, e l'individuazione di anomalie per la prevenzione delle frodi nei sistemi di pagamento online.

Sono tutte situazioni nelle quali abbiamo moltissimi dati, non abbiamo una teoria, ed è improbabile che ne avremo per via dell'ampia variabilità dei casi, ma vogliamo ottenere dei risultati. In questi e tanti altri casi la macchina può intervenire in piena autonomia cercando **correlazioni tra i dati** e offrendoci dei risultati che con l'approccio teorico umano non riusciremmo ad avere. Solo la macchina può arrivare a quei risultati. L'uomo non può arrivarci, o forse ci arriverebbe troppo tardi. E **non si tratta di un approccio irrazionale**: questi risultati infatti sono nella maggior parte dei casi corretti, ovvero utili, cioè diagnosticano la patologia, riconoscono il significato della parola, riducono i rischi, ma non spiegano il fenomeno. O meglio ci danno per i diversi fenomeni una sola spiegazione universale: c'era una correlazione tra i dati.

Grazie all'autonomia decisionale delle macchine abbiamo dunque la prospettiva di poter ottenere qualsiasi tipo di risultato che ci è utile per ogni possibile scopo partendo dalla semplice osservazione di dati e senza la necessità di sviluppare una teoria.

Lo slancio creativo che manca all'intelligenza artificiale

Ma questa possibilità ha un prezzo molto elevato. A fare le spese è infatti la capacità di **spiegare i fenomeni attraverso un ragionamento causale**. E non è poco. L'intero metodo scientifico è infatti fondato sulla ricerca di un nesso di causalità che cerchi di interpretare con congetture sempre falsificabili, come ha spiegato benissimo Karl Popper[6], la ragione dei fenomeni.

Se l'uomo, grazie alle sue teorie, osserva il sorgere del sole e sente il canto del gallo non ha nessun dubbio su quale fenomeno sia la causa e quale l'effetto. Se la macchina autonoma, in assenza di teorie, osserva oggi il sorgere del sole e sente oggi il canto del gallo, domani sarà certa che uno dei due fenomeni annuncia certamente l'altro, cioè saprà fare previsioni, ma non sarà in grado di dirci quel dei due fenomeni determina l'altro, cioè non saprà fare scoperte. Bisogna non sottovalutare questa limitazione.

Un paradigma conoscitivo guidato da macchine che agiscono senza intervento dell'uomo e interamente basato sulla ricerca delle correlazioni scandaglia tutti i dati esistenti e ci mostra legami tra fenomeni assolutamente invisibili all'uomo, consentendo di ottenere risultati impensabili, ma in definitiva lascia il mondo invariato, **venendo a mancare ogni slancio creativo**. Anche l'approccio creativo umano alla decisione, a dire il vero, non è privo di rischi, ma esistono molti meccanismi di tutela oltre alla sussistenza di una teoria spiegabile, quali l'assunzione di precise **responsabilità sui risultati**, la possibilità di ricorrere in giudizio. Un analogo insieme di garanzie per le decisioni assunte da una macchina autonoma non esiste ancora ed è tutto da

costruire. **Questo il fenomeno tecnologico che vogliamo regolamentare quando parliamo di intelligenza artificiale.**

Protezione dei dati perno della regolamentazione sull'AI: il concetto di privacy by design

In un contesto come questo, nel quale la ricerca di correlazioni tra dati è il presupposto unico per le decisioni, la protezione dei dati riveste un ruolo estremamente importante come perno della regolamentazione. Infatti, se nella fase in cui la macchina agisce in autonomia il margine per l'intervento regolatorio dell'uomo è limitatissimo, e forse inesistente, lo spazio di regolamentazione si sposta tutto sulla fase immediatamente precedente alla decisione, ossia quella della **raccolta e della strutturazione del dato**, e sulla fase successiva, ossia quella delle conseguenze sulla persona.

Dunque, prima che la macchina agisca, è necessario rafforzare le tutele intervenendo direttamente sul dato, che deve risultare ben protetto in modo da non indurre risultati sbagliati. In altri termini, condizione necessaria, ma non sufficiente, per potersi fidare dei risultati di una macchina è che i dati abbiano **un livello di qualità adeguato** allo scopo per cui il risultato che da essi verrà desunto sarà impiegato.

L'importanza dei dati di partenza

Non si può pensare di avere risultati accurati se i dati di partenza non sono accurati. La qualità del dato si può solo disperdere, non si crea cammin facendo. Si tratta di **un obiettivo molto complesso da raggiungere**, che richiede considerazioni a più livelli e che non può essere raggiunto unilateralmente.

Qui si possono solo richiamare **alcune questioni che devono essere affrontate per avvicinarsi a questo obiettivo**.

Cosa è una fonte di dati attendibile?

Chi ne certifica la qualità per lo scopo per cui i dati saranno impiegati?

Come verificare che i dati siano utilizzati per ottenere risultati non discriminatori e per non avvantaggiare un soggetto a discapito di un altro?

Come organizzare, su un piano più tecnico, un processo decisionale automatizzato in modo da non disperdere la qualità del dato in trattamenti che coinvolgono più soggetti?

Come evitare che dal trattamento emergano più risultati di quelli strettamente necessari?

A ben vedere sono modi diversi per ribadire i principi (di accuratezza, correttezza, minimizzazione, necessità) della protezione di dati personali.

Privacy by design

Nel contesto dell'intelligenza artificiale essi dovranno essere declinati in una forma tecnologica adeguata a essere compresa dalle macchine. È il concetto di **privacy by design**, la cui essenza è proprio l'integrazione dei principi per via tecnologica all'interno dei trattamenti. Inoltre, poiché il dato determina le sfumature del risultato, e nelle sfumature risiede la tutela, il dato prima di essere trattato deve assumere tutte le possibili sfaccettature per consentire risultati calibrati. Oggi il dato può essere pseudonimizzato, generalizzato, randomizzato, cifrato (e addirittura cifrato in modo omomorfo), e infine anonimizzato. Sono tutte forme nuove che i dati possono assumere e che stiamo imparando a conoscere, ognuna delle quali incorpora già una precisa forma di tutela. Occorre, e lo si sta facendo, comprendere cosa significa impiegare questi tipi di dati nel contesto dell'intelligenza artificiale. C'è un promettente settore che studia l'impiego di tecniche di anonimizzazione nell'ambito del machine learning. Questi studi vanno seguiti molto attentamente[7]. Magari potremmo scoprire che attraverso l'impiego di un certo tipo di dati tra quelli richiamati si può mantenere l'efficacia dei risultati riducendo però significativamente gli impatti sui diritti e le libertà della persona. Quanto più cresce l'autonomia decisionale della macchina tanto maggiore è dunque la necessità di rafforzare la protezione intervenendo direttamente sul dato.

I meccanismi di tutela ex-post

Sull'impatto che deriva dal risultato, il rischio connesso alle conseguenze di ogni tipo di decisione è ineliminabile, anche se è l'uomo ad assumerla, e con uno spirito privo di preconcetti dobbiamo riconoscere che l'uomo con le sue decisioni è spesso un generatore di disuguagliane e discriminazioni. Si parla molto di intervento umano come elemento di salvaguardia rispetto alle decisioni automatizzate, ma questo ci indurrebbe a domandarci a quale archetipo ideale di uomo ci si riferisca. Più pragmaticamente, piuttosto, è necessario, di pari passo con l'applicazione sempre più diffusa di decisioni assunte dalle macchine in autonomia, prevedere meccanismi di tutela ex-post che riducano i rischi che derivano da risultati già ottenuti, e che per loro natura sono incontrollabili.

Anche qui il **GDPR** ci dà ottimi spunti di riflessione da cui vale la pena partire. Esiste l'istituto della **notifica dei data breach**, che è proprio una misura ex-post di mitigazione dei rischi a incidente avvenuto. In ugual modo, anche per le decisioni già assunte dalle macchine in autonomia dovrebbe essere lasciata alla persona la facoltà di notificare situazioni penalizzanti a un soggetto *super partes*, con adeguati poteri di intervento tecnico e giuridico capaci di invalidare la decisione della macchina. Sarebbe una forma di notifica distribuita rispetto a quella prevista dal GDPR, oggi limitata ai titolari, che naturalmente richiederebbe ragionati interventi regolatori di incentivazione e ingegnerizzazione, che sarebbe effettuata nell'interesse collettivo di limitare, anche se soltanto ex-post, l'asimmetria ineliminabile tra uomo e macchina sul modo in cui dai dati si perviene ai risultati.

Il "senso" da attribuire ai risultati dell'AI

Ma oltre a questi interventi di carattere procedurale è sicuramente necessaria, e forse decisiva, una riflessione più generale, di carattere culturale, sul "senso" che si può attribuire ai risultati prodotti dall'intelligenza artificiale. L'assenza di una teoria spiegabile nei termini causali tipici del ragionamento umano dovrebbe farci sempre assumere un **atteggiamento critico** nei confronti dei risultati prodotti da una macchina che agisce in autonomia.

A metterci in guardia è la stessa comunità scientifica che oggi studia le possibili evoluzioni dell'intelligenza artificiale. Nel mondo della ricerca c'è un detto: "*artificial intelligence is doing more with less*".

È proprio così: l'intelligenza artificiale consente di fare un passo in più di più rispetto all'osservazione dei dati, ma bisogna stare attenti sulla valenza conoscitiva di questo passo ulteriore. Perché ci sia il "più" è sempre necessario il "meno", ossia la creatività umana che ha generato i dati osservati che la macchina riceve in ingresso, e questo "meno" è un grande **margine di manovra** che l'uomo ha per indirizzare l'azione autonoma della macchina.

Noi spesso ci soffermiamo sul "più", perché lì si concentra l'aspetto teatrale dell'intelligenza artificiale, ma il risultato del "più" molto spesso è già iscritto ed evidente nel "meno". Se il "meno", ossia il primo passo verso l'interpretazione dei fenomeni compiuto dall'uomo, è di qualità, allora il "più" ottenuto in autonomia dalla macchina diventa un vero valore aggiunto sulla conoscenza del mondo. Se noi ci concentriamo troppo sul "più" rischiamo di lasciarci abbagliare dall'effetto, e se divarichiamo la distanza tra il "più" e il "meno" allora il risultato diventa un fake sganciato dalla realtà, un **abbaglio** che non genera fiducia sull'impiego della macchina, e finiamo per consegnarci all'arbitrio della macchina.

Non è un caso che la frontiera della ricerca scientifica oggi sia maggiormente concentrata sul "meno", ossia sui **fondamenti causali del ragionamento** per verificare se sia possibile ipotizzare una macchina creativa[8]. La ricerca più avanzata sembra dirci che c'è un limite oltre il quale l'intelligenza artificiale rischia di diventare maniera, un puro esercizio di stile che potrebbe non trovare il favore collettivo e la necessaria fiducia sui risultati.

Conclusioni

Siamo dunque in presenza di un settore strategico, in grandissimo fermento sul piano della ricerca scientifica, ma anche desideroso di buone regole. Grande e motivata è dunque l'attesa per il Regolamento che la Commissione sta per presentare.

L'analisi del fenomeno che è stata qui condotta dal punto di vista della persona ci porta a concludere che ciò che occorre non sono macchine che puntino su *coup de théâtre* non comprensibili dalla ragione umana, ma **macchine che somiglino sempre di più all'uomo**, che si avvicinino al suo modo di ragionare.

E una protezione dei dati che punti sui principi e che sappia **guardare negli occhi l'intelligenza artificiale** offrendo soluzioni tecnologiche per mantenere **al centro la persona** anche in presenza di macchine con capacità di decisione autonoma è assolutamente indispensabile per costruire uno spirito di **fiducia e di ottimismo** nei confronti di questi strumenti che, se saggiamente impiegati, possono costituire un reale fattore di progresso e di crescita per l'umanità.

Bibliografia

High-Level Expert Group on Artificial Intelligence set up by the European Commission: The Assessment List For Trustworthy Artificial Intelligence (ALTAI), July 2020 ↑

European Parliament resolution of 20 October 2020 with recommendations to the Commission on a civil liability regime for artificial intelligence ↑

ENISA AI Threat Landscape, 15 December 2020 ↑

https://www.europarl.europa.eu/news/en/headlines/society/20201015STO89417/ai-rules-what-the-european-parliament-wants ↑

A European approach to Artificial intelligence, European Commission 9 March 2021, https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence ↑

K. Popper, *Congetture e confutazioni, Lo sviluppo della conoscenza scientifica*, Il Mulino, 2009. ↑

E. De Cristofaro *An Overview of Privacy in Machine Learning*, ArXiv:2005.08679, 18 May 2020. http://arxiv.org/abs/2005.08679 ↑

Si vedano J. Pearl, D. Mackenzie, *The Book of Why: The New Science of Cause and Effect*, Penguin 2019 (Judea Pearl è il vincitore del premio Turing, l'equivalente del premio Nobel dell'informatica, nel 2011) e B.A Richards, T.P Lillicrap, P Beaudoin, Y Bengio et alii, *A deep learning framework for neuroscience*, Nature neuroscience, 2019 (Yoshua Bengio è invece il vincitore del premio Turing nel 2018). ↑

L’errore umano dell’intelligenza artificiale: ecco perché dobbiamo imparare a conviverci

Gli algoritmi e i sistemi di IA sono fatti dall'uomo, e saranno sempre modellati dai nostri valori culturali e condizioni tecniche e sociali che li hanno creati. Invece di cercare di risolvere il bias dei sistemi IA e il loro errore umano, dobbiamo trovare il modo di coesistere con esso

di **Veronica Barassi**, Professor in Media and Communications Studies in the School of Humanities and Social Sciences at the University of St.Gallen in Switzerland

Nel gennaio 2020, **Robert Julian-Borchak**, un uomo di colore, è stato arrestato a Detroit per un crimine che non aveva commesso e soltanto per un errore di un sistema IA di riconoscimento facciale. Questo, secondo il New York Times, è stato il primo caso conosciuto di un cittadino americano arrestato per un errore dell’algoritmo.

Il caso di Robert Julian-Borchak racconta la storia dei nostri tempi. Oggi, sempre di più, sistemi di **intelligenza artificiale** vengono utilizzati da forze dell’ordine e tribunali per monitorarci, tracciarci e profilarci in modo da determinare la nostra possibile innocenza o colpevolezza. Queste trasformazioni non riguardano solo gli Stati Uniti, ma anche in Europa la logica algoritmica è sempre più diffusa nei **contesti di giustizia penale** [1] o **prevenzione del crimine** [2].

L’uso di queste tecnologie da parte di tribunali e forze dell’ordine può avere un **impatto deleterio sui nostri diritti e le nostre libertà**, come dimostra l’esempio di Robert Julian-Borchak, ed è per questo che questi sistemi devono essere capiti, studiati, collaudati e a volte proibiti.

In aprile 2021, la Commissione europea ha pubblicato la nuova Proposta di un regolamento sull’Intelligenza Artificiale in Europa, che cerca di far proprio questo. La nuova proposta stipula che i sistemi IA che vengono utilizzati per profilare gli individui - da parte delle forze dell'ordine, educatori, risorse umane ecc. – siano considerati **ad alto rischio**, e suggerisce la proibizione in Europa di pratiche come il social scoring o la sorveglianza biometrica "in tempo reale". Il regolamento prevede anche **maggiore responsabilità** sull’uso di tutti i sistemi IA ad alto rischio e trasparenza sulle banche dati utilizzate per il loro training. Ciò che emerge chiaramente dalla proposta della Commissione UE è che i sistemi di IA volti alla profilazione umana sono esposti a diversi tipi di errori e bias impliciti (pregiudizi) e quindi possono amplificare le disuguaglianze nella nostra società e possono avere un impatto negativo sui diritti umani.

La proposta della Commissione Europea per un regolamento sull'intelligenza artificiale è - a mio parere - **un passo avanti molto importante**. Tuttavia, credo anche che dobbiamo essere realistici, imparare dall'esperienza del **GDPR** e ricordare che l'implementazione di proposte legislative come questa in Europa non è affatto facile. La verità è che la nostra vita quotidiana è ormai definita da

una quantità incredibile di decisioni algoritmiche, che non sono facilmente controllabili o governabili.

La storia di Robert Julian-Borchak sembra **assurda, distopica, e kafkiana**. È un caso estremo. Nella sua estremità, però, ci mette di fronte ad una storia molto più banale, molto più ordinaria: **ogni giorno veniamo giudicati da sistemi IA** e ogni giorno i loro errori, i loro bias possono limitare le nostre libertà.

Vite intere giudicate dagli algoritmi?

I cambiamenti tecnologici degli ultimi anni, le innovazioni nel campo dei big data e intelligenza artificiale, ci hanno portato ad una situazione storica in cui i nostri dati personali, vengono utilizzati per **decidere aspetti fondamentali** della nostra vita quotidiana.

Quando cerchiamo un lavoro, stipuliamo una assicurazione, chiediamo un finanziamento, iscriviamo i nostri figli a scuola e in altre innumerevoli situazioni questi dati - decontestualizzati, sterilizzati e confrontati con benchmark e algoritmi standard - vengono usati per giudicarci. Per anni ho studiato questa trasformazione concentrandomi soprattutto sulla datificazione dei cittadini da prima della nascita e la racconto nel mio ultimo libro pubblicato da MIT Press e intitolato Child | Data | Citizen: How Tech Companies are Profiling Us from before birth (2020) e anche nel mio nuovo libro – intitolato **I Figli dell'Algoritmo** - che verrà pubblicato in Italia da LUISS University Press, quest'anno.

La mia ricerca mi ha portato alla conclusione che per la prima volta nella storia dell'uomo - stiamo creando una generazione di **cittadini che vengono datificati dalla nascita**. Dal momento in cui i bambini vengono concepiti, le loro informazioni mediche vengono spesso condivise sulle App di gravidanza o sui social media. Nella vita di tutti i giorni i loro dati personali vengono raccolti e archiviati da tecnologie domestiche e assistenti virtuali nelle loro case, da piattaforme educative nelle scuole, da documenti e portali online presso lo studio del medico, dai loro giocattoli connessi a Internet, dai loro giochi online e da molte, molte, molte altre tecnologie. Tutti questi dati personali vengono aggregati, scambiati, venduti e trasformati in profili digitali che possono seguirli per una vita.

Una delle idee principali della mia ricerca è la realizzazione che al giorno d'oggi non c'è più confine tra i dati del consumatore, che vengono raccolti per proporre pubblicità più mirate, e i dati del cittadino, che vengono raccolti per decidere se possiamo avere accesso o meno a determinati diritti (Barassi, 2020). Un esempio chiave che racconto in I Figli dell'Algoritmo viene dagli Stati Uniti.

Il 26 febbraio del 2021, il Washington Post ha riportato la notizia che ICE (il corpo di polizia che controlla l'immigrazione negli Stati Uniti) ha avuto accesso ad un database che si chiama CLEAR – di Thomson Reuters. Il database include più di **400 milioni di nomi**, indirizzi e registri di utenti creati con i dati dei consumatori raccolti da più di 80 società che si occupano delle bollette dell'acqua, gas, elettricità, telefono, Internet e TV via cavo. Quando sono andata sul sito di CLEAR ho scoperto che il database è usato per diversi tipi di investigazioni governative e istituzionali che vanno ben oltre le violazioni dell'immigrazione. Infatti, il database viene usato per combattere la frode fiscale, sanitaria, il riciclo di denaro oppure per prendere decisioni riguardo l'affidamento dei bambini.

Senza che gli utenti lo sappiano, i loro dati domestici vengono incrociati e condivisi, e vengono usati per decidere i loro diritti.

Stiamo vivendo una nuova realtà dove - sempre di più- le nostre informazioni che accettiamo di offrire come consumatori, vengono usate da sistemi IA di analisi predittiva e profilazione per determinare i nostri diritti umani. A differenza di altri paesi - come l'America e la Cina - **l'Europa e l'Italia offrono maggiori protezioni** per quanto riguarda l'uso dei nostri dati personali e la privacy. Eppure, anche in Italia e in Europa stiamo confrontandoci con una rapida digitalizzazione delle infrastrutture governative e un aumento dell'analisi predittiva a livello istituzionale. In Italia, un segno di questa trasformazione la troviamo nelle dichiarazioni programmatiche del Presidente Draghi.

La cosa che a me stupisce di più quando penso a queste trasformazioni, è la velocità con cui **sistemi IA volti alla profilazione umana** vengano addottati in diversi contesti, da scuole a ospedali, da forze dell'ordine a infrastrutture governative. Sembra quasi che nell'ultima decade ci siamo davvero convinti che le tecnologie usate per l'analisi predittiva, l'incrocio dei dati, il riconoscimento facciale, la classificazione dell'emozioni e tutte le altre tecnologie che vengono utilizzate da sistemi IA offrano una conoscenza più approfondita sui comportamenti e psicologia umana. La domanda che dobbiamo porci è: possiamo fidarci di queste tecnologie quando si tratta di leggere gli esseri umani? La mia risposta è no.

L'errore umano nei nostri dati

Nel 2018, mi sono seduta in un ristorante affollato a Los Angeles, dove ho incontrato Cara, uno dei genitori che ho intervistato per la ricerca per il mio libro Child Data Citizen. Il ristorante era rumoroso, e rifletteva tutta la vivacità sensoriale a cui eravamo abituati in un mondo pre-Covid 19. Durante l'intervista Cara ha definito le compagnie pubblicitarie su internet come dei "parassiti" e "pettegoloni", e ha spiegato: "quando la gente sembra dedurre qualcosa su di te sulla base di una certa informazione o voce di corridoio, allora è sbagliato, sembra un pettegolezzo. Quando vengo presa di mira per una ricerca che ho fatto su Google mi sento esattamente così; come se qualcuno stesse spettegolando su di me".

Nella sua semplicità, l'analogia di Cara mostra uno degli aspetti più problematici del **profiling digitale**. Ci ricorda che le tecnologie dei dati costruiscono immagini approssimative di chi siamo sulla base di pezzi di informazioni che aggregano e incrociano. Dobbiamo renderci conto che **molte volte c'è una mancanza di legame tra le nostre pratiche digitali e i dati che produciamo**. Le nostre pratiche digitali sono complesse, contraddittorie; riflettono diverse intenzioni, valori e situazioni, e non sono l'espressione di comportamenti precisi. A volte non usiamo le tecnologie come dovremmo, e le usiamo in modo tattico. Per esempio, nelle famiglie in cui ho lavorato, pratiche come l'"autocensura" o il "giocare con l'algoritmo" avevano proprio lo scopo di confondere il tracciamento di dati online.

Gli algoritmi non sono attrezzati per comprendere la complessità umana, e finiscono per fare ipotesi riduzioniste ed errate sull'intenzione dietro una specifica pratica digitale (una ricerca sul web, un acquisto o un post sui social media). Questo perché i dati che vengono raccolti dalle nostre pratiche digitali - e utilizzati per creare i nostri profili digitali - sono spesso privati dei sentimenti, dei pensieri e del contesto che li ha prodotti. Per fare luce su questo problema, nel 2018 ho scritto un rapporto firmato da Gus Hosein, CEO di Privacy International e supportato da

Jeff Chester, direttore del Centre For Digital Democracy di Washington DC, dove parlo della **complessità antropologica dei dati che vengono raccolti dalle nostre case** (home life data). Nel rapporto mostriamo che i sistemi di IA come gli assistenti virtuali - che sono utilizzati da Big Tech per "profilare" gli individui, trovare "soluzioni personalizzate" o "mitigare i rischi futuri" (come nel caso della profilazione vocale per determinare uno stato mentale) - si basano su dati raccolti attraverso pratiche domestiche digitali che sono disordinate, incomplete e contraddittorie. Concludiamo quindi che essere profilati sulla base di queste tracce di dati porterebbe inevitabilmente a un'analisi riduzionista e imprecisa dei nostri comportamenti.

Il problema dei nostri sistemi IA, però, non è solo che imparano da dati umani che sono spesso separati dall'intenzione e decontestualizzati, ma anche che vengono molto spesso addestrati con **banche dati pieni di errori e pregiudizi culturali e sociali**. Nel 2019, per esempio, i ricercatori dell'AI Now Institute di New York hanno pubblicato un rapporto terrificante sugli usi delle tecnologie IA da parte della polizia negli Stati Uniti. Il rapporto ha rivelato che in diverse giurisdizioni, le tecnologie IA utilizzate dalla polizia per analisi predittiva si sono basate su "dati sporchi", o in altre parole dati che erano stati raccolti durante un periodo storico in cui i corpi di polizia erano affetti da pratiche corrotte e ideologie razziste [4]. Questo è uno studio che io cito spesso, e che trovo particolarmente importante perché ci dimostra che se i sistemi di IA sono addestrati con set di dati che sono biased e imprecisi, finiscono per rinforzare e amplificare le disuguaglianze presenti nella nostra società. Quindi dobbiamo mettere in discussione i set di dati che vengono utilizzati per costruire i nostri sistemi di IA. Solo così possiamo cominciar a capire meglio perché questi sistemi sono molto spesso sbagliati e biased nella profilazione umana. In Italia, per esempio una domanda importante che dovremmo porci è perché nel database utilizzato dalle forze dell'ordine per il riconoscimento facciale (Sari), che comprende 16 milioni di cittadini, l'80% delle voci sono stranieri (2019 inchiesta Wired Italia).

Quando pensiamo agli errori e bias impliciti dei nostri database, dobbiamo renderci conto che purtroppo **non esiste una vera soluzione**. Le nostre banche dati non possono davvero essere corrette con dati "puliti e senza errori", perché molti dei nostri dati personali - e come vengono raccolti - riflettono il contesto sociale e culturale che li ha creati, e quindi sono intrinsecamente biased. Abbiamo la responsabilità di ridurre i nostri errori e bias nella raccolta dati, ma non possiamo evitarli completamente. Questa per me è la chiave di lettura. Se vogliamo davvero capire e limitare l'impatto sociale dei sistemi IA, dobbiamo **riconoscere l'errore umano dell'intelligenza artificiale**. È per questo motivo che, nell'ultimo anno, ho lanciato un nuovo progetto di ricerca all'Università di San Gallo, che si intitola *The Human Error Project: AI, Human Nature and the conflict over Algorithmic Profiling*. Nella mia ricerca, io e il mio team usiamo il concetto di "errore umano" proprio per sfruttare la sua ambivalenza e ambiguità. Le tecnologie IA sono spesso utilizzate per rendere il processo decisionale più efficiente e oggettivo e per "evitare l'errore umano". Eppure, paradossalmente, quando si tratta di leggere le persone queste tecnologie sono piene di "errori" sistemici, "bias" e "imprecisioni". Credo che sia fondamentale studiare questi errori perché fanno luce sul fatto che la corsa all'innovazione dell'IA è spesso plasmata non solo da interpretazioni stereotipate e approssimative della natura umana, ma anche da **teorie scientifiche problematiche che si basano su una lunga storia di riduzionismo umano**.

L'errore umano nei sistemi IA: tra riduzionismo e pregiudizio scientifico

A febbraio del 2021, la CNN ha pubblicato la notizia di un software IA, chiamato **4 Little Trees**, che viene usato in alcune scuole di Hong Kong per analizzare le espressioni facciali dei bambini,

determinarne le emozioni e intervenire a livello pedagogico. Pochi giorni dopo Kate Crawford, fondatrice dell'AI Now Institute a New York ha scritto un articolo su Nature [5], usando proprio questo esempio. Nell'articolo, Crawford ci fa notare come la scienza dietro questo sistema IA (e dietro la maggior parte di sistemi di classificazione delle emozioni facciali) sia basata sulle teorie di Ekman secondo cui esistono **6 emozioni "universali"** che sono innate, interculturali e coerenti - paura, rabbia, gioia, tristezza, disgusto e sorpresa - e che possono essere lette attraverso l'analisi delle espressioni facciali.

Crawford cita l'antropologa **Margaret Mead** evidenziando come la teoria di Ekman sia fallimentare perché non tiene in conto del contesto, della cultura e altri fattori sociali. Se pensiamo alla conoscenza antropologica, infatti, ci rendiamo conto che l'idea che le nostre emozioni siano universali e abbiano una corrispondenza oggettiva nel nostro corpo (come le espressioni facciali) è ancora tutta da provare. Qui non sto solo pensando al lavoro di **Michelle Rosaldo** sulla costruzione culturale dell'emozioni [6], ma anche al lavoro di **Brian Morris** sull'antropologia dell'io che dimostrano come psicologia e emozioni non sono solo determinate puramente da processi cognitivi ma da processi culturali [7].

L'idea che esistano 6 emozioni universali da mappare a espressioni facciali è stata screditata non solo dalla conoscenza antropologica ma anche dal lavoro di psicologi come Gendron et. al che hanno condotto **uno studio comparato** tra partecipanti americani e partecipanti della tribù dell'Himba facendo notare come le ricerche passate sull'universalità delle emozioni avessero usato metodi di ricerca non corretti [8]. Barrett et al., invece hanno dimostrato che ci sono ancora molte domande aperte sulla relazione tra espressioni e emozioni e che molte volte le persone non esprimono una sola emozione con l'espressione facciale [9]. La domanda, quindi, nasce spontanea: perché il mercato di IA di classificazione emotiva cerca validità scientifica nelle teorie di Ekman? La risposta secondo Crawford è ovvia: **la teoria di Ekman è stata adottata perché si adatta a ciò che i sistemi IA possono fare** [5]. Sei emozioni coerenti possono facilmente essere standardizzate e automatizzate su scala - a patto che i problemi più complessi siano ignorati.

La maggior parte dei nostri sistemi di IA sono basati sulla scienza occidentale, che è plasmata da una comprensione spesso distorta e riduzionista della natura umana. **È per questo che i nostri sistemi non interpretano male solo le nostre emozioni, ma anche i nostri corpi**. Negli ultimi 6 mesi io e la mia squadra sul **The Human Error Project** ci siamo resi conto che la maggior parte di esempi di errore algoritmico che venivano citati in più di 100 articoli di media internazionali in Europa prevedevano errori di **lettura sul corpo umano** e gli algoritmi venivano definiti razzisti o sessisti. Se vogliamo apprezzare davvero perché i sistemi IA e gli algoritmi sembrano sbagliare così tanto quando si tratta di leggere il corpo umano, dobbiamo guardare la storia del pregiudizio scientifico del pensiero occidentale (Poux-Berthe and Barassi, 2021). Nel 1981, Gould, per esempio, ha scritto un libro intitolato The Mismeasure of Man, dove dimostra come il pensiero scientifico occidentale si sia spesso basato su misure scientifiche che partivano dalle analisi biologiche modellate sull'uomo bianco [10]. Gould era particolarmente affascinato dai test su IQ e l'idea che l'intelligenza potesse essere misurata biologicamente (per esempio la misura del cranio) e dimostra come i benchmark di questi test usino **l'uomo bianco come punto di riferimento**. Una simile interpretazione la troviamo anche nel libro della sociologa Strings **Fearing the Black Body** che dimostra come i calcoli sull'indice di massa corporea, non sono stati raggiunti da studi che hanno misurato cosa voglia dire avere un peso-forma salutare in diverse etnie e contesti culturali, ma sono invece modellate su idee culturali e riduzioniste che prendono come riferimento il corpo caucasico [11].

Le tecnologie che stiamo creando si basano su dati e misure scientifiche che molto spesso portano con sé una lunga storia di riduzionismo umano e bias impliciti. Per questo non dobbiamo sorprenderci di tutti gli errori che stanno emergendo, quando si tratta della profilazione degli esseri umani. Un esempio chiave e particolarmente importante al momento che dimostra il riduzionismo umano dei nostri sistemi si trova nelle **app di tracciamento dei contatti** Covid-19. Nel suo affascinante lavoro, Milan ha dimostrato che la maggior parte di queste app si basa su un soggetto sperimentale "standard" che difficilmente permette di esplorare il ruolo di variabili come il genere, l'etnia, la razza o il basso reddito [12]. Milan mostra come le radici di questo riduzionismo, derivano dalla pratica stessa del design. È per questo motivo che si rifà all'antropologo Arturo Escobar che ha avanzato una nuova visione della teoria del design che tiene conto del pluriverso complesso e intersezionale in cui viviamo [13]. Pensare al pluriverso nel design delle tecnologie IA è sicuramente un passo avanti, come dice Escobar. Un altro passo importante, tuttavia, è quello di **imparare a coesistere con l'errore umano dell'IA** e riconoscere che questi sistemi saranno sempre in parte imprecisi e biased quando si tratta della profilazione umana.

Imparare a convivere con l'errore umano dell'intelligenza artificiale

Oggi si parla molto di bias dei sistemi IA, sempre più aziende tech stanno cercando di trovare soluzioni "etiche" per combattere i loro pregiudizi nei loro prodotti e tecnologie. È per questo motivo che stanno finanziando la ricerca e istituendo comitati consultivi che hanno lo scopo di esaminare gli impatti etici e politici delle loro tecnologie. Al centro di queste strategie e pratiche adottate dalle aziende, c'è la comprensione stessa che gli algoritmi sono distorti perché sono stati alimentati con "cattivi dati" e quindi, al fine di correggere il bias algoritmico, le aziende devono addestrare gli algoritmi con "dati equi" o "imparziali". Le attuali strategie per 'combattere il bias algoritmico' nel settore sono profondamente problematiche perché spingono alla conclusione che gli algoritmi possano essere corretti, ed essere imparziali [14].

Le attuali strategie per 'combattere i bias algoritmici', a mio parere non sono solo problematiche, ma sembrano non centrare il problema. Non esiste un sistema informatico che non sia biased. Nel 1996, Friedman e Nissenbaum, per esempio, hanno identificato tre tipi di bias nei sistemi informatici: bias preesistenti (il bias degli esseri umani che progettano sistemi informatici e il bias prodotto dal contesto culturale che influenza il design) bias tecnico (spesso c'è una mancanza di risorse nello sviluppo di sistemi informatici, e gli ingegneri lavorano con limitazioni tecniche, basta pensare all'esempio del riconoscimento emotivo) bias emergenti (la società è sempre in cambiamento e quindi le tecnologie progettate in un dato momento o contesto culturale potrebbero diventare biased in un tempo e contesto diverso) [15].

Gli algoritmi e i sistemi di IA sono fatti dall'uomo, e saranno sempre modellati dai nostri valori culturali e condizioni tecniche e sociali che li hanno creati. Invece di cercare di risolvere il bias dei sistemi IA e il loro errore umano, dobbiamo trovare il modo di coesistere con esso. L'antropologia qui ci può aiutare molto. Gli antropologi hanno cercato a lungo di affrontare il fatto che gli individui interpretano necessariamente i fenomeni della vita reale secondo le loro credenze culturali e la loro esperienza [16], e che i pregiudizi culturali si traducono necessariamente nei sistemi che costruiamo, compresi quelli scientifici [17]. Da una prospettiva antropologica, non c'è niente che possiamo davvero fare per "correggere" o combattere il nostro pregiudizio culturale, perché sarà sempre lì. L'unica cosa che possiamo fare è **riconoscere l'esistenza dei pregiudizi** attraverso una pratica autoriflessiva e ammettere che i sistemi, le rappresentazioni e gli artefatti

che costruiamo non saranno mai veramente “oggettivi”. Questa stessa comprensione dovrebbe essere applicata a tutti i nostri sistemi IA.

Bibliografia

[1] Završnik, Aleš. 2019. « Algorithmic Justice: Algorithms and Big Data in Criminal Justice Settings ». *European Journal of Criminology*: 1477370819876762.

[2] Figini, Silvia, et Vittoria Porta. 2019. « Algoritmi anti-crimine: tutte le tecnologie in campo ». *Agenda Digitale*. https://www.agendadigitale.eu/cultura-digitale/algoritmi-anti-crimine-tutte-le-tecnologie-in-campo/ (6 mai 2021).

[3] Barassi, Veronica. 2020. *Child Data Citizen How Tech Companies Are Profiling Us from before Birth*. MIT Press. https://mitpress.mit.edu/books/child-data-citizen (6 mai 2021).

[4] Richardson, Rachida, Jason Schultz, et Kate Crawford. « Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice ». *NYU Law Review*. https://www.nyulawreview.org/online-features/dirty-data-bad-predictions-how-civil-rights-violations-impact-police-data-predictive-policing-systems-and-justice/ (6 mai 2021).

[5] Crawford, Kate. 2021. « Time to Regulate AI That Interprets Human Emotions ». *Nature* 592(7853): 167-167.

[6] Rosaldo, Michelle Zimbalist. 1980. *Knowledge and Passion*. Cambridge University Press.

[7] Morris, Brian. 1994. *Anthropology of the Self: The Individual in Cultural Perspective*. Pluto Press.

[8] Gendron, Maria, Debi Roberson, Jacoba Marietta van der Vyver, et Lisa Feldman Barrett. 2014. « Perceptions of Emotion from Facial Expressions are Not Culturally Universal: Evidence from a Remote Culture ». *Emotion (Washington, D.C.)* 14(2): 251-62.

[9] Barrett, Lisa Feldman et al. 2019. « Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements ». *Psychological Science in the Public Interest* 20(1): 1-68.

[10] Gould, Stephen Jay, Steven James Gold, et The Alexander Agassiz Professor of Zoology Stephen Jay Gould. 1996. *The Mismeasure of Man*. Norton.

[11] Strings, Sabrina. 2019. *Fearing the Black Body The Racial Origins of Fat Phobia*. NYU Press. https://nyupress.org/9781479886753/fearing-the-black-body (6 mai 2021).

[12] Milan, Stefania. 2020. « Techno-Solutionism and the Standard Human in the Making of the COVID-19 Pandemic ». *Big Data & Society* 7(2): 2053951720966781.

[13] Escobar, Arturo. 2018. *Designs for the Pluriverse : Radical Interdependence, Autonomy, and the Making of Worlds*. Durham: Duke University Press.

[14] Gangadharan, Seeta Peña, et Jędrzej Niklas. 2019. « Decentering technology in discourse on discrimination ». *Information, Communication & Society* 22(7): 882-99.

[15] Friedman, Batya, et Helen Nissenbaum. 1996. « Bias in computer systems ». *ACM Transactions on Information Systems* 14(3): 330-47.

[16] Clifford, James, et George E. Marcus. 1986. *Writing Culture: The Poetics and Politics of Ethnography : A School of American Research Advanced Seminar*. University of California Press.

[17] Latour, Bruno, et Steve Woolgar. 1986. *Laboratory Life: The Construction of Scientific Facts*. Princeton University Press.

Fruizione e produzione delle immagini nella realtà virtuale: l'interiorità è un mito?

Quali sono le tendenze dell'immaginario attuale dell'interiorità? Si configura come un mito, seguendo la definizione di Benjamin? Un'analisi tra teoria ed ermeneutica delle immagini e sociologia dei media

Antonio Rafele, CEAQ, Université Paris La Sorbonne

Che cos'è l'interiorità nel mondo contemporaneo? Sembra assumere le fattezze di un mito, raggiungendo una configurazione altamente problematica.

In una pagina collocata ad inizio de "Per la Critica della violenza", il filosofo Walter Benjamin definisce il **mito come il contenitore di un fondo arcaico**, di cui anche la società moderna continua ad essere la riproduzione. Il mito, scrive Benjamin, adempie nelle società arcaiche una funzione indispensabile: rivela la realtà originaria, garantisce l'efficacia delle feste e dei culti, codifica le credenze, fonda regole morali, determina le pratiche della vita quotidiana (Benjamin, 1996, p. 75).

Il tema del mito e delle immagini

Il mito è una sorta di metalinguaggio: veicola contenuti e idee che travalicano il racconto, che paiono sostare oltre la vicenda narrata per illuminare un aspetto della realtà. Per mezzo di questa dinamica, "un ambiguo e perenne sostare dentro e fuori il testo", il mito agisce nella quotidianità, piegando i grandi quesiti esistenziali o eziologici alla gestione del quotidiano. Il mito diviene propriamente tale soltanto nell'attimo in cui si verifica un repentino spostamento dalla sfera della rappresentazione al mondo dei gesti e delle pratiche quotidiane.

Questo passaggio ha la forma di un riflesso meccanico e istantaneo, e qui risiede per Benjamin la forza e la peculiarità del mito, la cui potenza regressiva è espansa e amplificata dalla società moderna: "l'immediata affermazione di realtà e natura", ovvero **un racconto che si impone come una verità indiscussa, celando le intenzioni e le circostanze che l'hanno costruito**. Da questo preliminare piano di proiezione si dispiega in Benjamin **un'opposizione peculiare tra mito e immagine:** le immagini del ricercatore costituiscono una progressiva messa in tensione degli elementi che garantiscono la stabilità e l'efficacia del mito, offrendo analisi circoscritte.

Il metodo che organizza questa indagine attraversa due aree di ricerca contigue: **la teoria e l'ermeneutica delle immagini e la sociologia dei media**.

In particolare, seguendo la più esaustiva e recente classificazione di Belting (Belting, 2001), le **immagini** vengono colte nell'attimo in cui si stratificano come tracce viventi, **memoria vivente, dello spettatore**; ad essere messo in rilievo è dunque il loro essere riflesso o **sdoppiamento fantasmatico del reale**: uno spazio e un intervallo che, all'interno della storia e della

fenomenologia delle immagini, precede la narrazione in senso compiuto, facendo prevalere il lato tattile, epidermico, ma non meno profondo, dell'esperienza estetica.

L'analisi è stata pertanto condotta su due livelli: uno interno al **tessuto delle immagini,** mediante selezione e interpretazione delle scene che giungono a colpire in un sistema coerente l'attenzione e l'identità dello spettatore; l'altro usando l'**oggetto come filtro** per portare alla luce **alcune tendenze dell'immaginario attuale della interiorità**, e di quello ad essa concomitante del dolore. Il tema è stato inserito in una più ampia storia delle relazioni sociali, delle loro forme e variazioni, in una fitta trama di rimandi tra la sociologia della metropoli (Benjamin, 1983; Crary, 2000) e la sociologia dei media (McLuhan, 1964; Peters 1999; Morton, 2013).

Il piano di intersezione tra le tradizioni e gli autori menzionati si costruisce intorno a due presupposti teorici essenziali: **la centralità del fruitore e la centralità della forma narrativa.**

La centralità assunta dallo spettatore discende dal valore attribuito agli oggetti (Belting, 2001; Benjamin, 1983): g**li oggetti si completano nell'atto del consumo**, ovvero raggiungono un significato soltanto nell'uso che il fruitore ne compie a posteriori. Fuori da questo rapporto, essi non possiedono un'esistenza autonoma e perdono ogni funzione nell'ambito dell'esperienza. Ciò significa che ad acquisire un'importanza strategica è proprio **il rapporto, profondo anche se individuale, che il singolo stabilisce con un oggetto**, mostrandone in seguito la configurazione, gli effetti e le tendenze culturali che racchiude.

In un simile punto di osservazione, senso e figura coincidono, e l'analisi delle immagini non può prescindere da questa coincidenza che tiene insieme **in un'unica gerarchia il linguaggio e il vissuto** (Belting, 2001; Benjamin, 1983). L'analisi, entrando nella compiuta configurazione mediale, mette lo spettatore "nei panni" dell'autore; la destrutturazione dei temi e delle metafore, che costituiscono il tessuto soggiacente al testo multimediale, conduce lo spettatore a riconoscere nel testo la propria voce, un dettaglio del vissuto o dell'identità.

Il rapporto "organico" che si stabilisce tra immagine e spettatore configura le forme estetiche come i momenti in cui l'esperienza raggiunge la più alta evidenza. All'interno della narrazione il vissuto acquista una configurazione ricca e problematica, rivelando anche i rapporti e le rotture che il tempo presente stabilisce con le forme antiche dell'esperienza (Benjamin, 1983). La vita della metropoli, nell'interpretazione compiuta da Benjamin, diviene leggibile in una densa e sottile analisi del linguaggio: l'allegoria è il medium mediante cui illuminare i funzionamenti della moda, della storia, dell'identità, delle esposizioni universali, della pubblicità, dei rapporti sociali.

Le maggiori tendenze culturali divengono visibili dentro le forme narrative che ne realizzano un'esposizione attuale; il cinema e la realtà virtuale sono alcune di queste forme che, raffigurando e inventando le strutture dell'immaginario, "contengono" l'esperienza.

Realtà virtuale e stati di transitorietà accelerata: l'interiorità come riflesso meccanico

Il video di Ridley Scott (2017, "Alien: Covenant In Utero"), della durata di due minuti circa, quasi interamente scandito dal rumore del battito cardiaco, ricostruisce l'esperienza della nascita di un deforme e si compone di due parti distinte: una visione compiuta dall'interno del corpo, che sospinge per qualche frazione di secondo ad una sovrapposizione tra scena e spettatore (a questo paiono alludere le pieghe e i gemiti delle membrane), a cui succede **un'improvvisa – istantanea,**

come l'origine – scomparsa dell'immedesimazione; così, lo spettatore si colloca all'esterno della scena, e l'illusione appena vissuta conferma per contrasto una linea saldissima che si frappone tra testo e utente.

Questa polarità di fondo sul piano narrativo, come anche l'oggetto della scena, si ripetono in modo simile in alcune produzioni recenti: "Invasion" (2017), nel quale si assiste ad una ossessiva costruzione di un punto di vista "esterno", dall'alto, rispetto alla scena; "The Invisible Man" (2018) e "VR Dream Treatment Follow Up" (2018), nei quali la scena è interamente costruita sulla presenza/assenza dello spettatore; ugualmente, la ricerca di un effetto di credulità, una **visione talmente ravvicinata da celare le barriere di separazione**, è la strategia che contraddistingue "It: Float", 2017, "White Room", 2019, "The Conjuring 2", 2019 e "The Dream Collector", 2019.

"In Utero" è un video girato a 360° e si distingue, almeno in parte, da "The Limit", **cortometraggio in realtà virtuale** di Robert Rodriguez (che al momento si pone come lo spettro più avanzato del genere, seguito dei due esperimenti "Star Wars Rogue One Recon" e "Star Wars Hunting of the Fallen", entrambi del 2018), **che innesta nella narrazione cinematografica alcune funzioni tipiche del videogioco**.

Qui, **la coincidenza con la camera** (che, reiterata, come in molti videogiochi in VR, procura disagio fisico), i movimenti della testa e quelli del corpo nello spazio (in questo caso ad entrambi corrispondono modifiche nella vicinanza o lontananza rispetto al centro della scena), l'uso complementare della manetta che permette di partecipare allo svolgimento della trama (ma dentro un modello prestabilito: ad es., pur trattandosi di un game "sparatutto", lo spettatore non può prendere parte alle azioni violente o più spettacolari, limitandosi ad indicare ai protagonisti/collaboratori le due o tre opzioni rese disponibili), offrono una gamma più estesa di punti di vista o angoli di osservazione.

Se in "The Limit" l'effetto di illusione è dovuto ad una coincidenza tra sguardo e camera, ovvero le risposte dell'immagine corrispondono ai movimenti compiuti con la testa o con il corpo, tuttavia, in ogni sguardo o movimento, si impone una linea di demarcazione che nega il desiderio di una presenza più particolareggiata, ripetendo, come in "Alien In Utero", **la posizione di uno spettatore che occupa la scena e al contempo si riguarda dall'esterno come attore**.

"In Utero" è la messa in scena di un dolore insormontabile della carne, compiuta sfruttando le tecniche dello shock. L'ossessiva ricerca delle origini, la sopravvivenza della carne, in una subitanea e repentina oscillazione tra vita e morte (Costa, 2002), che è l'oggetto della scena ma anche il riflesso delle reazioni emotive dello spettatore (l'immagine si configura come la dilatazione di un istante che "ruba", al pari di uno stupefacente, la vita dello spettatore, sospingendolo in uno stato pre-coscienziale, attimo a cui succede la sensazione desolata della fine, dove l'immagine si riapre, allo sguardo retrospettivo, come un insieme di materiali secchi, aridi), rivela, in continuità con i ritmi della serialità televisiva, **uno stato di transitorietà accelerata**.

Queste sollecitazioni, che portano lo spettatore alle massime prestazioni nervose, sono anche **l'immagine capovolta di un corpo indebolito**, infiacchito, precocemente invecchiato: **un corpo reso dalle incessanti novità come immobile, "stordito"** (Benjamin, 1983), **nell'attesa di uno shock** che possa ricondurlo in vita, spostando ancora più in là il muro dell'indifferenza. Shock e tattilità delle immagini paiono congiuntamente delineare uno stato di sospensione delle funzioni mnemoniche, della loro necessità, approssimando lo spettatore all'immagine del deserto o della

polvere (Hansen, 2007). Si dovrebbe piuttosto parlare di "ripetizione" (Hansen, 2007), compiuta dentro il testo, o fuori dal testo come memoria vivente dello spettatore (è il caso della memoria come "rinnovata ripetizione" nella pornografia), ma presupponendo che una simile funzione è in primissima istanza esterna alla macchina narrativa, dato che il tutto si consuma in un istante, e dunque abbandonata al caso.

L'interiorità, come nelle seguenti descrizioni di Walter Benjamin, **assume la qualità di un riflesso tattile, meccanico**: "Con questo immenso sviluppo della tecnica una miseria del tutto nuova ha colpito gli uomini. [...] Sì, ammettiamolo: questa povertà di esperienza dell'umanità in generale. E con questo una specie di nuova barbarie. Barbarie? Proprio così. Diciamo questo per introdurre un nuovo positivo concetto di barbarie. **A cosa mai è indotto il barbaro dalla povertà di esperienza? È indotto a ricominciare da capo; a iniziare dal nuovo; a farcela con il poco**: a costruire a partire dal poco e inoltre a non guardare né a destra né a sinistra" (Benjamin 2002, 257). Il filosofo sottolinea come quando il lavoro umano è unicamente distruttivo, allora è realmente di un lavoro umano, naturale, nobile: l'inumano è ambasciatore di un umanesimo più reale, e solidarizza non con l'abete slanciato, ma con la pialla che lo consuma, non col metallo nobile, ma con la fornace che lo raffina (Benjamin 2002, 261).

La molteplicità delle immagini è il riflesso di un'esistenza priva di personalità: un modo di vita che cresce a stento, giorno per giorno, sulla scorta delle circostanze. È l'immagine della solitudine, di un io abbandonato ai grandi e minuti risvolti della vita quotidiana, ma anche il segno di una crisi irreversibile (un essere inattuale, antiquato) della formazione e delle agenzie predisposte rispetto a un tempo non codificabile, classificabile. La soggettività a cui Benjamin allude, e che sembra costituire sul piano storico l'antecedente delle attuali soggettività della rete, ha un tratto essenzialmente barbarico: essa riproduce una temporalità che si articola mediante riflessi rapidi, meccanici, modalità del tutto distanti dunque dalla tradizionale costruzione dell'interiorità.

La sovrapposizione che si determina tra immagini e spettatore, in una compresenza inverosimile di azione e reazione, causa ed effetto insieme del successo duraturo della tecnica dello shock, si giustifica sulle imponenti trasformazioni sensoriali che la metropoli, e poi in seguito le immagini, tra XIX e XX secolo, introducono nella vita quotidiana.

In uno stato di transitorietà accelerata, la distrazione non è una forma degradata di esperienza, bensì uno strumento attivo ed efficace: essa permette al singolo di seguire, senza eccessivi sconvolgimenti interiori, un ritmo veloce e dispersivo.

L'attenzione, che un un'immagine può destare nello spettatore, avviene in una interruzione del continuum: una momentanea e improvvisa sospensione del tempo mediante cui il singolo si immerge, anche se per brevi istanti, in una nuova illusione. Per abbattere il muro dell'indifferenza e del "già vissuto", le immagini devono intervenire con sempre nuove sorprese, possedere cioè la parvenza di uno shock sensoriale ed emotivo, che è al contempo un'innovazione tecnica e un "rimpasto" inedito dell'immaginario. Il rapporto circolare che si instaura tra distrazione e attenzione costituisce lo spazio psichico in cui si insinuano le immagini, la loro produzione come anche la loro riuscita nel contesto della metropoli e dei media.

Sul piano gnoseologico, le esperienze sopra descritte, riconfigurano i modi con cui l'osservatore raggiunge un'immagine sui fenomeni vissuti. Le immagini che compongono la trama dello spettatore sono colte nell'attimo in cui si stratificano come tracce della mente. Lo sguardo

retrospettivo colloca, in una vertiginosa torsione all'indietro, l'origine dell'evento narrativo nell'attimo in cui l'immagine libera i suoi effetti: il passaggio repentino, pressoché istantaneo, dal gesto presente nella scena al ricordo di chi guarda. **Non si tratta di una posizione preesistente all'analisi, bensì dell'avvenuta, in concomitanza col replay, centralità dello spettatore.** Per via dei salti che dissemina durante la fruizione, il replay configura le immagini come allegorie della visione, su cui occorre rivenire e indugiare per ricostruire l'esperienza vissuta. Piacere e memoria sono i due momenti inscindibili della visione, il costituirsi di un modo d'essere interamente proiettato sull'attualità.

Nonostante l'uso della soggettiva, **l'esperienza avviene in un tempo parallelo rispetto a quello della narrazione,** in un intervallo dominato dalla memoria e dalla ripetizione ossessiva di pochi attimi o secondi. Così, il tipo di esperienza compiuta rileva nitidamente che l'immersione non avviene in una presunta vicinanza con l'immagine, cioè nel fare proprio il punto di vista della camera; piuttosto, essa si colloca altrove, nell'attimo in cui una scena giunge a procurare nello spettatore un ricordo, una memoria vivente, da cui soltanto discende in seconda battuta il coinvolgimento sensoriale ed emotivo. Pur trattandosi strettamente di rappresentazioni, queste immagini vengono vissute come un campo di possibilità, una riproduzione fantasmatica del reale.

Interiorità e realtà virtuale: come cambiano le rappresentazioni del corpo e del sociale

Le modifiche intervenute nella rappresentazione dell'interiorità e della sofferenza sembrano strettamente legate ad una riconfigurazione dei rapporti preesistenti tra storia e natura, verso cui i media agiscono come potenti acceleratori.

In primo luogo, **la rappresentazione del corpo si slega dalla dicotomia storica sano/malato per divenire, essa stessa, un'immagine o una proiezione della tecnologia dominante**: si pensi alla differenza tra la medicina o la fisica ottocentesche fondate sulle tecnologie meccaniche e quelle novecentesche basate, invece, sulle tecnologie elettriche (McLuhan, 1964). La dimensione biologica è una proiezione del tempo storico, nel senso che la si può concepire e rappresentare solo a partire dai media di cui si dispone. Così, nella rappresentazione medica il corpo appare, in perfetta coincidenza con le potenzialità delle tecnologie elettriche, un sistema complesso, in cui tutto simultaneamente si tiene, e, allo stesso tempo, precario, in quanto tende all'errore (Sontag, 1978). Ogni corpo è potenzialmente malato, e la malattia è un evento che determina un'inaspettata caduta del sistema. La dimensione che può seguire all'evento della malattia consiste in una riconfigurazione del proprio mondo quotidiano, anche in base alle eventuali nuove caratteristiche biologiche; ci si "ricolloca" conservando la sensazione e la coscienza di morte dell'antico sistema di vita.

In secondo luogo, nelle attuali pratiche quotidiane e mediali sembra irreversibilmente entrare in **crisi il tradizionale concetto di comunità** (Morton, 2007). Se il termine "comunità" evoca una postura umanistica, ovvero l'immagine di un tempo lineare in cui presente e futuro si intrecciano, il mondo dei media delinea al contrario un'immagine radicale, profondamente anti-umanistica, della storia: la storia è abbandonata alle minute storie delle tecnologie e dei rapporti di forza che di volta in volta si determinano attorno ad esse: una storia di desideri, interessi e violenze che i gruppi umani dispiegano all'interno degli ambienti mediali – che così rivelano la loro natura ambivalente: media come campo di possibilità e nuove assuefazioni, ma anche media come armi

(McLuhan, 1964) – annullando qualsiasi proiezione del tempo. Ne consegue una visione che supera le precedenti distinzioni di gusto tra normalità e deforme, spostando di volta in volta l'attenzione, in una postura di radicale neutralità, su ciò che è in grado di procurare piacere o interesse. Così, la dimensione estetica, interamente proiettata sulla tecnica dello shock e sulla riuscita performativa degli oggetti, prolifera di esperienze in cui gli elementi disgustosi o deformi acquistano una nuova, autonoma legittimità.

Più in generale, in queste esperienze **il sociale pare sottrarsi dalla vista di chi osserva, lasciando un paesaggio di rovine**: la catastrofe dell'umanesimo ad opera di un angelo disumano "che preferirebbe liberare gli uomini privandoli di qualcosa, piuttosto che allietarli donando loro qualcosa" (Benjamin 2002, 341). Un angelo sterminatore, che comprime in una sola potenza le immagini fotografiche, della pubblicità o del cinema. L'immagine divora fino ad inghiottire la vita dello spettatore, che, sollecitato senza sosta, appare infine a se stesso "più morto che vivo" (Simmel, 1995). Se in un autore come McLuhan continuano ad agire tensioni di matrice umanistica, seppur come residui rispetto ai problemi posti dall'immagine (le configurazioni, le protesi, le amputazioni, il mito di narciso, il rigor mortis, i linguaggi del sentire, le reazioni in profondità, le assuefazioni), quei residui sono un invito ad uscire da se stessi, nella speranza di congiungere i diseredati e i "cadaveri" della tecnica moderna. Ma ciò non è in fondo una presa di coscienza, che proprio quegli "stralci" umanistici rivelano in modo nitido, di una impossibilità del sociale? O almeno l'apertura "in negativo" di una riflessione sulle condizioni di possibilità della vita sociale, al di là dello spirito di affermazione che pare contraddistinguere i suoi scritti?

Bibliografia

Belting H. (2001). "Bild-Anthropologie. Entwürfe für eine Bildwissenschaft". München: Fink.

Benjamin W. (1983). "Das Passagen-Werk". Frankfurt am Main: Suhrkamp (trad.it.: "I passages di Parigi". Torino: Einaudi, 2002).

Byung-chul H. (2013). "Digitale Rationalität und das Ende des kommunikativen Handelns". Berlin: Matthes & Seitz (trad.it.: "Razionalità digitale. La fine dell'agire comunicativo". Firenze: GoWare, 2014).

Boltanski L. (2000). "Lo spettacolo del dolore. Morale umanitaria, media e politica". Milano: Raffaello Cortina.

Calleja G. (2011). "In-Game. From Immersion to Incorporation". Cambridge: MIT Press.

Crary J. (2000). "Suspensions of Perception: Attention, Spectacle and Modern Culture". Cambridge: MIT Press.

Grau O. (2004). "Virtual Art: From Illusion to Immersion". Cambridge: MIT Press.

Hansen M. B. N. (2007). "Bodies in Code: Interfaces with Digital Media". New York: Routledge.

Illouz E. (2008). "Saving the Modern Soul: Therapy, Emotions, and the Culture of Self-Help". Berkeley: University of California Press.

McLuhan M. (1964). "Understanding media". New York: McGraw Hill (trad.it.: "Gli strumenti del comunicare". Milano: Net, 2002).

Mittel J. (2017). "Complex TV. Teoria e tecnica dello storytelling delle serie TV". Roma: Minumun Fax.

Morcellini M. (2020). "Antivirus. Una società senza sistemi immunitari alla sfida del Covid-19". Roma: Castelvecchi.

Morton T. (2013). "Hyperobjects: Philosophy and Ecology after the End of the World". Minnesota: University of Minnesota Press.

Pecchinenda, G. (2018) "L'essere e l'io. Fenomenologia, esistenzialismo e neuroscienze sociali". Roma: Meltemi.

Peters J. D. (1999). "Speaking Into the Air: A History of the Idea of Communication". Chicago: University of Chicago Press.

Sontag S. (2006). "Davanti al dolore degli altri". Milano: Mondadori

Fermiamo la cyber-war, prima che sia troppo tardi: la soglia da non attraversare

Dopo il lancio della prima bomba atomica su Hiroshima, il fisico Robert Oppenheimer scrive: "i fisici hanno conosciuto il peccato". Lo stesso rischio lo stanno correndo gli informatici. I rischi sono inaccettabili, ecco perché

Di **Norberto Patrignani**, Politecnico di Torino

Quando un settore della scienza e della tecnologia rischia di essere usato in applicazioni militari, i policy maker e la società civile in generale chiedono conto a chi progetta questi strumenti del loro possibile "dual use". Sono quindi le persone del mondo della scienza e della tecnologia a dover riflettere sugli aspetti sociali ed etici, questo è già molto chiaro a Leonardo Da Vinci nel 1506[1].

https://www.agendadigitale.eu/cultura-digitale/se-lintelligenza-artificiale-uccide-senza-controllo-umano-ecco-i-punti-chiave-del-dibattito/

Famosa è anche la riflessione del fisico **Leo Szilard** (1898-1964) dopo aver visto la prima reazione atomica a catena la notte del 3 marzo 1939: "Spegnemmo tutto e tornammo a casa. Quella notte, nella mia mente non vi era il minimo dubbio che il mondo era diretto verso un grande dolore" (Klein, 1992). La storia purtroppo conferma i presentimenti di Szilard e, dopo il lancio della prima bomba atomica su **Hiroshima**, il fisico Robert Oppenheimer (1904-1967) scrive: "i fisici hanno conosciuto il peccato".

Lo stesso rischio lo stanno correndo gli **informatici**. Lo sviluppo di cyber-weapon nelle reti informatiche, robot autonomi dotati di armi letali, sensori e sofisticati algoritmi di intelligenza artificiale rischia di scatenare una nuova **corsa agli armamenti in versione cyber-war**, spingendo gli scienziati dei computer e l'umanità intera verso una soglia che forse non dovremmo attraversare (Patrignani, 2018).

Computer e guerra

Il legame tra computer e guerra risale alle origini stesse di queste tecnologie. Il grande matematico del secolo scorso **John Von Neumann** (1903-1957) è a Los Alamos proprio negli anni in cui viene sviluppata la bomba atomica (LANL, 2021). Grazie ai faraonici investimenti del governo USA nel **Progetto Manhattan** e perfezionando l'idea di macchina universale introdotta nel 1937 da **Alan Turing** (1912-1954), Von Neumann sviluppa l'architettura dei computer (input, memoria, cpu, e output) che ancora oggi porta il suo nome.

Da questa parte dell'Atlantico, negli stessi anni, a Bletchley Park, Turing, con l'aiuto di centinaia di altre persone, decifra **Enigma**, la macchina crittografica usata dai nazisti contribuendo alla fine della Seconda guerra mondiale (Bletchley, 2021). Le enormi potenzialità delle nuove tecnologie digitali sono chiare anche a **Norbert Wiener** (1894-1964), professore al MIT e fondatore della cibernetica, ma con una visione completamente diversa rispetto a quella di Von Neumann. Infatti,

mentre Von Neumann collabora alle applicazioni militari, Wiener usa quasi le stesse parole usate da Leonardo Da Vinci secoli prima: "I do not expect to publish any future work of mine which may do damage in the hands of irresponsible militarists..." (Wiener, 1947). Lo scambio di lettere tra, il "super-falco" Von Neumann e il "pacifista" Wiener, rappresenta una delle più importanti testimonianze del **dibattito** sulle relazioni tra tecnologia e società per quanto riguarda il mondo digitale (Heims, 1980).

Computer professionals for social responsibility

Negli anni '60 i grandi movimenti pacifisti nelle università statunitensi riprendono il tema del coinvolgimento della ricerca scientifica nella guerra. **Joseph Weizenbaum** (1923-2008), altro grande scienziato dei computer, denuncia esplicitamente i rischi connessi con l'applicazione dei computer in ambito militare: "the question is not whether such a thing can be done, but whether it is appropriate to delegate this hitherto human function to a machine" (Weizenbaum, 1976).

La rivista "Science for the People" contribuisce a sviluppare questo dibattito (Weizenbaum, 1985), in seguito al quale, nel 1983 al Palo Alto Research Center, uno dei più grandi centri di ricerca informatica, nasce l'associazione internazionale Computer Professionals for Social Responsibility. L'opposizione al grandioso progetto del governo USA SDI (Strategic Defence Initiative), noto come "**Star War**" emerge in tutta la sua forza con le dimissioni dello scienziato dei computer **David Parnas** dal comitato scientifico che doveva supervisionare il progetto (Parnas, 1985). L'argomento principale: delegare al software il controllo del lancio di missili intercontinentali introduce **rischi inaccettabili**.

La cyber-war diventa realtà

Purtroppo, tutte le voci contrarie all'uso delle tecnologie digitali in ambito militare non hanno trovato ascolto. Arriviamo così ai giorni nostri dove ormai tutte le strategie militari prevedono uno **scenario "multi-domain"**: i cinque domini maritime, land, air, cyberspace, space sono fittamente interconnessi da reti di computer. La cyber-war diventa realtà. La combinazione delle conseguenze inimmaginabili di una guerra atomica e le probabilità di eventi scatenanti (introdotti dalla complessità del digitale) porta a rischi inaccettabili (Unal, 2021). Infatti, le probabilità sono accresciute dalla combinazione di minacce provenienti da stati o da gruppi (sponsorizzati dagli stati stessi) e dalle cyber-vulnerabilities. Si pensi alla fitta interconnessione di migliaia di server, alle vulnerabilità inevitabili del software, ai sistemi che controllano i silos missilistici, etc. In questi scenari emergono **domande difficilissime**: in quali condizioni un cyber-attack va considerato un "atto di guerra"? A chi e come va "attribuito"? Quali risposte predisporre come "difesa"?

Eppure la convocazione di **una convenzione internazionale** su questi temi riceve sempre l'opposizione da parte di molti stati (Eilstrup-Sangiovanni, 2018). L'argomento principale di queste resistenze è quello della difficoltà a trovare accordi internazionali in uno scenario dove le tecnologie digitali cambiano rapidamente. Eppure, molti stati registrano in continuazione **cyber-attacchi** contro i sistemi militari e le infrastrutture nazionali critiche. Uno dei casi più recenti: l'attacco alla rete dell'energia nel Nord Est degli USA in maggio 2021 (Herman, 2021).

Stop killer robot

L'evento più inquietante è quello descritto in un recente rapporto del marzo 2021 al Consiglio di Sicurezza dell'ONU che segnala, forse per la prima volta nella storia in modo ufficiale, l'uso di **droni autonomi**, le cosiddette Lethal Autonomous Weapons Systems (LEWS). Definite come armi automatiche che: "... una volta attivate, sono capaci di selezionare e attaccare un obiettivo senza ulteriori interventi da parte degli esseri umani" (DOD, 2012). Infatti il rapporto del "Panel of Experts on Libya" scrive: "The lethal autonomous weapons systems (LAWS) were programmed to attack targets without requiring data connectivity between the operator and the munition: in effect, a true “fire, forget and find” capability." (UN, 2021, pag.17).

Questa soglia non va attraversata: la delega alla macchina di uccidere, pone gli informatici, tutti i computer professionals, di fronte a interrogativi molto simili a quelli affrontati dai fisici nel 1939. Diventa importante ricordare il "**Roboethics Manifesto**", del 2004, una delle prime posizioni contro lo sviluppo di sistemi digitali e robot "against human beings and the environment". In questo storico documento la comunità di computer scientist definiva, per la prima volta, il "design, building and use of 'intelligent machines' against human beings" un **crimine contro l'umanità** (Veruggio, 2004). Le posizioni più lucide contro le LAWS si basano sulla **impossibilità di applicazione del diritto umanitario in guerra** (distinzione e proporzionalità) e di un'attribuzione di responsabilità (chi è responsabile di un crimine di guerra compiuto da un LAWS?). Viene inoltre negata la dignità umana: la vittima di un LAWS non può fare appello all'umanità di "qualcuno che si trovi dall'altra parte" (Tamburrini, 2018).

Queste sono anche le principali argomentazioni della campagna internazionale per **fermare le armi autonome:** "Campaign to stop killer robots" (stopkillerrobots.org).

Infine va segnalato un recente documento delle "Pugwash conferences", un movimento internazionale fondato nel 1957 a partire dal manifesto con cui Russell, Einstein, e Rotblat esortavano la comunità degli scienziati a denunciare i pericoli della guerra nucleare e a promuovere il disarmo. Nel Novembre 2020 hanno rilasciato un documento dedicato proprio alla cyber-war dove raccomandano:

di **proibire** cyber-attack a infrastrutture critiche e installazioni nucleari,

di **studiare** le cyber-vulnerabilities delle armi nucleari e del rischio del loro uso accidentale,

di **approfondire** gli aspetti etici e legali delle LAWS basati su "intelligenza" artificiale,

di **supportare** la UN Global Commission on the Stability of Cyberspace (Pugwash, 2020).

Tutto questo dovrebbe essere di aiuto a sensibilizzare le giovani generazioni di computer professionals a contribuire al rispetto dei primi due principi del "Codice Etico e di condotta professionale" adottato recentemente dall'IFIP (International Federation for Information Processing): "a computing professional should contribute to society and to human well-being, acknowledging that all people are stakeholders in computing; a computing professional should avoid harm" (ACM, 2018; IFIP, 2021).

Bibliografia

- ACM (2018). ACM Code of Ethics. www.acm.org.

- Bletchley (2021). Alan Turing at Bletchey Park. https://bletchleypark.org.uk/learn/resources/faq-alan-turing visitato 5 Giugno 2021.

- Da Vinci, L. (1506). Codice Leicester (f.15A-22v). Da Vinci L. (2002), The Notebooks of Leonardo Da Vinci, Konecky & Konecky, 2002.

- DOD (2012). Directive 3000.09/2012: Autonomy in Weapons Systems, pp.13-14.

- Eilstrup-Sangiovanni, M. (2018). Why the World Needs an International Cyberwar Convention. Philosophy & Technology. 31, 379-407.

- Heims, S.J. (1980). John Von Neumann and Norbert Wiener, from mathematics to the technologies of life and death. MIT Press.

- Herman, C. (2021, 8 May). Cyber attack shuts down vital fuel pipeline to Norteast US. Forbes.

- IFIP (2021, 28 Janury). The global conscience of the profession. IFIP Code of Ethics and Professional Conduct.

[1] "Perché io non iscrivo il mio modo di star sotto l'acqua?…

Questo non pubblico o divulgo per le male nature delli omini,

li quali userebbono li assassinamenti nel fondo de' mari…"

Leonardo Da Vinci, 1506

https://www.ifipnews.org/media-release-ifip-launches-global-code-ethics-ict-sector/

- Klein G. (1992), The Atheist and the Holy City: Encounters and Reflections. MIT Press.

- LANL (2021). Los Alamos National Labs. www.lanl.gov visitato 5 Giugno 2021.

- Parnas L.D. (1985, Sept-Oct). Why the SDI software system will be untrustworthy. American Scientist, 73:5, 432-440.

- Patrignani, N. (2018, 11 Ottobre). Perché vanno fermati i robot killer. L'Adige.

- Pugwash (2020). Pugwash document on cyber security and warfare. Pugwash conferences.

- Tamburrini, G. (2018). Le armi autonome e le ragioni etiche per il controllo umano significativo dei sistemi d'arma. Rovereto, 13 ottobre 2018, Festival Informatici senza frontiere.

- Unal, B. (2021). Strategic Stability and Cyber and Space Dependency in Nuclear Assets. Chatham House (UK).

- Veruggio, G. (2004). First International Symposium on ROBOETHICS. The ethics, social, humanitarian and ecological aspects of Robotics, 30th - 31st January, 2004, Villa Nobel, Sanremo, Italy.

- Weizenbaum, J. (1976). Computer Power and Human Reason. Freeman.

- Weizenbaum, J. (1985, March-April). Computers in uniform: a good fit? Science for the People, Vol(17), 1-2.

- Wiener, N. (1947, January). A Scientist Rebels. Atlantic Monthly.

L'information war in Ucraina (2013-15): una narrazione agitprop

Il conflitto civile ucraino tra il 2013 e il 2015 è stato contraddistinto da una guerra di informazioni che ci dà l'occasione di riflettere sulla sostanza del legame fra la fonte e il destinatario di una notizia e sulla natura mimetica della propaganda da una prospettiva semiotica e post-strutturalista

Di **Luigi Giungato**, ricercatore della Socint (Società italiana di intelligence)

Possono le information war essere ricondotte a conflitti di narrazione osservabili attraverso il ricorso ai paradigmi narratologici? Se sì, è plausibile pensare che i costrutti narrativi della propaganda non abbiano l'obiettivo di perseguire la rappresentazione della realtà, ma di costruire la più avvincente narrazione di essa?

Il presente studio tenterà, attraverso l'esame della guerra di informazioni associata al conflitto civile ucraino tra il 2013 e il 2015, di rispondere a questi quesiti da una prospettiva semiotica e post-strutturalista.

Ucraina 2013-'15: il conflitto delle interpretazioni

Nel 2013, su un piccolo **portale russo** dedicato a temi militari, viene pubblicato un articolo firmato dal Capo di Stato Maggiore delle Forze Armate Russe, **Valery Gerasimov**, nel quale viene esposta una nuova dottrina per affrontare i **nuovi scenari internazionali**. Fra gli assunti più rilevanti, la presa d'atto che il principale campo di battaglia sul quale avvenivano – e sarebbero avvenuti – i conflitti armati negli anni a venire, sarebbe stato quello immateriale delle **informazioni**. Nel frattempo, in Ucraina, una guerra di narrazioni contrapposte iniziava a coagularsi attorno alla **questione della lingua ufficiale**. Il dibattito pubblico veniva esacerbato talmente tanto che, come riportato da Pogrebinskiy (2015), da un certo punto in poi, anche solo l'utilizzo in pubblico dell'idioma russo o ucraino era sufficiente a determinare una forma dicotomica di appartenenza, in un territorio, fino a quel momento, pacificamente bilingue.

Nel novembre dello stesso anno, il Governo ucraino, guidato dal Presidente Yanukovyc, sospende i negoziati per l'inizio della procedura di ingresso nella UE, scatenando il cosiddetto **Euromaidan** dal nome Maidan Nezalezhnosti, la Piazza dell'Indipendenza di Kiev, teatro della rivolta ovvero un'ondata di manifestazioni e scontri popolari filoccidentali, sfociati nella cosiddetta **Rivoluzione della Dignità** (Revoliutsiia hidnosti), la fuga in Russia di Yanukovyc e la sua conseguente destituzione da parte del Parlamento ucraino.

È l'inizio di una **guerra civile e simbolica** combattuta sia sul terreno fisico sia mediatico, sui social network e sui telefoni cellulari, fra due macro-narrazioni contrapposte: una, che vedeva il libero e

democratico Stato dell'Ucraina bersaglio di un'offensiva propagandistica, politica e militare russa; l'altra, che vedeva negli Stati Uniti una nazione imperialista impegnata a fomentare una rivolta al fine di destabilizzare l'Europa Orientale.

Tra febbraio e marzo 2014, avviene lo scoppio della **rivolta filorussa in Crimea**, in conseguenza della quale il Governo russo, su invito delle autorità regionali, autorizza la mobilitazione delle proprie forze speciali e annette i territori della penisola, dopo che truppe paramilitari non identificate avevano già occupato luoghi chiave e punti strategici, sostanzialmente senza sparare un colpo. Contemporaneamente, la guerra civile scoppia anche nelle province orientali del Donbass, più specificatamente nei territori di Donetsk e Luhansk, al confine fra Russia e Ucraina. Dopo una prima fase vittoriosa per le truppe paramilitari filorusse, la controffensiva delle forze armate ucraine legittimiste conduceva a uno stallo che sarebbe poi sfociato nell'**accordo di Minsk** per un cessate il fuoco il 5 settembre 2014, lasciando tuttavia il territorio in una situazione instabile che si protrae tuttora.

Lo scenario ucraino subito appare agli analisti una pedissequa applicazione della **dottrina Gerasimov** (2013), **una guerra ibrida** (Nemeth, 2002), in cui la soluzione militare detiene un rapporto di 1 a 4 nei confronti di qualunque soluzione non militare e nella quale la battaglia determinante viene combattuta sul campo informativo e non su quello materiale. All'inizio del 2015, quasi a sottolineare il cambiamento di paradigma avvenuto nella strategia militare russa, viene firmato il nuovo **documento sulla strategia per la sicurezza nazionale** dal Presidente Vladimir Putin, in linea con la teoria espressa dai vertici militari. In essa, la soluzione più strettamente militare viene indicata come ultima istanza, a vantaggio di una più consistente applicazione del soft power, attraverso il ricorso ai media digitali.

Agitprop 2.0: old wine in a new bottle

Come riportato da Veebel (2015), nel conflitto ucraino, le narrazioni si sono focalizzate su **alcune principali direttrici**:

demonizzazione dell'avversario, posto nel ruolo di antagonista, attraverso un surplus di connotazioni negative;

legittimazione degli alleati e delle proprie élite politiche;

mobilitazione della popolazione verso valori e narrazioni identitarie e culturali;

deterrenza e demoralizzazione;

confusione o, per meglio dire, compresenza di fatti precisi e verificabili associati a fatti falsi, sì da costringere, sia il destinatario sia l'avversario, a un continuo e stressante tentativo di confutazione e verifica;

overload informativo, attraverso la somministrazione di un sovraccarico di notizie, tese a inibire il contraddittorio e l'iniziativa;

controllo dell'iniziativa: la prima versione di un fatto, solitamente, è quella che resta in cima alle successive confutazioni, come per una sorta di imprinting cognitivo.

In tutti i casi, la modalità prevalente della propaganda da parte di entrambi i contendenti in Ucraina, si è manifestata attraverso la veste di notizia, veicolata su determinati media specifici, secondo un meccanismo di **mimetizzazione fra codici differenti**, generatore di corto circuiti informativi.

Durante il conflitto ucraino, le operazioni di informazione russe hanno utilizzato una serie di canali, tra i quali quelli più tradizionali che hanno convissuto con quelli più contemporanei (Sazonov, Mölder, Müür, Saumets, 2017). Il **controllo delle stazioni radio e televisive, i giornali, l'utilizzo di agenti, spie, agitatori e attivisti** di parte, l'uso di altoparlanti nelle zone nevralgiche urbane in altre parole, tutti mezzi del vecchio armamentario sovietico si è accompagnato a un'intensa attività propagandistica condotta su internet, sui social media, nonché tramite il controllo dei principali operatori telefonici (aspetto, questo, che si configura, in effetti, come il più interessante dal punto di vista dell'innovazione dei mezzi di diffusione dei messaggi).

Narrazioni confliggenti hanno, dunque, attraversato un territorio ibrido in cui sia la produzione dei contenuti sia la fruizione non costituivano che meri elementi di una catena di contagio mediatico che, da un supporto all'altro e soprattutto mediante i personal media, coinvolgevano fasce sempre più consistenti di popolazione. Molte delle narrazioni erano accomunate dal ricorso a elementi della realtà fattuale col solo scopo di attribuire i ruoli di una struttura narrativa in grado di cambiare specularmente, a seconda del punto di vista del lettore. Nella maggior parte dei casi, l'immagine era usata come realtà e la parola come schema narrativo, alla stregua del funzionamento dei meme (Ciracì, 2021). Emblematico, in tal senso, appare l'uso del topos rappresentato dai **cosacchi**, simbolo della nazione ucraina per il Governo di Kiev, simbolo della difesa dell'impero russo e della chiesa ortodossa per i separatisti.

Uno dei metodi più utilizzati per la fabbricazione di narrazioni sul teatro di guerra è così illustrato (Sazonov, Mölder, Müür, Saumets, 2017): sabotatori e ufficiali dell'intelligence giungono in una particolare location, accompagnati da almeno due giornalisti, uno specializzato in questioni militari, l'altro in materie civili. Iniziano col ricreare determinate condizioni che saranno poi riprese in video e immediatamente postate su Youtube, da dove saranno, poi, ricondivise sugli altri social e sulle televisioni ucraine e russe. Il metodo utilizzato non è quello della ripresa cinematografica – nella quale il linguaggio della fiction crea una barriera di incredulità nei presenti – bensì rientra in quelle che sono le tecniche e il linguaggio tipici della **messinscena agitprop** (Brown, 2013), in cui la drammaturgia è spesso costituita da un canovaccio a struttura narrativa fissa, nella quale l'intreccio è improvvisato, di volta in volta, in virtù del coinvolgimento attivo e partecipato dei soggetti presenti. Ciascuna situazione inscena necessariamente un conflitto che separa nettamente su due assi i personaggi e invita gli spettatori/attori a prendere parte alla rappresentazione in qualità di aiutanti. Nel caso ucraino, grazie alla condivisione digitale in Rete mediante il personal medium, la messinscena oltrepassa gli angusti confini di spazio e di tempo a cui è tradizionalmente legata. D'altra parte, una fortunata espressione per descrivere il contesto mediatico nei regimi neoautoritari, è quella di **staged pluralism** (Tolz & Teper, 2018), sebbene, in effetti, anche la stessa narrazione giornalistica occidentale dei conflitti armati si basi su un preconcetto e quindi su un discorso ideologico (Foucault, 1997) che riconosce al giornalista, in qualità di corrispondente di guerra, una sorta di imparzialità e, quindi, di obiettiva aderenza alla realtà.

Molte narrazioni impiegate per la costruzione dei contenuti poggiano su **discorsi profondi sedimentati** e difficilmente smantellabili, come quelle basate sulla retorica della Seconda guerra

mondiale, della Grande guerra patriottica contro i nazifascisti o della fratellanza dei popoli slavi. Alcune utilizzano schemi sicuri e affidabili, come la creazione di eroi positivi e di martiri del popolo ucraino. Altre tendono a generare panico e depressione, come le notizie non confermate di cortei di **tank russi diretti su Kiev**, i messaggi inviati sui cellulari da parte di conoscenti e amici che invitano alla fuga immediata (Sazonov, Mölder, Müür, Saumets, 2017) o le riprese di cadaveri di soldati governativi legittimisti vilipesi e lasciati insepolti nelle strade. Alcune particolari narrazioni non si preoccupano affatto di raggiungere sempre livelli soddisfacenti di verosimiglianza, avendo come obiettivo principale il raggiungimento di un sovraccarico informativo, generatore di ansia (Giungato, 2020): soldati ucraini dipinti come assassini, criminali e neonazisti, testimonianze di esecuzioni di bambini crocifissi e di donne violentate, sparse nella Rete come sussurri, non verificabili e non cancellabili, **fantasmi narrativi** che ottengono il risultato di fungere da supporto per le frange della popolazione disposte ad accogliere e condividere narrazioni più radicali mentre, a ben guardare, nello stesso tempo fungono da generatore di confusione per i moderati. L'esacerbazione del discorso utilizza vecchie tecniche su mezzi nuovi, con il risultato di rendere il conflitto civile quasi una soluzione inevitabile e catartica. Ogni tentativo di interpretazione più approfondita dei fatti o di verifica delle fonti e delle dinamiche da parte di esperti, analisti e fonti alternative, viene regolarmente stigmatizzata da entrambe le fazioni come il tentativo di inquinamento del discorso da parte di collaborazionisti, vecchi opinion leader disimpegnati, nichilisti scettici, disfattisti o agenti provocatori e, in quanto tali, collocati convintamente su uno dei due assi narrativi, solitamente quello dell'antagonista.

La stessa figura di **Vladimir Putin** rappresenta perfettamente una tipologia di personaggio intercambiabile, a seconda delle narrazioni soggiacenti, sia della fonte sia del destinatario. Egli rappresenta la sintesi di tre delle grandi narrazioni profonde proprie dell'identità culturale e politica russa: **autocrazia, ortodossia e nazione** (Sazonov, Mölder, Müür, Saumets, 2017). Le immagini di Putin intento a compiere esercizi di judo, a cavallo o mentre assiste alla parata militare in commemorazione della vittoria contro le truppe nazifasciste, rappresentano uno schema narrativo generale da riempire con il testo di ogni singola notizia, verso la quale vi sarà adesione – o meno – nel processo di appropriazione da parte del destinatario del messaggio, a seconda della corrispondenza alla narrazione ideologica di fondo. Non è da escludere, da questo punto di vista, che il culto della personalità inscenato sulla figura del Presidente russo possa agire anche come disinnesco di contronarrazioni più superficiali, come quelle nazionaliste, espresse come reazione da parte del Governo ucraino. È da evidenziare, infatti, che le strutture narrative profonde agiscono a volte in maniera non agevolmente prevedibile (Horkheimer, 1936).

Gaming e serial

Se, quindi, l'agitainment si fonda sul fatto che l'azione narrativa viene rappresentata direttamente nel medium attraverso una coinvolgente messinscena dei conflitti, allora possiamo dire che il **messaggio propagandistico** può plausibilmente essere visto come un'avventura, nella quale il soggetto, nei panni dell'eroe o del suo aiutante, intraprende il suo viaggio verso il ristabilimento dell'equilibrio.

Tale meccanismo è estremamente simile a quello proprio del **gaming** (Vos & Perreault, 2020), in particolare dei Massively Multiplayer Role Playing Game (MMORPG), in cui decine di milioni di utenti, attraverso i loro alter ego digitali, vivono quotidianamente le loro avventure e partecipano alle storie create per loro da migliaia di storyteller e programmatori, secondo canovacci a struttura fissa e intrecci variabili.

Un altro genere che, significativamente, accomuna tali dinamiche di produzione e di fruizione è quello dei serial. Non a caso, la serialità è anche uno degli elementi che più caratterizzano l'industria videoludica, oltre che cinematografica e televisiva (Allen & Van den Berg, 2014). Come sottolineato da **Bernardelli** (2016), il quale, a sua volta, si muove sulle tracce di Eco (1984), il desiderio della serialità scaturisce proprio da una fame di ridondanza da parte del fruitore. Lo schema narrativo seriale, determinando un abbattimento dell'entropia del messaggio, genera piacere o, comunque, una reazione emotiva nella ripetizione e, nello stesso tempo, nell'individuazione degli elementi di novità.

Wagner PMC: plot & com-plot

Per comprendere meglio quanto esposto e, più in generale, per analizzare la guerra ibrida come discorso, passeremo in rapida rassegna un particolare fenomeno nato in concomitanza con la guerra ucraina, che ha, in seguito, giocato un ruolo estremamente importante nella narrazione bellica degli anni successivi: ci riferiamo alla **Compagnia privata mercenaria** (PMC) Wagner.

La storia della Compagnia Wagner è ricostruibile attraverso una sorta di caccia al tesoro mediatica in Rete, nella quale è impossibile distinguere il mito fondativo – di quello che a tutti gli effetti è da considerarsi un personaggio narrativo collettivo – dalla realtà.

Nata nel 2014, proprio durante i combattimenti nel Donbass, la PMC Wagner è una tipica compagnia mercenaria corsara, priva di un riconoscimento ufficiale da parte della comunità internazionale e che, tuttavia, opera nell'interesse – e probabilmente al servizio – del **Ministero della difesa russo**. Da questo punto di vista, essa non è altro che un corpo paramilitare in linea con la vecchia dottrina sovietica degli interventi armati all'estero da parte di reparti paramilitari clandestini non direttamente riconducibili al Governo. Dopo essere stata protagonista in Crimea e nelle autoproclamate Repubbliche di Luhansk e Donetsk, la Wagner si sposta in **Siria** e diventa artefice sul campo di battaglia della lotta al gruppo terroristico internazionale Daesh. Comparsa per la prima volta sul sito **fontanka.ru**, un portale filorusso di reportage sulle forze armate russe, in breve tempo la compagnia diventa una sorta di **soggetto fantasmagorico** che esiste solo come entità non precisamente identificabile. Foto, testimonianze, articoli, reportage, video, imprese: la Wagner incarna –se incarnare può costituire un verbo adeguato – la nuova dottrina russa: la narrazione viene prima della realtà, il mito prima della storia, la forma prima della sostanza. Ai combattenti della Wagner, sia su siti più o meno accreditati, sia tramite testimonianze condivise sui social (delle quali molto spesso è impossibile stabilire la veridicità) vengono accreditate imprese narrative che li collocano dal lato dell'eroe della storia, come accadrebbe per il personaggio di un'avventura a fumetti: dall'uccisione di leader ribelli, alla liberazione di Palmira, città simbolo del terrorismo jihadista, fino all'uccisione di tre giornalisti russi nella Repubblica Centrafricana, intenti a compiere delle indagini sulle attività della Compagnia.

Senza alcun riscontro ufficiale, compiendo poche ricerche in Rete e sui social, la Wagner viene segnalata in tutto il mondo. Dalla Libia al Mozambico, dalla Germania al Nagorno-Karabakh, essa appare più simile alla **Spectre** dei romanzi su James Bond o allo Shield della serie a fumetti della Marvel che a un soggetto reale, ancora più simile al personaggio collettivo di un videogioco, nel quale il soggetto entra in contatto con la storia, ne viene coinvolto, può decidere di interagire, entrare nella storia stessa e cambiarla. Basta seguire la strada selciata dei mattoncini gialli, seguire le istruzioni, contattare qualcuno, superare dei test, accedere a mondi sempre più preclusi, per salire di livello e ritrovarsi, un giorno, su una delle foto disseminate sui social, su Telegram o su un

grande news magazine. Trovare esempi nella cinematografia, nei fumetti o nei romanzi che, in qualche modo, somiglino alla narrazione della Wagner significa reperire elenchi infiniti e, in realtà, sarebbe peraltro fuorviante. La Compagnia è portatrice di un fascino esattamente in linea con il target giovanile di riferimento al quale la sua propaganda è rivolta e quel linguaggio è quello dei videogame, dei social, del dark web, in **un territorio ibrido al confine fra complottismo, ideologia e mito**. In fin dei conti, non è importante diventare realmente un combattente effettivo: come il Vassili Zaitsev narrato ne "Il Nemico alle Porte", eroe leggendario della Grande Guerra Patriottica contro i nazisti, la Wagner è un eroe virtuale in cui può riconoscersi anche solo chi combatte seduto nella sua stanza al sicuro, a patto che ne condivida le gesta. Non a caso, ricercando "Wagner PMC" su Facebook, ci si imbatte nell'account di una community di giocatori online dello sparatutto 3d PlayersUnknownBattleGround.

In un'intervista comparsa sul portale **znak.com**, la presunta moglie di un mercenario della Wagner, rimasta vedova, mostra alla telecamera un video dello scontro a fuoco in cui il marito è rimasto ucciso, inviatole dai compagni sopravvissuti su un canale Telegram riservato. Il video raffigura una sorta di ripresa a infrarossi dall'alto, girata presumibilmente da un drone. "Sembra un videogioco", commenta l'intervistatore.

Eroi e antagonisti: le maschere di un conflitto

Il 22 marzo del 2020, il portale ucraino in lingua inglese "Kharkiv Human Rights Protection Group", sedicente portale di informazioni per i diritti umani in Ucraina, pubblica un articolo a firma di **Halya Coynash** dal titolo: "Luhansk militant court admits Donbas "hero" was a criminal who killed a family for money". Il presunto eroe in questione è **Alexey Mozgovoy**, leader separatista deceduto in un agguato il 23 maggio del 2015, comandante della **Compagnia "Prizrak"** meglio conosciuta in ambito internazionale come la Compagnia "Ghost" descritto come un efferato criminale di guerra, accusato e condannato da una Corte della stessa Repubblica Popolare di Luhansk nonostante – si riporta – i media russi abbiano censurato la notizia. Lo stesso racconto viene ripreso da altri portali filoccidentali, incluso il Telegraph.

Il 4 maggio del 2021, un reportage scritto per commemorare la morte dello stesso Alexey Mozgovoy compare sul portale filorusso lenta.ru. L'articolo ha per titolo "Debaltseve – una delle località teatro della guerra civile-ndr) – somigliava a Stalingrado" e racconta la storia del capo guerrigliero separatista Alexey, appassionato di poesia, innamorato del suo popolo, amante della natura e degli animali. L'eroe della storia si imbatte in alcuni cadaveri di giovanissimi soldati dell'esercito regolare, morti dopo un combattimento dal quale egli è uscito vincitore. Così, **l'eroe recupera i cellulari delle vittime e contatta le loro famiglie**. A un padre che gli chiede ragione del figlio ucciso, Alexey Mozgovoy risponde: "Non l'abbiamo ucciso noi. Noi siamo la Verkhovna Rada (il partito separatista filorusso-ndr). Sono sempre felice di morire per la verità. E loro ti mentono tutto il tempo. Il Governo non ti ha nemmeno informato che i tuoi figli sono già stati uccisi! Ti viene sempre detto che stai vincendo. Ti mentono continuamente. Accendi il cervello, amico!".

Conclusioni: the willing suspension of disbelief

La sostanza del legame fra la fonte e il destinatario, a questo punto, non risiede nell'aderenza – o meno – della notizia a una verificabile realtà, quanto in un fatto ancor più interessante ai nostri fini: se il discorso soggiacente al testo, il suo codice, appunto, sia riconosciuto o meno dalla fonte

come affidabile. Come se, in altre parole, il rapporto che si stabilisce fra i due interlocutori del messaggio fosse del tutto simile al meccanismo di sospensione volontaria dell'incredulità (willing suspension of desbelief) (Coleridge, 1817), proprio del genere della fiction. Perché tale meccanismo si inneschi, ciò che deve funzionare, dunque, è **l'interazione dei codici**, non la verità fattuale.

Come suggerito da Veebel (2015), ci si trova di fronte ad alcune contraddizioni quando si tenta di analizzare una guerra di informazioni. Innanzitutto, non è agevole scorgerla, poiché il prerequisito di una information war ben condotta è proprio quello di essere invisibile; né è agevole verificarne gli effetti, poiché risulta molto critico, anche in presenza di un'indagine approfondita, basata su dati quantitativi, coglierne le dinamiche. Le tendenze più recenti della comunicazione politica una disintermediazione sempre più marcata, unitamente all'uso diffuso delle applicazioni di messaggistica interpersonale, prime fra tutte Telegram e Whatsapp, per loro natura difficilmente tracciabili rendono ancora più opaca l'osservazione dei fenomeni comunicativi. D'altronde, come sottolineato da Nye (2012), "la migliore propaganda non è la propaganda: la credibilità è una risorsa scarsa".

Eppure, la stessa **natura mimetica della propaganda** – ingrediente sempre più pervasivo nei conflitti internazionali in contesti in cui gli Stati occupano solo una parte degli attori in campo – può condurre verso un'aberrazione. La **costruzione di mondi immaginari**, sempre più raffinatamente costruiti, potrebbe rendere a un certo punto superfluo il ricorso a strumenti di controllo della verità, poiché, se possiamo tecnicamente riprodurne una ottimale, a nostro uso e consumo, per quale motivo indagarne altre? Una volta che la costruzione artificiale della verità è stata messa in moto e la macchina della verosimiglianza potenziata al meglio delle sue possibilità tecniche, ogni nuovo pezzo di notizia potrebbe condurci sempre più lontano dalla realtà, con la conseguenza che a venire meno potrebbe essere proprio la nostra stessa **capacità autocritica** e di ricerca di una soluzione al conflitto.

D'altronde, il fascino del falso, come descritto da Eco (2002), è e resta un'arma seducente.

Eppure quest'ovvia considerazione non deve portarci a concludere che non esiste un criterio di verità, e che racconti detti falsi e racconti che riteniamo oggi veri si equivalgano, appartenendo entrambi al genere letterario della finzione narrativa. Esiste una pratica della verificazione che si basa sul lavoro lento, collettivo, pubblico di quella che Charles Sanders Peirce chiamava la Comunità. […] In fondo, il primo dovere dell'uomo di cultura è quello di tenersi all'erta per riscrivere ogni giorno l'enciclopedia.

Le bugie – già! – che si chiamano anche storie.

Ma non ha mica nessuna colpa, sa? di non esser vera, questa storia.

Importa assai che non sia vera; se poi è bella!

L. Pirandello, Vestire gli ignudi, Atto III

Bibliografia

Allen, R., & Van den Berg, T. (2014). Serialization in Popular Culture. Routledge.

Bernardelli, A. (2016). Eco e le forme della narrazione seriale. Alcuni spunti per una discussione. In Between. 11. DOI:10.13125/2039-6597/2498

Brown, K. (2013). Agitprop in Soviet Russia. In Constructing the Past. 14.

Ciracì F. (2021). Per una teoria critica del digitale: fake-news e postverità alla luce della logica della verosimiglianza. In F. Ciracì, Fedriga R. & Marras C. Filosofia digitale. Mimesis: 87-112.

Eco, U., (1964). Apocalittici e integrati. Comunicazioni di massa e teorie della cultura di massa. Bompiani.

Eco, U. (2002). La Forza del Falso. In Sulla Letteratura, RCS.

Foucault M., (1997), L'archeologia del sapere. Una metodologia per la storia della cultura. RCS.

Gerasimov, V. (2013) Ценность науки в предвидении. Новые вызовы требуют переосмыслить формы и способы ведения боевых действий. https://vpk-news.ru/articles/14632.

Giungato L. (2020). Niente sarà più come prima. Il Covid-19 come narrazione apocalittica di successo. Nothing will ever be the same: COVID 19, an apocalyptic narrative of success". H-ermes. Journal of Communication. 16:99-122. DOI.org/10.1285/i22840753n16p99

Horkheimer, M. (1936). Studien über Autorität und familie. Alcan, Parigi.

Nemet W. J. (2002). Future war and Chechnya: a case for hybrid warfare. Calhoun. The NPS Institutional Archive.

http://hdl.handle.net/10945/5865.

Nye, J. (2012). China's Soft Power Deficit To catch up, its politics must unleash the many talents of its civil society. In The Wall Street Journal.

Pogrebinskiy, M. (2015). Russians in Ukraine: Before and After Euromaidan. In Ukraine and Russia: People, Politics, Propaganda and Perspectives. Agnieszka Pikulicka-Wilczewska & Richard Sakwa.

Rumiz P., (2000). Maschere Per un Massacro. Editori Riuniti.

Sazonov, V.; Mölder, H.; Müür, K.; Saumets, A. (2017). Russian Information Operations Against Ukrainian Armed Forces and Ukrainian Countermeasures (2014–2015). ENDC Occasional Papers. 6.

Tolz V. & Teper. (2018). Broadcasting agitainment: a new media strategy of Putin's third presidency: Post-Soviet Affairs, 34:4, 213-227, DOI: 10.1080/1060586X.2018.1459023.

Veebel, V. (2015). From Psychological defence to Propaganda War. Latvian Institute of International Affairs. <http://liia.lv/en/blogs/from-psychological-defence-to-propaganda-war/>.

Vos T.P & Perreault G.P. (2020). The discursive construction of the gamification of journalism. In Convergence.
26(3):470-485. DOI:10.1177/1354856520909542.

https://en.wikipedia.org/wiki/Wagner_Group

https://www.fontanka.ru/2015/10/16/118/

https://lenta.ru/news/2015/01/02/batman/

https://www.kyivpost.com/ukraine-politics/list-separatist-leaders-killed-donbas-motorola.html

https://www.economist.com/europe/2017/11/02/how-wagner-came-to-syria

https://www.bbc.com/news/world-europe-45030087

https://www.znak.com/2018-02-13/intervyu_s_suprugoy_pogibshego_v_sirii_uralskogo_boyca_chvk_vagnera

http://khpg.org/en/1586112422

https://www.telegraph.co.uk/news/worldnews/europe/ukraine/11201446/Rebels-in-Ukraine-post-video-of-peoples-court-sentencing-man-to-death.html

https://lenta.ru/articles/2021/05/04/mozgovoi/

https://jamestown.org/program/war-business-and-hybrid-warfare-the-case-of-the-wagner-private-military-company-part-two/

https://digitalcommons.iwu.edu/constructing/vol14/iss1/4

Deepfake: così ti rovino la web reputation aziendale. Rischi e strategie di difesa

La tutela dell'identità della persona giuridica online costituisce una protezione del marchio della stessa e si configura come vero e proprio asset intangibile. Il rischio e il danno reputazionale sono perciò elementi cruciali da considerare in ottica di cybersecurity. Come difendersi da attacchi deepfake e cosa si rischia

Di **Simone Bonavita**, Professore a contratto in "Sensitive Personal data Processing" at Università degli Studi di Milano e **Elisabetta Stringhi**, Lawyer Trainee at Perani Pozzi Associati

Il fenomeno del **deepfake** si presta facilmente ad attacchi reputazionali a danno di una persona giuridica, così vanificando gli sforzi compiuti per curarne la **web reputation,** un asset oggi cruciale.

Ecco perché cercheremo di approfondire di seguito i rischi derivanti da un uso strumentale dell'IA per cagionare un danno reputazionale, nonché possibili strategie per prevenire e rimediare all'incidente.

Il deepfake e la minaccia alla web reputation

Come noto, l'evoluzione dell'intelligenza artificiale (IA) consente allo stato dell'arte attuale di attaccare la reputazione online di una persona giuridica, con conseguenze notevoli per l'identità online, così vanificando gli sforzi compiuti fino a quel momento per proteggerla.

Le tecniche di IA di **autoencoder** (T. T. Nguyen et al., 2020; Hasson et al., 2020) o di **Generative Adversarial Networks** (I Goodfellow et al., 2014) consentono di sintetizzare file audio, fotografici e/o audiovisivi accurati e realistici agli occhi degli utenti.

Per quanto riguarda i file fotografici, la tecnologia **StarGAN** (Y. M. Choi et al., 2018) realizzava immagini facilmente riconoscibili come false. Tuttavia, le architetture **StyleGAN** (T. Karras et al., 2019) e **StyleGAN 2** (T. Karras et al., 2020) hanno consentito di generare contenuti visivi estremamente credibili, come una sovrapposizione di volti su corpi differenti ("face swap") o una sintesi di immagini originali ("image-generation"). Con tale tecnologia possono essere sintetizzati anche nuovi contenuti audio ("speech synthesis"). Inoltre, sono noti anche i "shallowfakes", ossia i contenuti audio-visivi di rozza manifattura, mediante il ricorso a tecniche meno sofisticate di IA.

Questo fenomeno è comunemente definito con la parola inglese "deepfake", termine che combina "deep" e "fake", dal nome dell'omonimo utente di Reddit "Deepfakes" che ne ha reso popolare la creazione (H. Ajder, et al., 2019). Il deepfake è stato autorevolmente definito come una "tecnica che utilizza l'intelligenza artificiale per **combinare e sovrapporre immagini o video originali**, ritraenti una persona, con quelli ritraenti qualcun altro, o per generare immagini o video completamente falsi e difficilmente riconoscibili come falsi" (G. Ziccardi, P. Perri, 2020).

Tale tecnica si presta ad essere utilizzata per molteplici scopi come, a titolo illustrativo, la produzione creativa ed artistica (T. Shen et al., 2018).

In ambito cinematografico, il deepfake potrà consentire di agire in **post-produzione** per creare immagini di attori e/o attrici scomparse durante il corso della ripresa del film, andando a sostituire la tecnica di CGI. Difatti, l'utente anonimo "deepfakes" ha utilizzato le Generative Adversarial Network" proprio per ricreare celebri scene di "Guerre Stellari" ritraenti l'amato personaggio Leia Organa (H. Ajder, et al., 2019).

Risulta allora evidente come la tecnica deepfake possa essere utilizzata da potenziali attaccanti per ledere la web reputation non soltanto di persone fisiche, ma anche di persone giuridiche. Come altresì emerso dal **Rapporto CLUSIT** del 2020 (F. Bertoni, 2020), nel lungo periodo è stimato che gli attacchi fondati sulla combinazione di tecniche di ingegneria sociale e di intelligenza artificiale come il deepfake cresceranno in maniera sostenuta, prettamente a danno di organizzazioni economiche.

Lo scopo del presente contributo è di **approfondire il ricorso alla tecnologia deepfake** per attaccare la reputazione delle persone giuridiche, analizzandone lo stato dell'arte attuale, le prospettive tecniche e giuridiche, come possibili strategie preventive e rimediali all'incidente. Le tecniche deepfake possono essere utilizzate per perpetrare un attacco reputazionale al rappresentante legale e/o al CEO della persona giuridica, come per perfezionare una impersonificazione diretta della stessa, per lo svolgimento di operazioni di tipo bancario, anche di home-banking (Bonavita et al., 2021). In questa analisi, esamineremo la prima tipologia di attacco.

L'attacco reputazionale alle persone giuridiche tramite deepfake: tecniche e prospettive

Il rischio primario di attacco reputazionale alle persone giuridiche tramite deepfake è rappresentato dalla diffusione di immagini e/o video ritraenti una o più persone fisiche (riconducibili all'organizzazione economica), a seguito della combinazione e sovrapposizione di immagini o video originali oppure della generazione di immagini o video falsi e difficilmente riconoscibili come tali, in atteggiamenti o contesti lesivi della reputazione.

Nonostante sia ancora una tecnica relativamente poco diffusa, è ragionevole supporre che il deepfake sarà sempre più utilizzato dagli attaccanti in ragione di diversi fattori:

l'agevole ricorso a **software gratuiti** per sintetizzare tali file;

la presenza di un vero e proprio **marketplace online** nel quale è possibile acquistare o commissionare la realizzazione di immagini e/o video deepfake;

lo sviluppo e la distribuzione di applicativi per smartphone con cui si possono semplicemente creare (A. Robertson, 2018, A. Carman, 2019 Agence france-presse, 2019);

la mole di **dati digitali disponibili online relativi alla persona**, tali per cui risulta possibile ricostruire nel dettaglio ogni singolo aspetto della stessa, fino alla creazione di un vero e proprio "corpo digitale", potenzialmente separato dal "corpo fisico" (cfr., inter alia, S. Bonavita, E. Stringhi, 2021; S. Bonavita, 2017; G. Ziccardi, 2017).

Per questo, è possibile parlare di un **processo di commodification dei deepfake**, se non, addirittura, di deepfake-as-a-service, essendosi venuto a creare un mercato illecito annesso (M. Rosario Fuentes, 2020). Sebbene sia stato evidenziato (EUROPOL, 2020) come, ad oggi, casi documentati di ricorso alla tecnologia deepfake per commettere attacchi informatici ai danni di organizzazioni economiche siano relativamente limitati per una serie di ragioni, come i) la durata breve di tali contenuti, segnalati celermente; ii) la necessità di competenze altamente tecniche, iii) la presenza di metodi alternativi "tradizionali" per perfezionare gli attacchi, comunque preme sottolineare tale fenomeno è in espansione. Quanto più tale tecnologia diviene nota al grande pubblico e capace di sintetizzare contenuti convincenti, tanto più gli attaccanti non necessiteranno di elevate conoscenze e competenze da un punto di vista tecnico e, dunque, i deepfake diventeranno degli **efficaci strumenti di attacco** della web reputation.

Alla luce di ciò, la tecnologia deepfake potrebbe essere agevolmente strumentalizzata da attaccanti organizzati al fine di perfezionare delle azioni di **ingegneria sociale** per danneggiare la reputazione di una società target e, quindi, manipolare il mercato azionario di riferimento ovvero realizzare degli schemi fraudolenti.

Basti pensare, ad esempio, al **furto di identità** di un amministratore delegato di una società che viene ritratto artificiosamente in un video, successivamente diffuso a mezzo social network, mentre rilascia dichiarazioni offensive, a carattere razzista, oppure false, in relazione a presunte perdite finanziarie subite o ad una violazione di dati mai avvenuta.

Ancora, in riferimento ad un episodio avvenuto concretamente, basti pensare alla diffusione sul social network Instagram di un video, sintetizzato con le GANs, ritraente Mark Zuckerberg mentre dichiarava di essere affiliato all'organizzazione criminosa "Spectre" (H. Ajder, et al., 2019). Seppur tale video fosse stato realizzato per scopi prettamente artistici e satirici, l'episodio è dimostrativo della reale potenzialità della tecnologia deepfake ad essere sfruttata per causare un danno reputazionale ad un'organizzazione economica. Nonostante tali contenuti messi in circolazione nei social media possano essere naturalmente riconosciuti e dichiarati come **falsi**, nonché eventualmente rimossi tramite procedura di notice-and-take-down, ove ne ricorrano i presupposti, in conformità all'art. 512 DMCA, bisogna comunque evidenziare come si registrerebbe una flessione del valore economico attribuibile alla società *target* a seguito di tale incidente.

Un altro ipotetico scenario di rischio reputazionale in ambito corporate vede **l'estorsione** tramite la prospettazione di un attacco alla web reputation dell'organizzazione economica con materiale deepfake, azione che peraltro potrebbe essere – secondo alcune teorie - combinata ad un attacco informatico di tipo ransomware verso il pagamento di un corrispettivo in valuta o, eventualmente, in criptovalute (M. R. Fuentes, 2020).

La pericolosità di un attacco reputazionale di tipo deepfake deve, pertanto, essere inquadrata alla luce delle tecniche di ingegneria sociale ben note dagli attaccanti nonché dagli stessi studiate e riadattate al particolare contesto dell'organizzazione economica target.

I danni conseguenti ad un incidente deepfake

Nel contesto attuale, la tutela dell'identità della persona giuridica online costituisce al contempo una protezione del marchio della stessa, pertanto, configurandosi come vero e proprio asset

intangibile. Il rischio ed il danno reputazionale sono perciò elementi cruciali da considerare in ottica di **cybersecurity** in ambito corporate. Si intende, difatti, con "rischio reputazionale" quel pericolo, attuale o futuro, di flessione del capitale o degli utili derivante da una percezione negativa dell'immagine dell'organizzazione economica da parte di vari soggetti: clienti, controparti, azionisti, investitori, autorità di vigilanza.

Il rischio reputazionale costituisce dunque un rischio di cosiddetto secondo livello, perché derivante da eventi negativi, a loro volta causati da altri rischi verificatisi in concreto, come ad esempio irregolarità e perdite finanziarie, violazione di dati, incidenti di sicurezza, questioni e violazioni etiche. Secondo un report del World Economic Forum, **almeno il 25% del valore di mercato di una società è direttamente attribuibile alla sua reputazione**. Un'altra indagine del 2017 (KPMG, 2017) osserva come il rischio reputazionale sia il più temuto dagli amministratori delegati e dai consigli di amministrazione, in ben 10 paesi differenti e 11 diversi settori dell'industria. Analogamente, il Reputation Institute posiziona il rischio reputazionale al **secondo posto tra i rischi per gli investitori** (Reputation institute). Sempre secondo il Global Risk Survey (Aon, 2019), il danno da incidente reputazionale è al primo posto per quanto riguarda l'industria dei c.d. professional services, seguito dagli attacchi informatici e dall'incapacità di trattenere o attrarre valide risorse. Bisogna, infatti, sottolineare come, seppur il danno reputazionale sia sottovalutato a causa della propria natura "intangibile", esso abbia un impatto diretto sul valore della società per gli azionisti, incidendo sostanzialmente sulle relative prospettive finanziarie.

Gli attacchi reputazionali perpetrati tramite deepfake sollevano, inevitabilmente, questioni e riflessioni connesse alla diffusione di notizie non veritiere, capaci di ledere l'identità online e la web reputation delle persone giuridiche e dei brand o, più sinteticamente, legate al tema delle **fake news** in ambito commerciale ed industriale (cfr. G. Ziccardi, S. Bonavita, A. Barchiesti, Fake news in ambito commerciale ed industriale, conferenza organizzata da ISLC e Reputation Manager). Se con l'espressione "fake news" si identifica una particolare fattispecie, in cui risulta determinante la trasmissione delle informazioni a mezzo Internet così connotando la notizia diffusa, con il depfake tuttavia si delineano scenari di distorsione della realtà ben più profonde e senza precedenti. Difatti, tali attacchi reputazionali non soltanto minano la capacità di chi osserva di distinguere il contenuto audio-visivo falso da quello vero, ma anche di fidarsi di immagini e/o video effettivamente reali, in quel paradosso definito come "liar's dividend" (Chesney et al., 2020). Trattandosi di contenuti capaci di distorcere la rappresentazione dei fatti e di travisare un'identità online, la relativa diffusione costituisce fonte di responsabilità civile qualora sia cagionato un danno ingiusto.

A titolo esemplificativo, basti pensare alla condivisione tramite social network di un video deepfake relativo ad una **finta sponsorizzazione di un brand** che, anziché alla sua promozione, miri alla sua rovina, ad esempio esaltandone difetti e/o distorcendone le qualità. Si tratterebbe di informazioni false e diffamatorie relative ai prodotti e/o ai servizi offerti dalla persona giuridica attaccata, la quale patirebbe danni reputazionali ingiusti.

Un simile attacco potrebbe integrare il reato di **diffamazione aggravata d**all'uso dei social network per la diffusione dei contenuti audio-visivi diffamatori e non veritieri, nonché risultare suscettibile di tutela cautelare, come in un caso di lesione della reputazione tramite TripAdvisor (cfr. Tribunale di Venezia, ordinanza del 24 febbraio 2015) (L. Vizzoni). In circostanze analoghe, i giudici di merito hanno condannato l'autore del reato al risarcimento del danno conseguente alla condotta (Tribunale di Lecce, 12 settembre 2018).

Ad esempio, qualora le immagini di un CEO e/o di uno sponsor ufficiale siano distorte con tecnica deepfake e diffuse a mezzo social network, la diffusione di tali contenuti potrebbe configurare un atto di concorrenza sleale, perché il relativo brand potrebbe registrare una perdita finanziaria a seguito dell'incidente.

Come si evince da questi esempi, in questa sede trovano applicazione anche le norme relative alla disciplina della **proprietà intellettuale ed industriale**, come attinenti la tutela della concorrenza. Ciò è evidente soprattutto se si considera che ciascun utente in rete non fruisce passivamente dei contenuti online, ma al contrario è attivamente creatore e diffusore di post, foto, video ed informazioni – potenzialmente false. Così la diffusione di deepfake può alterare il comportamento economico, oltre che influenzare le scelte dei consumatori, a danno del brand vittima dell'attacco reputazionale.

Strategie per prevenire e rispondere a un attacco reputazionale deepfake

Dato che le conseguenze in termini reputazionali e finanziari di un attacco basato su deepfake possono essere quantitativamente ingenti e persistenti nel tempo, proteggere il brand e la reputazione online dell'organizzazione economica significa in primis adottare una strategia preventiva, che si articola in una serie di azioni.

Specialmente nella prospettiva di proteggere il bene intangibile della web reputation a lungo termine (Aon, 2019), sia in ottica di percezione pubblica del brand come di monitoraggio dell'andamento del mercato, è cruciale implementare un controllo costante dei canali comunicativi aperti: ciò ricomprende senz'altro un monitoraggio continuo delle pagine dei social network ufficiali del marchio, senza escludere tuttavia un'analisi di siti web di testate di giornale o di esperti del settore, come i profili social media di divulgatori, sponsor e/o influencer del settore. In tal senso, è opportuno che l'organizzazione economica adotti ed implementi, anche ricorrendo a società di consulenza e/o esperti esterni, **capacità di intelligence** nello svolgimento di tale attività di monitoraggio. È altrettanto utile che la reputazione online sia monitorata in modalità automatizzata, con il ricorso ad appositi software. L'attività di monitoraggio e di identificazione di eventuali contenuti deepfake lesivi del brand non possono prescindere da una preparazione tecnica nel campo dell'image processing (S. Aterno, 2019).

Una presenza attiva sui canali comunicativi social media è cruciale per mantenere il controllo sulla reputazione online dell'organizzazione e/o del marchio target. Difatti, la costruzione di un'identità online definita e solida in relazione al brand potrebbe ridurre il rischio che consumatori, utenti e stakeholder ritengano veritieri i contenuti deepfake diffusi, come potrebbe prevenire un sentimento di scetticismo nei confronti di contenuti pubblicati dagli account ufficiali. In particolare, l'avviamento di **campagne comunicative mirate** per coltivare e promuovere il sostegno degli stakeholder potrebbe minimizzare il rischio reputazionale derivante dalla diffusione di deepfake fortemente offensivi come, ad esempio, il rilascio di un video ritraente l'amministratore delegato in atteggiamenti razzisti.

Infine, è fondamentale che all'interno dell'organizzazione economica venga coltivata una **cultura di consapevolezza del rischio reputazionale**, soprattutto in relazione ad azioni coordinate di ingegneria sociale, tramite strumenti, formazione e simulazioni interne. Nello specifico, soci, dipendenti e collaboratori dovrebbero essere sensibilizzati all'uso cauto dei social network, perché

non svelino dati personali e informazioni sensibili a potenziali attaccanti, come delle app di messaggistica, evitando di condividere immagini, video e/o audio personali più del dovuto.

La consapevolezza, da un lato delle tecniche di ingegneria sociale, dall'altro del potenziale riutilizzo dei media condivisi per sintetizzare deepfake, dovrebbe essere al centro di questa cultura di consapevolezza del rischio reputazionale.

Per quanto concerne, invece, l'adozione di una **strategia comunicativa successiva e conseguente** al verificarsi di un incidente reputazionale basato sul deepfake, nonché durante tutto il corso della crisi reputazionale, è opportuno che l'organizzazione economica adotti una linea trasparente e proattiva. Anziché consentire ad altri soggetti, come attaccanti e/o concorrenti, di esprimersi sull'incidente, è bene che sia il brand stesso a spiegare l'avvenuto, smentendone la natura di falso profondo e chiarendo che non trattasi di contenuto ufficiale. Ad ogni modo, da tale comunicazione dovrebbero emergere una serie di elementi, in primis la volontà di esprimere pubblicamente delle scuse, per suscitare l'empatia nei confronti del destinatario della comunicazione; l'assunzione di responsabilità pubblica, senza risultare né impersonale né de-responsabilizzante, al fine di creare fiducia nei destinatari o, comunque, evitare una perdita di credibilità e affidabilità del brand; fornire le informazioni utili per illustrare il problema verificatosi ed eventualmente risolverlo, come contatti di riferimento e/o rimedi predisposti.

Infine, risulta ragionevole che le organizzazioni economiche inizino a valutare la **stipulazione di un contratto assicurativo** per essere mantenute indenni dalle perdite di profitto lordo causate dalla riduzione del volume di affari, nonché dall'aumento dei costi di esercizio, ivi incluse quelle spese imprescindibili per la comunicazione da svolgere nella gestione di una crisi; la comunicazione per ricostruire la reputazione online ed il brand; la cancellazione e la de-indicizzazione dei contenuti deepfake; la (eventuale) tutela in sede civile e/o penale; tenere monitorato il gradimento del marchio.

Considerazioni conclusive e prospettive future

Come approfondito, il deepfake si presta agilmente ad offendere la reputazione online di una persona giuridica e colpire i brand. Per fronteggiare al meglio le sfide che tale tecnologia ci riserva, soprattutto in ambito **commerciale ed industriale**, sarà essenziale compiere uno sforzo per accrescere la consapevolezza della sua esistenza deepfake ed attuare quelle strategie cruciali per proteggere il brand da tentativi di manipolazione e distorsione (Europol, 2019).

Bibliografia

A. Carman, 2019. *The Verge*. "FaceApp is back and so are privacy concerns", disponibile al link https://www.theverge.com/2019/7/17/20697771/faceapp-privacy-concerns-ios-android-old-age-filter-russia (sito web online e consultato il 26 aprile 2021).

A. Robertson, 2018. *The Verge.* "I'm using AI to face-Swap Elon Musk and Jeff Bezos, and I'm really bad at it", disponibile al link https://www.theverge.com/2018/2/11/16992986/fakeapp-deepfakes-ai-face-swapping (sito web online e consultato il 26 aprile 2021.

Agence france-presse, 2019. *The Guardian. The Guardian.* "Chinese deepfake app Zao sparks privacy row after going viral", disponibile al link

https://www.theguardian.com/technology/2019/sep/02/chinese-face-swap-app-zaotriggers-privacy-fears-viral (sito web online e consultato il 26 aprile 2021).

F. Bertoni (2020). "L'impatto dei deepfake sulla sicurezza delle organizzazioni economiche". *Rapporto CLUSIT 2020 sulla sicurezza ICT in Italia,* pp. 155-171.

Chesney, Citron, Farid (2020). "All's Clear for Deepfakes: Think Again", *Lawfare: Hard National Security Choices*, disponibile all'indirizzo <https://www.lawfareblog.com/alls-clear-deepfakes-think-again> (sito web online e consultato il 27 aprile 2021).

D. Lee, Deepfake Salvador Dalí takes selfies with museum visitors,10 maggio 2019, sito web online, consultato e disponibile all'URL: https://www.theverge.com/2019/5/10/18540953/salvador-dali-lives-deepfake-museum.

Europol (2018). "French Coder Who Helped Extort British Company Arrested in Thailand", disponibile all'indirizzo: <https://www.europol.europa.eu/newsroom/news/french-coder-who-helped-extort-british-company-arrested-inthailand>.

Europol European Cybercrime Centre (2019). DeepFake. Europol Platform for Experts – EPE.

Europol, Trend Micro, Unicri. (2020). *Criminal Misuse of AI and Deepfakes*, The Hague, The Netherlands.

G. Ziccardi (2017). "La morte nell'era dei "dati eterni": che ne sarà del nostro corpo digitale", in *Agenda Digitale*. Disponibile all'URL: <https://www.agendadigitale.eu/cultura-digitale/la-morte-nellera-dei-dati-eterni-che-ne-sara-del-nostro-corpo-digitale/>.

G. Ziccardi, P. Perri. *Dizionario Legal Tech: Informatica Giuridica, Protezione dei Dati, Investigazioni Digitali, Criminalità Informatica, Cybersecurity e Digital Transformation Law,* Milano, Giuffré Francis Levebvre, 2020, p. 333.

H. Ajder, G. Patrini, F. Cavalli, L. Cullen (2019). *The State of Deepfakes: Landscape, Threats, and Impact.*

I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, & Y. Bengio (2014). "Generative Adversarial Nets". *Advances in Neural Information Processing Systems,* 2672–2680.

L. Vizzoni, Recensioni genuine su TripAdvisor: quali responsabilità? Responsabilità Civile e Previdenza, fasc. 2, 1 febbraio 2018, p. 706).

M. Rosario Fuentes. (2020). Trend Micro. *Shifts in Underground Markets: Past, Present, and Future,* disponibile all'indirizzo https://www.trendmicro.com/vinfo/us/security/news/cybercrime-and-digital-threats/trading-in-the-dark (sito web online e consultato il 26 aprile 2021). Si stima, inoltre, che sarà in crescita anche il mercato relativo all'offeta di strumenti e di servizi online di noleggio.

S. Aterno, Deepfake in tribunale: ecco come la digital forensics smaschera il falso. Agenda Digitale. 6 novembre 2019. Disponibile all'URL: <Deepfake in tribunale: ecco come la digital forensics smaschera il falso | Agenda Digitale>.

S. Barnes, Artist uses AI to generate realistic faces of subjects from world's most iconic paintings, 15 luglio 2020, sito web online, consultato e disponibile all'URL: <https://mymodernmet.com/denis-shiryaev-neural-network-art/>.

S. Bonavita (2017). "Le ragioni dell'oblio", in *Ciberspazio e diritto,* (1-2017), 85-111.

S. Bonavita, A. Cortina, E. Stringhi, (2021). "Conosci il tuo nemico": un primo approccio tassonomico ai principali attacchi informatici nel settore del cybercrime bancario e finanziario", in *Ciberspazio e diritto*, vol. 21, n. 66 (3-2020), 457-501.

S. Bonavita, E. Stringhi (2021). "Identità digitale e personalità on-line: i profili sostanziali di tutela", in *Orientamenti della giurisprudenza,* Il Quotidiano Giuridico. Disponibile all'URL: <https://www.quotidianogiuridico.it/documents/2021/05/03/identita-digitale-e-personalita-on-line-i-profili-sostanziali-di-tutela>.

T. Karras, S. Laine, & T. Aila (2019). "A style-based generator architecture for generative adversarial networks". *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4401–4410.

T. Karras, S. Laine, M. Aittala, A. Hellsten, J. Lehtinen, & T. Aila (2020). "Analyzing and improving the image quality of StyleGAN". *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* 8110–8119.

T. Shen, R. Liu, J. Bai, Z. Li, "Deep Fakes" using Generative Adversarial Networks (GAN), ECE228 and SIO209 Machine learning for physical applications, vol. 16, Spring 2018.

T. T. Nguyen, C. M. Nguyen, D. T. Nguyen, S. Nahavandi, (2020). "Deep Learning for Deepfakes Creation and Detection: A Survey", in *Computer Vision and Pattern Recognition*. Disponibile all'URL: <https://arxiv.org/pdf/1909.11573.pdf>.

Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, & J. Choo. (2018). "Stargan: Unified generative adversarial networks for multi-domain image-to-image translation". *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* 8789–8797.

Fake news, così scienza e informazione interpersonale perdono credibilità

Le fake news causano perdita di credibilità nelle reti sociali di prossimità e nella scienza. Come si diffondono, qual è il ruolo dell'ignoranza attiva, come contrastarle

Di **Fabio Ciracì,** Docente di Storia della Filosofia Italiana e Informatica Umanistica - Università degli Studi del Salento

Uno dei maggiori pericoli contemporanei è la **progressiva erosione della credibilità** che investe tutto il mondo dell'informazione, a causa di numerosi fattori, tra cui il **sovraccarico informativo (information overload) e la mancanza di autorevolezza (lack of authority) delle fonti** scientifiche e istituzionali.

Se la lacuna di autorevolezza richiede una capacità di scelta delle fonti informazionali, tecniche di selezione e strumenti di filtering sempre più sofisticati, il sovraccarico informativo invece pone radicali problemi di orientamento nel vasto mare dell'informazione digitale.

La credibilità è la moneta sonante della comunicazione, la condizione di possibilità di qualsiasi impresa o commercio, la base necessaria per poter fondare relazioni sociali basate sulla fiducia, la condizione stessa del "munus" comunitario. Minare la credibilità degli attori sociali vuol dire sempre minare una società nelle sue fondamenta.

Chi lavora con l'informazione sa bene quanto sia complicato conquistare la fiducia dei lettori, siano di un giornale cartaceo o digitale, perché **è sulla base della credibilità che si acquisisce autorevolezza** ed essa si accresce ogni qualvolta esce vittoriosa dalla prova dei fatti, secondo quella che è abitualmente chiamata digital reputation.

Essere credibili, inoltre, rende visibili, permette alle testate giornalistiche di emergere fra i flutti dello tsunami informativo, di orientare l'opinione pubblica.

Analizzeremo l'erosione della credibilità in due campi, apparentemente lontani: le reti sociali di prossimità prodotte dalle app di messaggistica istantanea, con la conseguente perdita di credibilità dell'informazione interpersonale; la sfiducia nel mondo scientifico.

Come le fake news erodono la credibilità nei rapporti interpersonali

Come funziona la viralità delle fake news nelle nicchie reticolari di social-messaging istantanei come WhatsApp o Telegram?

Non si tratta semplicemente di mappare la circolazione di informazione cattiva (misinformation)[1] o mendace (disinformation), magari attraverso l'azione di bot, che consentirebbe al più di fotografare una situazione di fatto, descrivendo le forze agenti senza però individuarne le cause. Si

tratta invece di comprendere quali meccanismi presiedono alla diffusione delle fake news, con particolare attenzione alla questione della credibilità, all'interno delle reti di prossimità che si formano attraverso le app di messaggistica.

Un'ipotesi del presente contributo è che **le fake news, sviluppandosi in senso crossmediale,** ovvero tra piattaforme mediali diverse per struttura tecnologica, ibridandosi e agendo con un effetto domino nell'infosfera, **erodono in maniera sempre maggiore le relazioni fiduciarie fra i loro appartenenti**.

Le fake news circolano velocemente, senza soluzione di continuità, attraverso le diverse piattaforme digitali, passando da social media, come Facebook, ad app di micro-blogging come Twitter o a app di messaging, come WhatsApp, **con grande danno della "credibilità forte", quella interpersonale e relativa ai rapporti di prossimità** con persone della nostra cerchia dunberiana, differenti dai cosiddetti "legami deboli" dei social media su cui si basa la nostra rete digitale, secondo la logica dei gradi di separazione.

I fenomeni di diffusione di fake news su social media e via app di messaggistica sono simili ma diversi, proprio per la fiducia che si ripone nelle nicchie reticolari di queste ultime, per esempio di WhatsApp, che riproducono un network "di prossimità", di persone che in genere si conoscono e con cui esiste un rapporto personale: si dà il proprio numero di cellulare a persone fidate o accreditate, perché ci si frequenta spesso per lavoro o si condividono situazioni di vita ricorrenti.

Sembrerebbe quindi che le reti di prossimità attivate via numero di cellulare ispirino maggiore fiducia e possibilità di controllo di quelle costituite su piattaforme dei classici SNS-Social Network Sites Networking[2]. Ma non è così. Pare infatti che sia in atto un cambiamento, che anche i grandi social network come Facebook hanno colto, se è vero che si sono molto prodigati per acquistare WhatsApp, nonostante esistesse già Messenger.

Si tratta di un cambiamento messo in atto in particolare con la diffusione sempre maggiore di gruppi creati per app di messaging: è vero che si dà maggior credito (e quindi è più facile cadere in trappola) a informazioni condivise via WhatsApp o Telegram, fondamentalmente perché **si attribuisce maggiore credibilità alla propria cerchia sociale,** ma **nel sistema di diffusione entrano in gioco anche dinamiche di riconoscimento sociale,** non solo processi attivati dalla struttura tecnologica del dispositivo (perlopiù smartphone, ma anche web o desktop) o della piattaforma digitale. Si pensi, ad esempio, ai gruppi WhatsApp o Messenger delle scuole, dei gruppi di classe, oppure dei genitori, gruppi relativi a qualche attività condivisa dai figli con i loro compagni; oppure, attivati per team di lavoro.

In questi casi, si riproducono gerarchie sociali e rapporti di forza, logiche di gruppo tipici di un contesto sociale complesso anche se semplificato e ridotto dalla forma digitale. **Come nella realtà, non si mette in discussione l'informazione del proprio docente o del capo e del capoclasse**, sebbene il velo di Maja della comunicazione digitale asincrona possa talvolta disinibire comportamenti che altrimenti rimarrebbero inespressi. E tuttavia, non si può nemmeno evitare di essere aggiornati su dinamiche di gruppo, dalle quali un tempo si rifuggiva, semplicemente non partecipandovi (si pensi alle assemblee di condominio o ai consigli didattici).

Un esempio è la massiccia diffusione delle notizie no vax in particolare via WhatsApp, rimbalzando poi su Facebook con un effetto virale potente. I messaggi anti-vaccino sui social media sono stati

recentemente studiati per contenuto, portata ed efficacia, ma la misurazione della loro viralità è cosa complessa e spesso osservabile solo a distanza di tempo[3]. Per di più, ancora molto poco si sa su chi crea le bufale e su come agisce, da quali interessi è mosso, se si tratta di information war commissionate oppure di forme di panico che attecchiscono attraverso la rete.

Anche il ruolo dei cosiddetti influencer è stato oggetto di numerose (e non sempre convergenti) indagini. Le patina di vicinanza degli influencer al pubblico dei follower e il loro comportamento friendly hanno comunque dei limiti.

Non a caso**, in un ecosistema sempre più inquinato da fake e postverità, l'infosfera è messa a dura prova, sicché la rete delle relazioni personali di prossimità, con il proprio doppio digitale dei gruppi social di messaging, come WhatsApp, agisce con maggiore forza persuasiva sulle opinioni individuali** di quanto spesso non facciano i guru del web.

Tuttavia, questo meccanismo ha una falla importante, un bug di sistema: se un anello della catena, cosciente o no, immette nella propria rete di prossimità una fake, lo infetta – per tenere ferma la metafora virale. **È possibile cioè iniettare dosi di falsità in un sistema di rapporti sociali collaudati e nei quali si ripone mediamente maggiore fiducia di quanta invece non si conferisca ai legami deboli instaurati via social network**. Come ha più volte ribadito Floridi, la nostra vita è del tutto permeata dalla dimensione onlife[4]. Ciò vuol dire che non è più possibile una separazione netta fra "realtà naturale" e "realtà digitale", né tantomeno si dà più una distinzione fra relazioni personali e sfera delle relazioni sociali e di lavoro, fra privato e pubblico, sicché vi è una ibridazione finanche delle relazioni sociali, delle reti di prossimità con il networking attivato dai social media.

Ovviamente, **alla struttura logica della piattaforma digitale si sovrappone quella della rete sociale, con dinamiche tipiche dei processi di riconoscimento**. Non stupisce, infatti, che anche là dove la credibilità di un'informazione venga meno, o addirittura possa recar danno a qualche elemento del network, la necessità di riconoscimento sociale ponga l'individuo di fronte all'aut-aut: o lasciare la rete (e quindi isolarsi, scelta de facto resa impossibile dalla dimensione onlife) oppure subirne la pressione sociale, finanche l'aggressività, come nel caso del cyberbulling messo in atto con app di messaggistica istantanea, con effetti nel migliore dei casi conformistici[5]. danah boyd ha ampiamente dimostrato come in realtà le piattaforme sociali, e in particolare quelle di messaggistica, rappresentino per gli adolescenti reali luoghi di ritrovo sociale.

Ciò vale vieppiù specialmente in situazioni come quella di attuale distanziamento forzato dovuto alla pandemia. La "generazione K" si riversa nei gruppi WhatsApp come un tempo i ragazzi degli anni Ottanta e Novanta si ritrovavano in oratorio per giocare a calcio balilla. E, come sostiene la boyd, riuscire a trovare forme di comunicazione efficace in un sistema di questo tipo "it's complicated"[6].

Come la scienza sta perdendo di credibilità: il ruolo dell'ignoranza attiva

Sarebbe però un errore immaginare l'erosione della credibilità dell'infosfera come una malattia endemica, circoscritta alle piattaforme digitali, ai social media o alle app di messaging. Si tratta invece di un'infezione sistemica, pandemica, che coinvolge l'intera infosfera e che trova una delle sue cause principali nella **perdita di autorevolezza delle fonti istituzionali e scientifiche.**

La natura complessa delle informazioni che circolano sulle piattaforme sociali è ambigua ed insidiosa, spesso capziosa. Inoltre, se la conoscenza, come afferma Floridi, racchiude sempre la verità fattuale, **la non-conoscenza** è invece un fenomeno più complesso e bivalente: da un lato **può derivare da una conoscenza distorta o da una fonte mendace**, che ha un potere negativo sulle azioni umane, perché la condiziona sulla base di falsi contenuti, credenze o pseudo-verità; dall'altro lato, **può essere una questione di semplice ignoranza** (non conoscenza della verità), quindi di mancanza di verità[7].

Per esempio: una cosa è dichiarare di non aver visto un ladro nel mentre commette un delitto (omissione colposa); un'altra è mentire affermando che il criminale è Tizio anziché Caio (falsa dichiarazione). Inoltre, **ignoranza e credenza non coincidono, ma si sostengono a vicenda, si nutrono reciprocamente**. Ad esempio, prima di Copernico non si sapeva che la terra fosse un pianeta in orbita attorno al sole (ignoranza); e si credeva che la terra fosse al centro dell'universo (credenza). Sebbene le due tesi si sostengano a vicenda, non coincidono. In effetti, il geocentrismo è anche compatibile con la teoria del sistema di Tycho Brahe (anch'esso errata).

Pertanto, oltre alle false informazioni, diffuse involontariamente o volontariamente (mis- e dis-information), **l'infosfera è condizionata da vere e proprie lacune informative (ignoranza, ignorance as un-information) e da ammassi di credenze o pseudo-verità**, che rimangono tali fino a prova della loro inesattezza o insensatezza; oppure, assurgono a rango di post-verità, oggetti sociali inscalfibili[8], anche se sottoposte a confutazione scientifica.

A tutto ciò si aggiunge che **le informazioni false o cattive non sono allo stato, per così dire, pure:** non si tratta cioè di menzogne chiare e distinte. Ovvero, non si tratta di risolvere un'espressione formale secondo la logica booleana, poiché le singole proposizioni non sono enucleabili e atomiche, dissolte e risolte, ma esistono insieme ad altro, come grumi concettuali, in cui vero e falso convivono, in un sistema granulare tipico dell'informazione in rete[9], nella forma della plausibilità, del verosimile[10]. Infatti, se le informazioni fossero palesemente vere o false, nessun utente cadrebbe nella trappola delle fake-news.

In tale contesto, un ruolo significativo è svolto dalla **comunità scientifica e dalla credibilità che le si conferisce, come fonte autorevole di verità**. La divergenza tra diversi punti di vista, che nella storia è stata ed è il sale della scienza, può però essere manipolata a favore di false costruzioni di realtà, dovute a verità convenienti, condizionate cioè da interessi economici o scientifici o politici o più semplicemente di parte. Con ciò si genera un fenomeno di discredito, di perdita della credibilità, che investe l'intera comunità scientifica, che può essere definito **"lacuna di autorità"**.

I punti dogmatici di divergenza o disaccordo tra presunti scienziati (o individui che, in virtù di una più o meno flebile fama mediatica si accreditano presso l'opinione comune come luminari) e la comunità scientifica internazionale **generano un fenomeno di perdita di credibilità e demolizione dell'autorità in senso ampio** (istituzionale, sociale e scientifico).

Di ignoranza non solo come vuoto informativo ma soprattutto come "costruzione sociale", quindi in senso attivo, si è discusso di recente in un corposo volume collettaneo a cura di Kourany e Carrier[11]. I curatori usano il termine **agnotologia** (lemma coniato invero da Robert Proctor) per riferirsi allo studio di ignoranza intenzionalmente prodotta, "creata, mantenuta e manipolata" da una scienza sempre più condizionata dalla politica e dagli affari, come costruzione sociale giustapposta.

L'agnotologia riguarda la costruzione attiva dell'ignoranza attraverso la progettazione faziosa e l'interpretazione distorta di esperimenti e studi empirici; ad esempio, le ricerche mendaci dei negazionisti del cambiamento climatico. Oppure, è possibile una costruzione 'virtuosa' dell'ignoranza, ad esempio limitando la ricerca sulle differenze cognitive legate alla razza e al genere; oppure, l'ignoranza come sottoprodotto non intenzionale delle scelte fatte nel processo di ricerca, quando regole, incentivi e metodi incoraggiano un'enfasi sugli effetti benefici (o nocivi) e commerciali dei prodotti chimici industriali e quando certi concetti e persino gli interessi di certi gruppi non possono essere sfruttati in un data struttura concettuale. Non si tratta di errori protocollari, ovvero di errori contemplati e controllati dalla rilevanza statistica, ma distorsioni vere e proprie del sistema.

Stando a questo nuovo approccio, insomma, l'ignoranza è molto più complessa di quanto si sia sinora pensato. **L'ignoranza non è cioè solo il vuoto che precede la conoscenza o la privazione che risulta da un'attenzione parziale. È anche – anzi e soprattutto – qualcosa che può essere costruito, realizzato socialmente costruito e attivo: ne sono un esempio la confusione prodotta quando interessi particolari bloccano l'accesso alle informazioni o addirittura creano disinformazione su di un tema di rilievo pubblico o globale**[12]. Nel 1854, il filosofo scozzese James Frederick Ferrier aveva si era già interessato alle forme di ignoranza attiva[13]. L'autore non solo usa per primo il termine "epistemologia" in senso moderno, ma anche il termine "Agnoiologia" (con la "i"; il libro è diviso in tre sezioni: dopo una lunga Introduzione, seguono le sezioni su Epistemologia, Agnoiologia e Ontologia). In un certo senso, l'agnologia di Proctor sembra essere influenzata dalla agnoiologia di Ferrier, che scrive come l'ignoranza sia un difetto intellettuale, imperfezione, privazione o mancanza[14]. Tuttavia, pur avendo una natura difettosa, costituisce una barriera contro la conoscenza, che svolge un ruolo attivo[15].

Un altro fenomeno rilevante, a tal proposito, è generato da una forma di ignoranza attiva derivata dalla **proiezione di autorità**, che avviene quando uno scienziato viene chiamato a discutere di questioni che esulano dalle proprie specifiche competenze scientifiche e dall'ambito dei propri studi. In tal caso, l'autorevolezza conquistata nel proprio ambito viene proiettata sulle opinioni dello scienziato che è legittimato a disquisire di temi e argomenti non pertinenti al suo campo di indagine, senza cioè adottare alcun metodo scientifico. È un po' come se si chiedesse a Cacciari che cosa pensa della trasmissibilità del virus, oppure a Burioni qual è il ruolo dell'immaginazione riproduttiva nello schematismo kantiano. In entrambi i casi, l'uditorio attribuisce al relatore un credito di autorevolezza senza che però essa derivi in alcun caso dalla materia di ricerca per il quale il relatore è riconosciuto. Nell'ambito delle piattaforme sociali questo fenomeno moltiplica il potere virale di un messaggio, polarizzando la discussione, generando cioè un'oggettivazione di opinioni frequenti e un ricorso **all'argumentum ab auctoritate** che abbassa il livello critico della discussione.

In tal senso, la cornice di delegittimazione scientifica in atto, tanto sulle piattaforme sociali quanto nel paese reale, si riflette anche sui comportamenti degli stessi scienziati, con conseguenze catastrofiche per il credito e l'autorevolezza della comunità scientifica: venendo meno a un aplomb istituzionale e professionale improntato al rigore scientifico, **lo scienziato ricerca il plauso e il consenso, si piega cioè a una polarizzazione delle opinioni e, per dirlo con Freud, ha luogo una "alleanza del sintomo", sicché egli si allinea a una fazione polarizzata dell'opinione comune.** In parte, la proiezione dell'autorità è effetto del condizionamento performante dei new-media che tendono a spettacolarizzare il messaggio, con effetti di semplificazione e polarizzazione delle

opinioni. Gli strumenti di comunicazione digitale non sono però solo performanti, bensì anche trasformativi[16], attivano pratiche e forme sociali impensabili senza la loro esistenza.

Un'ultima forma di ignoranza attiva è quella prodotta **dall'attività iper-specialistica che genera conoscenza in senso sapienziale.** A fianco a quella capziosa e orientata politicamente (le ricerche prezzolate di Trump sull'inquinamento ambientale, oppure di alcune case farmaceutiche sul fumo negli anni Settanta, per esempio), vi è cioè una forma di ignoranza del punto di vista "generale" prodotta dall'iper-specializzazione, nel senso di conoscenza preclusa ai più, secondo un modello aristocratico di cultura, che vorrei definire "sapere sacerdotale", consegnato a élite culturali chiuse, con potere esclusivo, orientativo e decisionale; si tratta di un iper-sapere specialistico che rifiuta di spiegarsi trincerandosi dietro la 'difficoltà della materia'.

Di grande interesse risulta infatti il discorso sulle grandi sacche di "**conoscenza pubblica nascosta"**[17]. Si pensi alla conoscenza "brevettata", privatizzata, ovvero – per dirlo con l'etimo – la conoscenza di cui è privata la comunità scientifica. Ne può essere un esempio efficace e cogente quello del **brevetto dei vaccini**, che hanno un interesse preponderante per la salute pubblica mondiale. Di recente, una critica al sapere specialistico è stata lanciata sulla nota piattaforma TED da David Epstein, con una conferenza intitolata "Why specializing early doesn't always mean career success", reperibile anche su YouTube. Epstein ha discusso la sua teoria in un volume, "Range"[18] di cui la LUISS ha pubblicato la traduzione italiana[19]: l'autore cerca di dimostrare come in realtà la specializzazione precluda molto spesso a un atteggiamento flessibile e come invece sia garantita da contesti di apprendimento "gentile", come li aveva definiti lo psicologo Robin Hogarth, ovvero da campi di azione controllati, ben definiti, che però sono ben estranei alla ricerca universitaria, la quale (si presuppone) si muove su sentieri inesplorati e in maniera spesso interdisciplinare.

La moneta della credibilità è, quindi, inflazionata anche nel mondo scientifico. Si pensi alle riviste peer review, inondate da continue richieste di valutazione, alla loro ricerca spasmodica di revisori affidabili, il cui lavoro – immane e ingrato –, fra l'altro, non è nemmeno titolo di valutazione, con il paradosso che c'è chi è riconosciuto come revisore, ma non può esibire questo credito come titolo valutabile.

Un caso a sé è invece quello del pre-print e delle repository che ospitano articoli pubblicati (questo è il punto) con beneficio di inventario perché comunque non passati al vaglio della revisione fra pari. Può quindi darsi il caso di scienziati specializzati che, in attesa che il loro paper sia accettato, rivisto o rigettato, lo mettano a disposizione della comunità scientifica, per un confronto aperto; ma si può dare anche il caso di numerosi sedicenti scienziati che fanno 'massa acritica', una sorta di mucillagine informativa inerte che inquina l'infosfera con informazione – non originale, falsa o mendace – e rende vischiosa l'infosfera, generando una vera pania informazionale. Le conseguenze possono essere catastrofiche.

Si pensi, a mo' di esempio, al noto fisico e informatico ungherese, Albert-László Barabási, della Northeastern University di Boston, Massachusetts, il quale, a febbraio, ha presentato un documento al server di pre-print di bioRxiv, vedendosi il *paper* rigettato in automatico. Difatti, il repository biomedico non accetterebbe più manoscritti che fanno previsioni sui trattamenti per COVID-19 sulla base esclusiva del lavoro computazionale. Il team di bioRxiv ha suggerito a Barabási di inviare lo studio a una rivista per una rapida revisione tra pari, invece di pubblicarlo come pre-print[20].

Qui si tace infine della laida suburra del cosiddetto **predatory publishing**[21], ovvero di tutti quei Journal o riviste che millantano riconoscimenti scientifici, talvolta emulando il titolo di qualche celebre testata con cui vorrebbero confondersi: un modello di sfruttamento della logica concorsuale accademica, a danno soprattutto dei giovani ricercatori, che imita l'editoria scientifica, con lo scopo di addebitare costi di pubblicazione agli autori, senza controllare la qualità scientifica dei loro paper.

Vengono detti predatori perché gli studiosi sono indotti a pubblicare con loro (sebbene alcuni possano talora essere consapevoli che la rivista è di scarsa qualità o addirittura fraudolenta). Sono, questi, i frutti della logica della quantità dei "prodotti di ricerca" da presentare in fase di selezione concorsuale. Il frutto perverso di un sistema di selezione della ricerca che ha cortocircuitato. Possiamo forse giudicare in maniera positiva un sistema che valuta "il contenuto dal contenitore", ovvero l'articolo dalla rivista che lo ospita – solitamente una rivista di "fascia A", il cui apriori determina spesso ab divino la bontà del contributo ospitato? Oppure, possiamo giudicare ottimale il più rigido rispetto della modalità double blind, che presume l'anonimato del candidato e del suo revisore (salvo i soliti espedienti), ma che invece mette in atto la più palese forma di abdicazione di responsabilità individuale: non sarebbe forse più corretto esprimere un giudizio, negativo o positivo, di un contributo, accollandosene l'onere del giudizio? E come giudicare la politica di premialità per le riviste assunte agli onori della prima fascia? Si glissa qui sui diversi criteri, ma sarebbe forse già interessante se qualche giornalista indagasse, per esempio, sulle compagnie di giro che compongono i board delle riviste di prima fascia, e quali e quante ristrette lobby siano presenti oramai all'interno del sistema universitario italiano, che vede una certa concentrazione, anche geografica, del potere di selezione delle leve.

In tutto questo, un ruolo fondamentale è svolto certamente **dai media, i quali rappresentano il mondo scientifico come un'arena di contraddizioni,** riducendolo spesso la discussione a mera canea televisiva. La neutralizzazione delle differenze, l'impossibilità di una prospettiva scientifica ancorata alla forza statistica dei numeri (la "pandemia dei dati"[22]), di una argomentazione complessa (spesso cassata perché noiosa), la costante presenza di opinionisti di ogni risma, hanno tutti generato l'idea di una scienza inaffidabile o perlomeno opinabile. **Oppure, viceversa e paradossalmente, l'immagine di una scienza dogmatica,** alla quale è necessario credere per decreto divino - là dove invece la conoscenza scientifica è tale se può dimostrare con prove e metodo scientifico, e non per articolo di fede. Un ruolo altrettanto importante è anche quello assegnato a ciò che danah boyd ha chiamato **silenzio strategico**[23], il palinsesto informativo che determina l'agenda setting del discorso pubblico.

Siamo al bivio: logoramento sociale o rafforzamento degli "anticorpi" alle bufale?

Per concludere, c'è ancora molto da studiare. E molto dipende anche dalla geografia ad arcipelago delle numerose piattaforme digitali, ciascuna con delle funzioni e logiche specifiche, dalla loro mutazione interna (con le diverse ibridazioni e segmentazioni fra piattaforme, come quella fra Facebook, WhatsApp e Instagram) che si ripercuotono sull'infosfera con una potenza virale, in senso crossmediale, davvero difficile da analizzare e controllare. Un ruolo centrale, imprescindibile è quello svolto dalla scienza, alla quale in un mare così vasto e periglioso sarebbe necessario affidarsi come ad una buona bussola. Ma di certo, occorre che la freccia dell'autonomia della ricerca segni il nord dell'interesse pubblico, sia libera da compromessi con l'interesse privatistico e

individuale, riconquisti l'autorevolezza che in parte ha perduto per propria colpa, per seduzione di sirene mediatiche e cortocircuiti interni, e **si dimostri autorevole anziché autoritaria**.

Di questo passo, l'analisi dell'erosione della credibilità nell'infosfera potrebbe portare a due risultati differenti: uno nefasto, ovvero il **logoramento, lento e irreversibile, del patrimonio di credibilità**, lo sperpero di capitale di fiducia sociale che è alla base del concetto di comunità, come dono reciproco e credito fiduciario fra pari; oppure, viceversa, **potrebbe rafforzare le 'difese immunitarie'** del nostro sistema di comunicazione e informazione, educandoci alla falsità, secondo un **processo di addomesticamento**[24], riuscendo a limitare l'inquinamento dell'infosfera. L'esito dipende in buona sostanza dagli attori politici e sociali, da un lato, ma anche da una formazione maggiore agli strumenti digitali. Nessuno è dispensato, ciascuno faccia la propria parte.

Bibliografia

Bleize, Daniëlle NM, et al. "The effects of group centrality and accountability on conformity to cyber aggressive norms: Two messaging app experiments". *Computers in Human Behavior* (2021): 106754.

boyd, danah. *"*It's complicated: The social lives of networked teens", Yale University Press, 2014.

Ciracì, Fabio, "Per una teoria critica del digitale: fake-news e postverità alla luce della logica della verosimiglianza", in "Filosofia e digitale", Quaderni di «Filosofia», a cura di Fabio Ciracì, Riccardo Fedriga, Cristina Marras, Mimesis 2021, pp. 87-112 (in corso di stampa).

Ciracì, Fabio. "Lacuna di autorità e costruzione dell'ignoranza attiva", in "AIUCD 2021 - DHs for society: e-quality, participation, rights and values in the Digital Age. Book of extended abstracts of the 10th national conference, Copyright ©2021 AIUCD, Associazione per l'Informatica Umanistica e la Cultura Digitale, ISBN: 9788894253559, pp. 343-347.

Dator, James A., John A. Sweeney, and Aubrey M. Yee. "Mutative media". Springer International Pu, 2016.

Donovan, Joan, and danah boyd. "Stop the presses? Moving from strategic silence to strategic amplification in a networked media ecosystem." American Behavioral Scientist 65.2 (2021): 333-350.

Epstein, David. "Range. Why generalist Triumph in a Specialized World", Macmillan Publishers 2019; tr. it., "Generalisti: Perché una conoscenza allargata, flessibile e trasversale è la chiave", Luiss University Press, 2020.

Ferraris, Maurizio, "Postverità e altri enigmi*"*. Il M;ulino, 2017. Kindle

Ferraris, Maurizio. "(forthcoming). Metafisica del web*"*. Roma-Bari: Laterza.

Ferrier, James Frederick. "Institutes of Metaphysic: The Theory of Knowing the Mind" London: Blackwood and Sons, 1854."

Floridi, Luciano, ed. “The Cambridge handbook of information and computer ethics”. Cambridge University Press, 2010. Id., “The ethics of information”. Oxford University Press, 2013.

Floridi, Luciano, “La rivoluzione dell’informazione”. Torino: Codice, 2012;

Iannelli, Laura, “Facebook & Co.”, Guerini, 2010

Kourany, Janet, e Martin Carrier. “Science and the Production of Ignorance: When the Quest for Knowledge Is Thwarted”. Cambridge: MIT Press, 2020.

Kwon D. “How swamped preprint servers are blocking bad coronavirus research”. Nature. 2020 May; 581(7807):130-131. doi: 10.1038/d41586-020-01394-6. PMID: 32382120.

Leader, Amy E., et al. "Understanding the messages and motivation of vaccine hesitant or refusing social media influencers.", Vaccine 39.2 (2021): 350-356.

Massarenti, Armando e Antonietta Mira. “La pandemia dei dati. Ecco il vaccino”, Mondadori, 2021.

Paini, Germano, and Maurizio Ferraris. “Scienza nuova: ontologia della trasformazione digitale”, Scienza nuova (2018): 1-230.

Quattrociocchi, Walter, e Antonella Vicini. “Misinformation. Guida alla società dell’informazione e della credulità”. Milano: Franco Angeli, 2016.

Recuero, Raquel, Felipe Soares, and Otávio Vinhas. "Discursive strategies for disinformation on WhatsApp and Twitter during the 2018 Brazilian presidential election.", “First Monday “(2021).

Roncaglia, Gino. «Tra granularità e complessità: contenuti digitali e storia della rete». Nuovi annali della Scuola speciale per archivisti e bibliotecari 31 (2017): 349-361.

Russo, Federica, et al. "Internet addiction disorder: nuova emergenza nel mondo dell’infanzia e dell’adolescenza". Quaderni di Psicoterapia Cognitiva-Open Access 47 (2021).

Selene Arfini, e Di Cecco. “Ignorant cognition”. Berlin: Springer International Publishing, 2019.

Terrone, Enrico, “Filosofia dell’ingegneria”. Il Mulino, 2019.

Si vedano Floridi Luciano, “La rivoluzione dell’informazione”. Torino: Codice, 2012; Quattrociocchi, Walter, e Antonella Vicini. “Misinformation. Guida alla società dell’informazione e della credulità”, Milano: Franco Angeli, 2016. ↑

Intendo qui avvalermi della definizione offerta da Iannelli, Laura, "Facebook & Co”., Guerini, 2010, p. 13: «I Social Network Sites sono dunque ambienti informativi che abilitano un networking individualizzato in reti di relazioni friend-driven attraverso un profilo, una lista di contatti e una

gestione negoziata di molteplici pratiche comunicative». La definizione di “reti di prossimità” è di chi scrive. ↑

Cfr. Leader, Amy E., et al. "Understanding the messages and motivation of vaccine hesitant or refusing social media influencers.", Vaccine 39.2 (2021): 350-356. ↑

Floridi, Luciano, ed. “The Cambridge handbook of information and computer ethics”, Cambridge University Press, 2010. Id., "The ethics of information". Oxford University Press, 2013. ↑

Cfr. Bleize, Daniëlle NM, et al. “The effects of group centrality and accountability on conformity to cyber aggressive norms: Two messaging app experiments”. Computers in Human Behavior (2021): 106754. ↑

boyd, danah. ”It's complicated: The social lives of networked teens”, Yale University Press, 2014. Su desiderio dell’intellettuale americana, teniamo in minuscolo la scrittura del suo nome. Per uno studio sull’impatto delle nuove tecnologie sul nostro modo di vivere, con particolare attenzione alle giovani generazioni, si veda Russo, Federica, et al. "Internet addiction disorder: nuova emergenza nel mondo dell’infanzia e dell’adolescenza", Quaderni di Psicoterapia Cognitiva-Open Access 47 (2021). ↑

Vengono riprese e sviluppate le ricerche presentate nel *paper* “Lacuna di autorità e costruzione dell’ignoranza attiva”, in “AIUCD 2021 - DHs for society: e-quality, participation, rights and values in the Digital Age. Book of extended abstracts of the 10th national conference”, Copyright ©2021 AIUCD, Associazione per l’Informatica Umanistica e la Cultura Digitale, ISBN: 9788894253559, pp. 343-347. ↑

Si veda la definizione di “oggetto sociale” di Ferraris, Maurizio, “Postverità e altri enigmi”, Il Mulino, 2017. Kindle, pos. 959: «Oggetto sociale = atto registrato. Un oggetto sociale è il risultato di un atto sociale (tale da coinvolgere almeno due persone) che ha la caratteristica di essere registrato su un supporto qualsiasi, dalla mente delle persone al web, passando per gli archivi cartacei. In base a questa legge, la documentalità si presenta come il fondamento della realtà sociale: è all’opera prima del capitale, ne costituisce il fondamento, e continua a valere anche dopo che il capitale ha ceduto il posto alla medialità e alla documedialità». Sul rapporto esistente fra digitale e mondo degli oggetti sociali, si veda anche Ferraris, Maurizio, “Metafisica del web”. Laterza, Roma-Bari, 2021 (in corso di stampa). ↑

Roncaglia, Gino. «Tra granularità e complessità: contenuti digitali e storia della rete». Nuovi annali della Scuola speciale per archivisti e bibliotecari 31 (2017): 349-361. ↑

Cfr. Ciracì, Fabio, “Per una teoria critica del digitale: fake-news e postverità alla luce della logica della verosimiglianza”, in “Filosofia e digitale”, Quaderni di «Filosofia», a cura di Fabio Ciracì, Riccardo Fedriga, Cristina Marras, Mimesis 2021, pp. 87-112 (in corso di stampa). ↑

Kourany, Janet, e Martin Carrier. “Science and the Production of Ignorance: When the Quest for Knowledge Is Thwarted”. Cambridge: MIT Press, 2020. ↑

Id., p. 3. ↑

Ferrier, James Frederick. “Institutes of Metaphysic: The Theory of Knowing the Mind”. London: Blackwood and Sons, 1854. ↑

Id., p. 397 ↑

Su questo argomento si vedano anche Arfini, Selene, Selene Arfini, e Di Cecco. “Ignorant cognition”. Berlin: Springer International Publishing, 2019. ↑

Sul tema, si veda il fondamentale Dator, James A., John A. Sweeney, and Aubrey M. Yee. “Mutative media.” Springer International Pu, 2016. Uno studio interessante dal punto di vista epistemologico è quello di Enrico Terrone, “Filosofia dell'Ingegneria”, Il Mulino 2019. ↑

Id., p. 163. ↑

Epstein, David. “Range. Why generalist Triumph in a Specialized World”, Macmillan Publishers 2019. ↑

Id., “Generalisti: Perché una conoscenza allargata, flessibile e trasversale è la chiave”, Luiss University Press, 2020. ↑

Kwon D. “How swamped preprint servers are blocking bad coronavirus research”, Nature. 2020 May; 581(7807):130-131. doi: 10.1038/d41586-020-01394-6. PMID: 32382120. ↑

Si veda la lista, in continuo aggiornamento, pubblicata su https://predatoryjournals.com/publishers/ ↑

Sul tema del covid si veda per esempio il bel volume di Massarenti, Armando e Antonietta Mira. “La pandemia dei dati. Ecco il vaccino”, Mondadori, 2021. ↑

Donovan, Joan, and danah boyd. "Stop the presses? Moving from strategic silence to strategic amplification in a networked media ecosystem." American Behavioral Scientist 65.2 (2021): 333-350. ↑

Iannelli, Laura,” Facebook & Co.”, cit., p. 11. ↑

Contro gli algoritmi che ci manipolano, l'educazione psicosociale degli utenti

Al netto delle responsabilità dei progettisti e della necessaria regolamentazione del controllo esercitato dai sistemi di AI, è fondamentale che siano soprattutto gli utenti a essere alfabetizzati a un uso che non è tecnico ma sempre più psico-sociale per conoscere processi ed effetti dell'uso delle tecnologie digitali

Di **Daria Grimaldi**, docente di psicologia sociale delle comunicazioni di massa, Università di Napoli Federico II

L'accelerata dei processi di **digitalizzazione** nell'ultimo anno ha portato una serie di riflessioni che non possono essere più rimandate e che riguardano la capacità degli individui di comprendere e controllare i processi in atto.

La disponibilità di enormi quantità di **dati** ha reso facile ricavare nuove informazioni e rimandare un gran numero di decisioni agli algoritmi: dal semplice profiling che spinge Netflix a suggerirci un film, passando ai sistemi che aiutano le banche a determinare la nostra **affidabilità creditizia**, fino ad arrivare agli strumenti per la diagnosi di problemi di salute o al **deep learning**, utilizzato per intervenire sulla questione climatica, tra le altre cose.

Se da una parte gli sviluppi hanno garantito importanti progressi, i rischi derivanti dall'automazione del bias sono esacerbati dalla quantità di dati ora generati, che gli utenti stessi condividono volontariamente in svariati modi e che si prevede quadruplicheranno entro il 2025. Questo enorme volume di dati se da una parte riduce i costi e sveltisce i processi decisionali, dall'altra rischia di lasciare carta bianca **all'utilizzo di algoritmi per scopi manipolatori.**

Il problema non è, chiaramente, l'algoritmo in sé, ma la crescita che l'ingerenza delle tecnologie sta avendo su tutti i livelli della nostra realtà sociale, dal nostro privato quotidiano, all'impatto sulla democrazia e sulla governance, senza che, contemporaneamente, la società ed i singoli individui siano in grado di comprenderne a pieno le **implicazioni** d'uso e senza sufficiente risonanza delle questioni etiche connesse a questo impatto.[1]

Il tema è ampio, come il dibattito che ormai già da anni, soprattutto dopo le elezioni del 2016, si è costruito attorno al peso che hanno gli algoritmi nelle realtà sociali in termini di manipolazione di massa.[2]

Gli algoritmi e le manipolazioni

Da un punto di vista sociale è importante considerare che l'impatto della rapida crescita digitale è legato al generarsi di un crescente **divario digitale** che può minare una ripresa inclusiva, sia a breve che a lungo termine. [3]

A questo si associano le **preoccupazioni etiche** più urgenti correlate al ruolo delle tecnologie come persuasori occulti: [4]

l'amplificazione dei pregiudizi insiti (anche implicitamente) nella progettazione

una crescita ineguale che aumenta pericolosamente il digital divide

la deregolamentazione e il monopolio in gruppi privati, i cui obiettvi sono essenzialmente costruiti sul profitto commerciale.

Amplificazione dei pregiudizi

Il problema che emerge, guardando al futuro è, anzitutto, di ordine etico e riguarda la probabilità che questi algoritmi possano **perpetuare ed inasprire bias decisionali**, come quelli basati su tratti sociali e razziali.[5]

Come sottolineò già nel 2018 la studiosa **Safija Noble** in Algorithms of Oppression[6], i programmatori hanno pregiudizi, così come chi **definisce le regole su cui loro lavorano.**

Tali pregiudizi possono derivare da dati di addestramento basati su inferenze errate, volutamente o inconsapevolmente distorte a monte, basati su informazioni di partenza non rappresentative o incomplete che lasciati alla discrezione dell'automatismo potrebbero crescere il numero di decisioni inique e potrebbero erroneamente riflettere le disuguaglianze storiche caratterizzanti il decision making, per così dire, "analogico" invece di andare ad ottimizzare le prestazioni. Se non controllati, algoritmi di parte possono portare a decisioni che possono avere un impatto collettivo e disparato su determinati gruppi di persone anche senza l'intenzione del programmatore di discriminare.[7]

Il lavoro tra controllo e digital divide

Un contesto in cui la distorsione dell'uso degli algoritmi può avere un impatto incredibile è, ad esempio, **il** mondo del lavoro dove sempre più aziende private si rivolgono alla gestione algoritmica per monitorare la produttività dei dipendenti. "Senza un'attenta considerazione, il luogo di lavoro algoritmico del futuro potrebbe finire come una **distopia basata sui dati**" scrive Mike Walsh.[8]

Essere gestiti algoritmicamente significa non solo essere soggetti più di prima a **monitoraggio e sorveglianza** costanti, ma anche avere meno probabilità di fare carriera cominciando dal basso.

Nelle organizzazioni basate sull'intelligenza artificiale, difatti, esiste un problema non indifferente noto come "limite di codice", riconosciuto come un vincolo nella crescita professionale e nella opportunità di carriera dei lavoratori che, non venendo monitorati da persone, non hanno la stessa possibilità che c'era in passato di avere successo nella stessa azienda.

La rapida digitalizzazione nelle interazioni sociali e nel lavoro ha ampliato le competenze digitali essenziali, tra cui comunicazione, sicurezza informatica ed elaborazione delle informazioni, rendendo così le lacune nell'alfabetizzazione digitale la possibile causa della nascita di una vera e propria "**sottoclasse digitale**". Così denuncia il rapporto Future of Jobs del World Economic Forum,

per il quale si stima che l'automazione potrebbe spostare 85 milioni di posti di lavoro in soli cinque anni e che i lavoratori esclusi dalle risorse digitali perderanno le opportunità di istruzione e occupazione costantemente create dall'economia digitale globale.[9]

Sia nelle economie sviluppate che in quelle emergenti, il rapido passaggio al lavoro a distanza rischia di creare **nuovi divari tra i lavoratori** della conoscenza e quelli in settori pratici che non possono lavorare a distanza e potrebbero non disporre delle competenze e degli strumenti digitali per trovare altro impiego. [10]

https://www.agendadigitale.eu/cultura-digitale/coded-bias-cosi-abbiamo-delegato-il-nostro-razzismo-agli-algoritmi/

La governance e i monopoli

Contestualmente all'aumento dell'uso dei social network, cresce una "economia dell'attenzione" che concentra nelle mani di pochi privati il controllo della maggior parte delle piattaforme, all'interno delle quali si concretizza il quotidiano della comunicazione di massa.

L'ambivalenza degli utenti è data dalla consapevolezza della necessarietà del progresso, che però si accompagna sia alla mancanza di fiducia nelle grandi aziende tecnologiche che alla preoccupazione per l'incapacità di usare la tecnologia in modo critico e responsabile, da parte della maggior parte delle persone.

Questa sfiducia nell'industria tecnologica ha avviato il fenomeno del **techlash normativo** e affonda le radici nella paura che si è generata intorno al potere stesso dei Big del digitale.

La risposta dei governi, orientata ad aumentare la protezione degli utenti o incrementare le pressioni normative sui mercati digitali, non è considerata una strada univocamente positiva: laddove i governi decidono di intervenire direttamente si teme la limitazione della libera espressione, dove rimandano la regolamentazione ai privati si teme l'aumento sproporzionato del potere di questi ultimi sull'opinione pubblica.[11]

Il Techlash normativo, difatti, rischia di considerare qualunque intervento come una restrizione di Internet, una sorta di **azione censoria** che limita la libera espressione, propria delle prime utopie su cui è nata la rete: di volta in volta in causa ci sono i governi o le aziende.

Nella ricerca sulla percezione del rischio (GRPS) si legge che uno dei rischi a lungo termine più probabile potrebbe proprio essere la **concentrazione del potere digitale**, ovvero la tendenza a limitare il discorso politico e sociale a un numero limitato di piattaforme che hanno la capacità di filtrare le informazioni, accumulando dati e potenzialmente prestandosi alla manipolazione degli stessi, senza che venga di fatto garantita una sufficiente trasparenza su come vengano utilizzati.[12]

Disinformazione e manipolazione sono di fatto un problema che nei social va ad influenzare l'opinione pubblica, riuscendo oggi ad orientarla concretamente ed agire sulla realtà sociale.[13]

Progettazione etica ed alfabetizzazione digitale

Il comun denominatore dei fenomeni appena descritti è nel ruolo persuasorio degli artefatti digitali, percepiti in maniera sempre più indipendente dai propri creatori e dagli stessi utilizzatori. Progettati con l'intento endogeno di rendere più facili da raggiungere gli obiettivi desiderati, semplificare ed ottimizzare processi decisionali e personalizzare le esperienze utente, attivano nel raggiungimento di questi scopi strategie manipolatorie che aumentano la motivazione degli utilizzatori, incentivando processi d'uso non sempre critici.

Allo strutturale intento persuasorio si affiancano altri due, quello esogeno, che prevede che la **modifica del comportamento dell'utente** sia voluta da un soggetto che cerca di manipolare un altro attore; e l'intento *autogeno,* che si manifesta quando è il soggetto stesso a decidere di **utilizzare una tecnologia per migliorare i propri comportamenti**.[14]

Pensiamo ad esempio alla **profilazione** propria dei social network nelle campagne di advertising: l'intendo endogeno del sistema è offrire la migliore esperienza utente al fine di aumentare la permanenza all'interno della piattaforma: più al cliente vengono proposti contenuti compatibili più avrà voglia di restare all'interno. Allo stesso modo i brand utilizzano i medesimi algoritmi per targettizzare le proprie **campagne** ed arrivare esattamente ai propri clienti obiettivo, al fine di indurli all'acquisto (intento esogeno). Gli utenti, infine, nell'iscriversi alle pagine o alle newsletter, generano l'intento autogeno della manipolazione, così da avere facilmente aggiornamenti sui propri prodotti preferiti.

Da una parte, quindi, una progettazione volutamente orientata a manipolare i comportamenti degli utilizzatori richiede una riflessione a monte, dall'altra una scarsa consapevolezza da parte degli utenti nel potere delle tecnologie della persuasione richiede una maggior consapevolezza d'uso.

È necessario un **accordo generale e sistemico** sul fatto che la tecnologia persuasiva sia realizzata in modo etico ed il ruolo della progettazione responsabile richiede un esame approfondito delle conseguenze intenzionali e non intenzionali degli strumenti tecnologici.[15]

Secondo Berdichevsky e Neuenschwander un parametro universalmente valido è che i creatori di tecnologia persuasiva non dovrebbero mai cercare di persuadere qualcuno di qualcosa di cui loro stessi non vorrebbero essere persuasi .[16]

Conclusioni

Al netto del ruolo di responsabilità dei progettisti e della necessaria regolamentazione del controllo esercitato dagli strumenti, diviene sempre più necessario che siano soprattutto gli utenti a **essere alfabetizzati ad un uso che non è tecnico ma sempre più psico-sociale**.

La soluzione a lungo termine per una crescita sostenibile risiederà nella capacità di fornire su vasta scala un sistema formativo che permetta di conoscere intrinsecamente processi ed effetti dell'uso delle tecnologie digitali, non tanto o solo tecnicamente, ma in termini di gestione psicosociale dei fenomeni connessi.

Bibliografia

Come espresso nella risoluzione del Parlamento Europeo del 20 gennaio 2021 l'Intelligenza artificiale è, ad esempio, un punto centrale nel Patto Verde Europeo ed ha anche un obiettivo nevralgico il rilancio dell'economia post COVID-19. *L'Intelligenza Artificiale nei programmi UE: rischi e benefici:* https://www.altalex.com/documents/news/2021/02/10/intelligenza-artificiale-programmi-ue-rischi-e-benefici; si veda anche:*Intelligenza artificiale: questioni relative all'interpretazione e applicazione del diritto internazionale*: https://images.go.wolterskluwer.com/Web/WoltersKluwer/%7Bbf393c12-ee6b-443a-80b1-f87d5892e4c6%7D_parlamento-europeo-risoluzione-20-gennaio-2021-intelligenza-artificiale.pdf ↑

Kaye D, *Libertà vigilata. La lotta per il controllo di Internet,* Treccani, 2021 ↑

Global Risk Report 2021: https://www.weforum.org/reports/the-global-risks-report-2021 Si veda: World Economic Forum. 2020. Markets of Tomorrow: Pathways to a New Economy. Insight Report. October 2020. http://www3.weforum.org/docs/WEF_Markets_of_Tomorrow_2020.pdf ↑

Si veda: https://en.unesco.org/artificial-intelligence/ethics ↑

Lee, N. T., Resnick, P. and Barton, G. 2019. "Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms". Brookings Institution. 22 May 2019.https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/ ↑

Noble S. in *Algorithms of Oppression: how search engines reinforce racism, NY University Press, NY 2018* ↑

Lopez, G.. The First Step Act, Congress's Criminal Justice Reform Bill, Explained." Vox, December 3, 2018. https://www.vox.com/future-perfect/2018/12/3/18122392/first-step-act-criminal-justice-reform-bill-congress ↑

Walsh, M. 2019. "When Algorithms Make Managers Worse". Harvard Business Review. 8 May 2019. https://hbr. org/2019/05/when-algorithms-make-managers-worse ↑

Walsh, M. 2020. "Algorithms Are Making Economic Inequality Worse". Harvard Business Review. 22 October 2020. https://hbr.org/2020/10/algorithms-are-making-economic-inequality-worse ↑

World Economic Forum, in partnership with Marsh & McLennan Companies and Zurich Insurance Group. 2020. The Global Risks Report 2020. Insight Report. Chapter 6 False Positive. January 2020. http://reports.weforum. org/global-risks-report-2020/false-positive/ ↑

Kaye, Op. cit ↑

ILO (International Labour Organization). 2020. Global Employment Trends for Youth 2020: Technology and the future of jobs. March 2020. https://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/---publ/ documents/publication/wcms_737648.pdf ↑

Per i ricercatori dell'Oxford Internet Institute, la propaganda computazionale è "un seria minaccia alla vita pubblica" che richiede delle azioni politiche da parte dei governi che vadano oltre il demandare alle aziende private una regolamentazione interna, che non fa altro che aumentare il potere oligarghico dei Big Tech. Bradshaw S., Howard P. N., Challenging truth and trust: a global inventory of organized social media manipulation, oxford Internet Institute, luglio 2018: https://demtech.oii.ox.ac.uk/research/posts/challenging-truth-and-trust-a-global-inventory-of-organized-social-media-manipulation/ ↑

B.J. Fogg, (2005) Tecnologie della persuasione. Un'introduzione alla captologia, la disciplina che studia l'uso dei computer per influenzare idee e comportamenti,Apogeo, Milano, ↑

Tucker, K.K., *Persuasive Technology: How Can We Make It More Ethical?*https://www.infusedinnovations.com/blog/secure-intelligent-workplace/persuasive-technology-how-can-we-make-it-more-ethical ↑

Berdichevsky D. Neuenschwander E., *Toward an ethics of persuasive technology*, *Communications of the ACM, Vol. 42, No. 5* https://dl.acm.org/doi/10.1145/301353.301410 ↑

Le competenze digitali dei docenti: quale scuola vogliamo dopo il Covid

Docenti maturi dal punto di vista pedagogico e digitale potrebbero creare quell'innovazione indispensabile al sistema educativo post-Covid e stimolare la partecipazione di tutta la comunità scolastica. Sarebbero, in sostanza, i pilastri di una scuola contemporanea basata sulla libertà di imparare e la libertà di insegnare

Di **Daniela Di Donato,** Docente di lettere, Dottoranda di ricerca presso Sapienza Università di Roma-Dipartimento di Psicologia dei processi di sviluppo e socializzazione, Collaboratrice del Crespi

Tredici competenze: neanche un ingegnere aerospaziale ne ha così tante quante deve averne un docente sulla base del Contratto Collettivo Nazionale di Lavoro che, all'art. 27 recita: "Il profilo professionale dei docenti è costituito da competenze disciplinari, informatiche, linguistiche, psicopedagogiche, metodologico-didattiche, organizzativo-relazionali, di orientamento e di ricerca, documentazione e valutazione tra loro correlate ed interagenti, che si sviluppano col maturare dell'esperienza didattica, l'attività di studio e di sistematizzazione della pratica didattica.

I contenuti della prestazione professionale del personale docente si definiscono nel quadro degli obiettivi generali perseguiti dal sistema nazionale di istruzione e nel rispetto degli indirizzi delineati nel piano dell'offerta formativa della scuola".

Non basta, perché queste competenze devono essere **correlate e interagenti**, svilupparsi nel tempo e definirsi in base ad obiettivi decisi dal Paese e dalla scuola di appartenenza, contemporaneamente.

Le competenze informatiche, o meglio le "competenze digitali", sono esattamente al pari delle altre e sono ritenute indispensabili allo svolgimento della propria attività lavorativa. Queste competenze dovrebbero svilupparsi nel tempo assieme a quelle pedagogiche o disciplinari, come segnale della propria maturità professionale.

Modelli di riferimento per le competenze digitali dei docenti

Le teorie e i modelli di riferimento per lo sviluppo delle competenze (non solo digitali) dei docenti sottolineano i medesimi aspetti. Lo studioso Wang nel 2009[1] propone una **commistione di abilità**: conoscere le proprietà educative degli strumenti tecnologici, le loro capacità di promuovere l'interazione sociale degli studenti e sapere anche le modalità attraverso le quali uno strumento è efficace ed efficiente nel permettere agli studenti di portare a termine specifici compiti. Nel framework **TPACK** (Technological Pedagogical and Content Knowledge) si evidenziano

conoscenze e competenze che il docente dovrebbe possedere per poter operare efficacemente nella società della conoscenza.

La **conoscenza pedagogica** delle discipline e la conoscenza dei contenuti, che non può prescindere da una conoscenza del rapporto che li lega alla tecnologia, dovrebbero poi generare competenze psico-pedagogiche, disciplinari, relazionali e tecnologiche in grado di renderlo attivo e creativo designer dell'esperienza di apprendimento dei suoi allievi. L'immagine che si ottiene è quella di un **professionista consapevole** di come gli strumenti tecnologici trasformano le strategie pedagogiche e le rappresentazioni dei contenuti disciplinari, per promuovere nei discenti la costruzione della conoscenza.

I framework che si concentrano anche sul contesto, che può influenzare le pratiche dei docenti e l'uso educativo delle tecnologie digitali, non sono invece molti: abbiamo il **DigCompEdu** (Digital Competence Framework for Educators) e il modello **Expert Teacher**, che a quello si ispira.

Il DigCompEdu

Nel DigCompEdu, elaborato nel 2017 dalla Comunità Europea, grazie al lavoro del Centro Comune di Ricerca (Joint Research Center- JRC), vengono presentate ventidue competenze, suddivise in sei macroaree. I livelli sono cumulativi e sono immaginati come un percorso di ampliamento e affinamento delle competenze, che si sviluppa grazie all'esperienza, alla riflessione e alla collaborazione fra docenti. **I profili professionali sono sei** e vengono identificati con le stesse lettere dei framework per la certificazione linguistica europea: da A1 (principiante) a C2 (pioniere). L'area dedicata all'impegno professionale è quella nella quale l'uso delle tecnologie digitali è anche orientato alla collaborazione con i colleghi, per mutuo aiuto, scambio di esperienze e avvio di pratiche riflessive e autoriflessive (Schön, 1993). Si tratta di un gruppo di azioni potenzialmente in grado di generare **innovazione didattica e organizzativa**. Nel profilo del docente principiante manca proprio l'aspetto legato allo sviluppo professionale e alla collaborazione sulle pratiche fra colleghi, mentre in quello del leader la riflessione sulle pratiche personali e la condivisione con i colleghi è divenuta prassi. Le competenze digitali del docente e del formatore riguardano la capacità di utilizzare le tecnologie digitali non solo per migliorare le pratiche di insegnamento, ma anche per svolgere altre funzioni fondamentali:

- per **interagire** a livello professionale con i colleghi, gli studenti, i genitori e altre parti interessate;
- per la **propria crescita professionale**
- per contribuire al **miglioramento** sia dell'organizzazione in cui si opera, sia del settore professionale dei docenti/formatori in generale.

Il modello Expert Teacher

Il modello Expert Teacher è stato invece elaborato da un gruppo del settore Ricerca e sviluppo della Erickson e ha identificato un *Syllabus* di competenze chiave del docente esperto, individuando tre macro aree di competenza: Professione (Area 1), Didattica (Area 2) e Organizzazione (Area 3).

Nell'**Area 1** troviamo sei competenze:

- praticare l'etica professionale,
- gestire le relazioni,
- formarsi e innovarsi,
- possedere competenze digitali,
- avvalersi della lingua inglese come strumento professionalizzante.

Nell'**Area 2**:

- progettare la didattica,
- valorizzare i talenti e orientare,
- organizzare le risorse educative,
- includere,
- gestire la classe e i gruppi,
- osservare gli studenti e valutare il percorso di apprendimento,
- valutare l'efficacia degli interventi educativi.

Nell'**Area 3** le competenze sono:

- collaborare,
- progettare e valutare,
- gestire e accompagnare.

In almeno due delle aree individuate troviamo citati elementi di contesto, come appunto la collaborazione con i colleghi: la competenza "Gestire le relazioni" dell'Area 1 e la competenza "Collaborare, condividere, gestire i conflitti" dell'Area 3. La novità del modello Expert Teacher consiste nell'aver individuato **quattro profili del docente esperto** (facendo anche leva sugli aspetti della didattica inclusiva) e aver progettato anche un percorso formativo per il conseguimento delle relative competenze:

- Esperto in didattica innovativa e inclusiva
- Esperto in sviluppo professionale continuo
- Esperto in organizzazione scolastica (ambienti di apprendimento integrati e processi di miglioramento)
- Esperto in orientamento formativo.

Il Piano per l'istruzione digitale

La Commissione europea ha **elaborato un piano d'azione per l'istruzione digitale** (2021-2027), che delinea la visione per **un'istruzione digitale di alta qualità, inclusiva e accessibile in Europa**. È un invito ad agire per rafforzare la cooperazione a livello europeo al fine di imparare dalla crisi della pandemia, durante la quale la tecnologia viene utilizzata su una scala senza precedenti nell'istruzione e nella formazione e adeguare i sistemi di istruzione e formazione all'era digitale. Tra giugno e settembre 2020 si è svolta una consultazione pubblica aperta sul nuovo piano. Dai risultati della consultazione è emerso che quasi il 60% degli intervistati non aveva utilizzato l'apprendimento a distanza e online prima della crisi, il 95% ritiene che la crisi pandemica

rappresenti **un punto di non ritorno**, per il modo in cui la tecnologia viene utilizzata nell'istruzione e nella formazione; gli intervistati affermano che le risorse e i contenuti didattici online devono essere più pertinenti, interattivi e di facile utilizzo; oltre il 60% ritiene di aver migliorato le proprie competenze digitali durante la crisi e oltre il 50% degli intervistati vuole fare di più.

Nel documento elaborato dalla Commissione Europea troviamo scritto che è necessario elaborare **orientamenti comuni per gli insegnanti e il personale didattico**, volti a promuovere l'alfabetizzazione digitale e a contrastare la disinformazione attraverso l'istruzione e la formazione.

Ciò dovrebbe avvenire in stretta collaborazione con le parti interessate attraverso un gruppo multilaterale, che riunisca le organizzazioni della società civile, le imprese e gli operatori tecnologici europei, i giornalisti, i media e le emittenti radiotelevisive, il gruppo di esperti sull'alfabetizzazione mediatica e l'Osservatorio europeo dei media digitali, le autorità nazionali, gli istituti di istruzione e formazione, i centri "Internet più sicuro", gli educatori, i genitori e i giovani, e in linea con il prossimo piano d'azione per i media. Si ritiene anche che sia necessario **aggiornare il quadro europeo delle competenze digitali** al fine di includere l'intelligenza artificiale e le competenze connesse ai dati, sostenere lo sviluppo di risorse di apprendimento in materia di Intelligenza artificiale per le scuole, le organizzazioni IFP e altri erogatori di formazione, nonché sensibilizzare in merito alle opportunità e alle sfide dell'IA per l'istruzione e la formazione.

Sviluppare maggiormente l'autoefficacia e promuovere il cambiamento

Bandura ci dice che l'autoefficacia è la convinzione che ciascuno ha di essere capace di dominare specifiche attività, situazioni o aspetti del proprio funzionamento psicologico e sociale. La mente non solo reagisce, ma agisce in modo trasformativo alle sollecitazioni esterne, anche seguendo queste convinzioni, tanto da determinare le proprie abitudini: sapere quanto si pensa di essere in grado di fare qualcosa è un aspetto che incide sulle proprie capacità personali, tanto più se il giudizio che si esprime è autentico. Credere di essere maggiormente **efficaci** in un determinato campo di azione può facilitare il **rendimento** nelle attività correlate a quei campi. Forse una delle competenze che i docenti dovranno sviluppare in futuro sarà proprio questa: **aumentare la propensione personale a promuovere nuove esperienze, a coordinare gli sforzi promuovendo sinergie positive, in modo informale.**

Questi **docenti digitalmente e pedagogicamente maturi potrebbero generare quell'innovazione indispensabile al nostro sistema educativo e stimolare la partecipazione di tutta la comunità scolastica**. La maggiore professionalità del docente potrebbe rendere migliore la sua azione educativa e aumentare la percezione della propria efficacia nell'uso delle tecnologie digitali per l'apprendimento.

Il grande collaudo del digitale immersivo forzato dalla didattica a distanza diventerebbe l'inaspettata opportunità che mancava, per fare quel passo verso la trasformazione delle pratiche didattiche e dell'idea di scuola.

Lavorare sull'autoefficacia significa anche acquisire competenze di orientamento, per progettare e costruire il futuro della scuola insieme agli studenti e non contro di loro: sarà necessario interrompere quella vocazione naturale allo statu quo e rompere le abitudini consolidate. Adattarci e cambiare è parte del nostro essere persone ed esseri umani: il primo passo lo abbiamo

fatto per colpa di un virus, il secondo e quelli successivi li potremo imputare a noi stessi e alla volontà di andare avanti e non indietro.

L'antropologo **Gregory Bateson** ha scritto: "Sembra che la storia dell'evoluzione dell'apprendimento sia stata, a grandi linee, quella di un lento arretrare del determinismo genetico verso livelli di tipo logico superiore". Dovremmo capire se vogliamo arrenderci alla genetica sociale, che ci riporta meccanicamente a procedure quasi innate come la lezione frontale, l'interrogazione e il controllo, dalle quali fatichiamo a separarci, oppure vogliamo passare ad un livello successivo. Si tratta di **affrontare un passaggio delicato**, in cui ci prepariamo a generare una scuola che non voglio definire nuova (quando lo dici sembra già improvvisamente invecchiata), ma **contemporanea**, basata sulla libertà di imparare e la libertà di insegnare. Penso ad una scuola sistemica, che non ragioni più in termini di alunno, docente, aula, ma si rappresenti come comunità che prepara ambienti, che accolgono persone, che sviluppano intelligenze, che crescono in armonia e che non temono il futuro, cioè una vera scuola plurale e coraggiosa.

Bibliografia

Bandura A. (2018), Autoefficacia. Teorie e applicazioni, Erickson, Trento.

Bateson G. (2018), Verso una ecologia della mente, Adelphi, Milano.

Digital Education Action Plan 2021-2027, 30 settembre 2020 in https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020DC0624 (Ultima consultazione: 2 aprile 2021)

Foà C., Saudino M. (2021), Cambiamo la scuola. Per un'istruzione a forma di persona, Eris Edizioni, Torino.

Ianes, D., Cramerotti, S., Biancato, L., & Demo, H. (2019), Il manuale dell'Expert Teacher. 16 competenze chiave per 4 nuovi profili docente, Erickson, Trento.

Koehler M. & Mishra P. (2009), "What is Technological Pedagogical Content Knowledge (TPACK)?". Contemporary Issues in Technology and Teacher Education, 9 (1), 60-70.

Schön, D. A. (1993), Il professionista riflessivo. Per una nuova epistemologia della pratica professionale, Edizioni Dedalo, Bari.

Wang, Q. (2009), Guiding teachers in the process of ICT integration: Analysis of three conceptual models, Educational Technology, 49 (5), 23-27.

La persona al centro: le nuove tecniche di privacy by design nei trattamenti dati

Molte delle tecniche per dare piena attuazione alla privacy by design, scoperte diversi anni fa, si sono rivelate troppo inefficienti e poco pratiche. Con l'implementazione del Machine Learning (ML), tuttavia, se ne è reso possibile un impiego efficace ed efficiente. Vediamo quali sono e come funzionano

Di **Alessandra Lucchini**, Avvocato cassazionista - DPO e **Salvatore Nucera** Judicial intern at Messina Appeal Court, Criminal Section - Junior fellow at DPO innovation - AI and Data ethics activist

In un periodo in cui si parla tanto di **sovranità dei dati** e di capitalismo della sorveglianza [i] è fondamentale, necessario e indispensabile mettere al centro di ogni sistema tecnologico l'individuo, come soggetto a tutto tondo capace anche di autodeterminarsi pienamente nella sua ormai imprescindibile dimensione digitale.

E a tal proposito viene in soccorso il concetto di privacy by design, disciplinato dal primo paragrafo dell'articolo 25 del **GDPR**, secondo il quale la persona fisica, i cui dati vengono trattati, deve essere al centro del sistema di tutela, e il titolare deve dimostrare un impegno effettivo e non solo formale.

La visione statica della protezione dei dati che conoscevamo si è ormai evoluta "in forme dinamiche, liquide, accelerate di protezione dei dati: l'art. 25 ha introdotto la privacy by design che bene si attaglia alla idea generale di una responsabilizzazione ex ante e tendenzialmente in movimento del trattamento, della conservazione e dell'utilizzo dei dati" [ii].

La Privacy by design nel GDPR

La nozione di privacy by design prevede, come noto, che, fin dalla fase di progettazione di nuovi prodotti, servizi o di qualsiasi iniziativa di business, progetti o tecnologie, il Titolare del trattamento dei dati debba:

- **individuare** i dati personali che saranno oggetto di trattamento per mezzo del prodotto, servizio, iniziativa o tecnologia realizzati;
- **determinare**, sin dall'origine, il periodo di conservazione dei dati;
- **individuare** i soggetti che, nelle rispettive aree, avranno accesso ai dati personali;
- **implementare** specifici presidi, in ottemperanza ai requisiti per la protezione dei dati personali, che possano mitigare eventi di violazione in seguito ad attacchi informatici esterni o comportamenti illeciti interni;

- **individuare** quei trattamenti che, presentando rischi elevati per i diritti degli interessati, in quanto trattano una categoria di dati o presentano un rischio residuo di trattamento elevato, sono soggetti alla valutazione d'impatto.

Per fare ciò il Titolare deve tenere conto dello stato dell'arte e dei costi di attuazione, della natura, dell'ambito di applicazione, del contesto e delle finalità del trattamento, come anche dei rischi aventi probabilità e gravità diverse per i diritti e le libertà delle persone fisiche costituiti dal trattamento. Il regolamento suggerisce alcune misure tecniche e organizzative adeguate, la pseudonimizzazione, l'anonimizzazione, e l'integrazione nel trattamento delle necessarie garanzie al fine di soddisfare i requisiti del regolamento e tutelare i diritti degli interessati, nel pieno rispetto del principio di minimizzazione.

Pseudonimizzazione e anonimizzazione

Cominciamo ad analizzare le due misure di pseudonimizzazione ed anonimizzazione per comprenderne caratteristiche e differenze.

La pseudonimizzazione è una tecnica che "consiste nel sostituire un attributo (solitamente un attributo univoco) di un dato con un altro, riducendo la correlabilità di un insieme di dati all'identità originale della persona" [iii]. In altre parole, vengono sostituiti uno o più attributi all'interno di un dato con altri attributi differenti, ovvero gli pseudonimi [iv].

Esistono, diversi metodi per generare pseudonimi, tra i quali:

- **l'utilizzo delle funzioni di hash**, che calcolano, a partire da un insieme di caratteri di lunghezza arbitraria, una stringa alfanumerica di lunghezza determinata;
- **la crittografia**: si tratta di una misura che, attraverso un apposito algoritmo matematico, rende illeggibili i dati personali a chiunque non abbia l'autorizzazione a visionarli, proteggendo i dati da trattamenti non autorizzati o illegali. Per accedere ai dati personali crittografati è necessario essere in possesso di una chiave di decriptazione;
- **la tokenizzazione**: è una misura sviluppata di recente che si basa sull'applicazione di un meccanismo di crittografia univoca o di assegnazione, tramite una funzione indicizzata, di un numero sequenziale o di un numero generato casualmente che non deriva matematicamente dai dati originali. Tale tecnica viene utilizzata in particolar modo nel settore bancario e finanziario [v].

Per **anonimizzazione** si intende, invece, una tecnica che viene applicata ai dati personali in modo tale che le persone fisiche interessate non possano più essere identificate in nessun modo: l'obiettivo è eliminare la correlazione tra i dati personali e una determinata persona fisica (l'interessato), rendendo impossibile l'identificazione della stessa.

L'anonimizzazione si può realizzare tramite la rimozione, la sostituzione, la distorsione, la generalizzazione o l'aggregazione degli identificatori diretti, come il nome completo o altre caratteristiche rilevanti della persona fisica, e indiretti, cioè attributi che combinati con altre informazioni disponibili rendono identificabile una persona.

Uno dei metodi più comuni è costituito dalla tecnica della **generalizzazione**, la quale comporta la riduzione del grado di dettaglio di una determinata variabile. In via esemplificativa, le date di

nascita possono essere generalizzate per mese o anno, producendo una riduzione del grado di identificabilità [vi]. Quindi, eliminare i nomi completi degli individui, mantenendo solo l'anno di nascita degli stessi, permetterebbe di de-identificare in modo irreversibile le persone fisiche, potendo comunque effettuare analisi statistiche sul campione di dati.

In questo senso anche il Parere n. 5/2014 del WP29 [vii] che individua due **macrocategorie**: la randomizzazione, che modifica il grado di verità del dato al fine di eliminare la correlazione che esiste tra lo stesso e la persona e la generalizzazione, che consiste nel diluire gli attributi delle persone interessate modificandone la rispettiva scala o il rispettivo ordine di grandezza [viii]

L'utilità della anonimizzazione è da ravvisare nella condivisione di set di dati, garantendo da una parte la privacy delle persone fisiche, dall'altra la possibilità di sfruttare il predetto set di dati per analisi e ricerche statistiche.

Questioni problematiche: il costante rischio di re-identificazione

Spiegate le due misure andiamo ora ad analizzarne **gli aspetti critici e problematici**.

Quanto alla tecnica di pseudonimizzazione i dati originali, come vedremo al par. III, vengono generalmente conservati in un database e le tabelle delle corrispondenze tra dati originali e gli pseudonimi utilizzati dovrebbero essere conservate in un luogo separato e adeguatamente protetto. Tale sistema, tuttavia, permette di re-identificare le persone fisiche, in quanto il titolare, o il responsabile del trattamento, possiede le "informazioni aggiuntive" che consentono in modo reversibile di risalire all'identità degli interessati.

Per quanto i dati pseudonimizzati non possano essere attribuiti ad un individuo senza utilizzare le predette "informazioni aggiuntive", la presenza di tali informazioni non garantisce di per sé che gli interessati non possano essere identificati. Anche nel caso di utilizzo della crittografia chi conosce la chiave può facilmente risalire all'identificazione di ogni persona interessata decrittando l'insieme di dati.

I **rischi principali dell'anonimizzazione** sono invece:

l'individuazione, che corrisponde alla possibilità di isolare alcuni o tutti i dati che identificano una persona all'interno dell'insieme di dati;

la correlabilità, ossia la possibilità di correlare almeno due dati concernenti la medesima persona interessata, o un gruppo di persone interessate, nella medesima banca dati o in due diverse banche dati;

la deduzione vale a dire la possibilità di desumere, con un alto grado di probabilità, il valore di un attributo dai valori di un insieme di altri attributi [ix].

Le critiche alla privacy by design avanzate dalla dottrina

Nonostante i buoni propositi riposti nel **Regolamento Europeo**, sono state espresse numerose opinioni contrastanti con riguardo alla formula adoperata per introdurre la privacy by design tra i

pilastri del GDPR [x]. E non avrebbe potuto essere altrimenti, stante la ben nota difficoltà della legge ad anticipare le novità fattuali e, in questo caso, tecnologiche in costante evoluzione [xi].

Preliminarmente, occorre evidenziare come ad occuparsi di privacy by design (fin dalla progettazione) sia solo il primo paragrafo dell'art. 25 GDPR, mentre il secondo definisce la privacy by default (per impostazione predefinita).

Il terzo paragrafo del medesimo articolo, invece, rinvia all'art, 42 GDPR (Certificazioni) quale possibile alternativa per dimostrare la conformità del trattamento alle prescrizioni del Regolamento [xii].

Con specifico riguardo al principio di **privacy by design**, il Titolare del trattamento dovrà, come anticipato, tenere in considerazione i rischi aventi probabilità e gravità diverse per i diritti e le libertà delle persone fisiche a cui i dati si riferiscono.

Tali diritti e libertà, a loro volta, necessitano di un puntuale bilanciamento da effettuarsi in concreto, a seconda del contesto (ancora, l'art. 25 par. 1 GDPR) in cui si ritenga necessario il trattamento delle loro informazioni personali. Si pensi al bilanciamento tra diritto all'informazione (art.21 Cost.) e diritto all'oblio (art. 17 GDPR, che trova le sue radici nel diritto alla riservatezza quale tutela della propria immagine), oppure all'attuale dibattito tra tutela della salute (art.32 Cost.) – anche nella sua dimensione collettiva – e libertà di circolazione (art.16 Cost.) [xiii]. In tal senso, la privacy by design si dovrebbe atteggiare quale chiave di volta del sistema, strumento finalizzato ad evitare che vi siano diritti tiranni e a perseguire l'equilibrio tra i molteplici interessi in gioco.

Tuttavia, il principio soffre di alcuni **limiti** legati alla sua applicazione pratica.

Un **primo profilo problematico** è stato rinvenuto nell'eccessiva vaghezza dell'art. 25 GDPR, optata per privilegiare il principio di neutralità tecnologica, con la conseguenza di lasciare i tecnici (del diritto e non) privi di precise coordinate operative su quali misure adottare (le quali sono state di volta in volta individuate per via di prassi, come quelle elencate nei paragrafi precedenti).

In secondo luogo, è stato detto che il testo dell'articolo si sovrappone a troppe parti del GDPR (art. 5, 24, 28, 32), complicando da un punto di vista sistematico il lavoro degli operatori e col rischio di svuotare il principio di un'autonoma portata applicativa.

Ancora, l'art. 25 menziona solo il principio di minimizzazione dei dati mentre, secondo autorevoli opinioni, il design delle nuove tecnologie dovrebbe implementare tutti i principi del GDPR e, in particolare, quelli enunciati all'art. 5 del Regolamento [xiv].

Le soluzioni della tecnica: secure multy party computation, homomophic encryption e blockchain preserving privacy

Per rispondere ai dubbi sollevati dalla dottrina, dalla commistione di scienza giuridica e tecnologica sono state ideate le **privacy enhancing technologies** (c.d. PETs), ossia delle strategie digitali volte a minimizzare, separare, aggregare, nascondere, informare e controllare grandi quantità di dati personali [xv].

Con riguardo a tali strategie, ci si riferisce alle sopra illustrate tecniche di anonimizzazione e di pseudonimizzazione (ma non solo, come si vedrà a breve), basate sui principi di integrità e riservatezza del dato, nonché di generalizzazione e della randomizzazione, basate sul principio di minimizzazione.

Volendo porre un focus sulla pseudonimizzazione, occorre evidenziare che la sua particolarità rispetto all'anonimizzazione consiste nella possibilità di poter accedere a delle informazioni aggiuntive, conservate separatamente e non consultabili da parte di terzi, con le quali è permesso re-identificare la persona a cui i dati pseudonimizzati si riferivano, con la duplice conclusione che:

- i dati speudonimizzati continuano a rientrare nella categoria dei dati personali e, di conseguenza, a dover essere trattati nel rispetto del GDPR;
- tali dati restano esposti a possibili data breach.

Di conseguenza, le tecniche di pseudonimizzazione - il cui utilizzo ad oggi è particolarmente diffuso - necessitano delle dovute strategie per poter limitare al massimo i rischi legati alle fasi del trattamento.

Nello specifico, la prassi ha conosciuto diverse tecniche per dare piena attuazione alla privacy by design. In particolare, ve ne sono alcune particolarmente promettenti come la **secure multy party computation** (MPC) e la **homomorphic encryption**. Tali tecniche sono state scoperte diversi anni fa. Tuttavia, al tempo esse si sono rivelate troppo inefficienti e poco pratiche. Fortunatamente, i recenti progressi algoritmici e, nello specifico, l'implementazione del Machine Learning (ML), hanno reso possibile un impiego efficace ed efficiente di tali tecniche, rendendole dei tools versatili ed in grado di eseguire analisi su set di dati molto più ampi [xvi].

La multy party computation o MPC è una tecnica che si rivela particolarmente utile nelle operazioni di trasferimento di dati tra soggetti diversi e, a volte, in competizione tra loro (c.d. adversaries). Essa assicura la c.d. privacy in ingresso (input privacy), poiché le informazioni fornite da ogni interessato al trattamento – ossia ogni parte – non vengono trasmesse ai restanti partecipanti. L'unico dato condiviso tra tutti riguarda l'output del trattamento, ossia il risultato dell'elaborazione, le cui fasi restano inaccessibili. Volendo fare un'esemplificazione, se Alfa, Alfa 1, Alfa 2 […], Alfa n, forniscono i propri dati x, y, z […], l'unico dato visibile ai partecipanti sarà il risultato della funzione f (x, y, z).

Tra i rischi, la MPC comporta il pericolo di re-identificazione partendo dall'output. Tuttavia, eventuali attacchi possono essere sviati grazie all'applicazione di ulteriori tecniche, quali la k-anonimity, t- closeness e la differential privacy [xvii].

Diverso è il funzionamento della homomorphic encryption. In sostanza mentre nella MPC le informazioni restano segrete durante il trattamento, la homomorphic encryption permette di elaborare dei dati già crittografati in precedenza. In tal modo, tale tecnica consente di tutelare le informazioni non solo durante la conservazione delle stesse, ma anche durante le fasi di elaborazione. Un importante vantaggio di tale tecnica consiste nell'aumentare la resilienza dei sistemi di elaborazione dati, evitando (o per lo meno diminuendo) ipotesi di data breach da parte dei processor, oltre a rendere ancor più complessa la realizzazione di eventuali cyber attacchi dall'esterno. La homomorphic encryption può essere di due tipi: **semi omomorfica** e **completamente omomorfica**. Nel primo caso, essa potrà svolgere le operazioni più semplici, come

la somma e la moltiplicazione; nel secondo, invece, essa potrà anche elaborare calcoli più complessi, come le operazioni aritmetiche e le funzioni boleane [xviii].

Una terza tecnica volta a garantire una piena attuazione della privacy by design e che potrebbe avere dei grandi sviluppi nel prossimo futuro sfrutta la tecnologia **blockchain**. Quest'ultima, come abbiamo accennato, grazie al modello decentrato che la caratterizza, permette di realizzare transazioni con un alto tasso di sicurezza per le informazioni oggetto di trattamento.

Infatti, grazie anche all'apporto del ML, tali modelli possono sfruttare le tecnologie blockchain che utilizzano anche strumenti crittografici per fornire un'archiviazione sicura dei dati. Allo stesso tempo, la combinazione tra ML e blockchain potrebbe ridurre al minimo i rischi per la privacy e la sicurezza legati all'elaborazione di dati non crittografati.

Un esempio è offerto da **Kairos**, un tool di riconoscimento facciale che ha incorporato le tecnologie blockchain. Esso combina la biometria facciale e la tecnologia blockchain per consentire agli utenti di proteggere meglio la propria privacy. Un algoritmo confronta l'immagine di una persona con i punti di riferimento (o identificatori) facciali in primo piano fino a creare una corrispondenza univoca. Questa corrispondenza viene quindi convertita in una stringa di numeri univoca e casuale, dopodiché l'immagine originale può essere scartata, non essendo più necessario conservarla. In tal modo, questa "blockchain biometrica" permette alle aziende o i governi di identificare l'interessato, senza dover effettivamente conoscere il volto dello stesso [xix].

Occorre evidenziare che le tecniche fin qui elencate, che prendono il nome di software based solutions, possono essere accompagnate da altre misure c.d. hardware based solutions - ossia delle misure di tipo fisico – comunque necessarie ad attuare il più generale principio di privacy by design. Tra queste, è possibile menzionare la decriptazione in un trusted execution environment, ossia uno spazio chiuso, a prova di manomissione dall'esterno [xx]-

Conclusioni

Ma queste soluzioni da sole non bastano.

Nel sistema digitale il dato diviene "un cardine della leva di produzione", la sua condivisione oltre che inserirsi in un insieme economicamente apprezzabile diventa un elemento di cessione di parti sempre più diffuse della propria personalità, quasi come se fossero atti dispositivi del proprio corpo [xxi].

È pertanto necessario e assolutamente imprescindibile che vi sia piena consapevolezza da parte del soggetto interessato che alla accettazione di determinati dispositivi conseguirà fisiologicamente una perdita di alcune sue libertà o almeno una loro "dequotazione".

Il cittadino deve quindi vantare il diritto ad una piena e libera informazione, non coartata mediante sistemi di profiling e di **data mining**: il servizio offerto può certamente fare utilizzo per fini promozionali dei dati legittimamente raccolti, ma deve sempre offrire una informazione che renda possibile per il singolo il non aderire e il non usufruire del servizio stesso e soprattutto che gli permetta di sapere, il quanto e il se del trattamento dei dati. L'informazione deve quindi essere garantita ex ante dalla piena conoscibilità dei pattern logici di strutturazione del servizio, le condizioni di utilizzo devono essere chiare e semplici.

Ed è per questo che appare opportuna l'approvazione di un piano generale di utilizzo e di governo dei dati, quello che viene definito "un autentico piano regolatore ombra da sovrapporre alle modifiche urbanistiche innervate di dispositivi tecnologici" [xxii].

Bibliografia

1. Su questi argomenti si veda in particolare SHOSHANA ZUBOFF, *Il capitalismo della sorveglianza, Il futuro dell'umanità nell'era dei nuovi poteri*, Luiss Press, 2019 e *Colpo di stato del capitalismo della sorveglianza*, Internazionale, 9 aprile 2021; LUCA BOLOGNINI, N*on lasciamo all'intelligenza artificiale il compito di prevenire i crimini*, *editoriale domani.it.*

2. ANDREA VENANZONI, *Smart cities e capitalismo di sorveglianza: una prospettiva costituzionale*, Forum di quaderni costituzionali, 20 ottobre 2019.

3. GIOVANNI ZICCARDI, PIERLUIGI PERRI (a cura di); *Dizionario Legal tech*, Giuffrè, Milano, 2020, pg. 786.

4. LUCA BOLOGNINI, ENRICO PELINO (diretto da), *Codice della disciplina privacy*, Giuffrè, Milano, 2019, commento art. 25; GIUSEPPE D'ACQUISTO, MAURIZIO NALDI, *Big Data e Privacy by Design, Anonimizzazione, Pseudonimizzazione, sicurezza*, Giappichelli, Torino 2017; MARCO MARTORANA, *Dati personali: anonimizzazione e pseudonimizzazione. Facciamo chiarezza alla luce delle indicazioni AEPD-EDPB*, *https://www.altalex.com/documents/news/2021/06/08/dati-personali-anonimizzazione-e-pseudonimizzazione;* EDOARDO SIMONE, *Pseudonimizzazione e anonimizzazione dei dati: differenze tecniche e applicative*, https://www.cybersecurity360.it/legal/privacy-dati-personali/pseudonimizzazione-e-anonimizzazione-dei-dati-differenze-tecniche-e-applicative/

5. GIOVANNI ZICCARDI, PIERLUIGI PERRI (a cura di); *Dizionario Legal tech*, Giuffrè, Milano, 2020, pg. 788. Cfr. anche il documento sulle tecniche e migliori pratiche di pseudonimizzazione pubblicato da Enisa nel novembre 2019. In questo documento vengono descritte le diverse modalità utilizzabili per la pseudonimizzazione e vengono confrontate le caratteristiche di attuazione. Vengono inoltre illustrati i criteri principali che un titolare del trattamento può utilizzare per selezionare una tecnica di pseudonimizzazione.

   - La prima ipotesi descritta è la pseudonimizzazione identificativo singolo, tramite le seguenti modalità:
   - Contatore
   - Generatore di numeri casuali (RNG)
   - Funzione di hash crittografico
   - Codice di autenticazione del messaggio (MAC)
   - Crittografia
   - La seconda ipotesi è costituita dalle politiche (o modalità) di attuazione della pseudonimizzazione che si dividono in
   - pseudonimizzazione deterministica,
   - pseudonimizzazione randomizzata dei documenti

- pseudonimizzazione completamente randomizzata.

6. RAMONA CAVALLI, *Anonimizzazione del dato: le tecniche possibili* cit.

7. Parere 05/2014 -WP 216 sulle tecniche di anonimizzazione.

8. Nella randomizzazione sono ricomprese le tecniche di:
   1) aggiunta di rumore statistico.
   2) permutazione.
   3) differential privacy.

Nella generalizzazione, invece, vengono classificate le tecniche di aggregazione e di "k-anonymity". Esse sono volte a impedire l'individuazione di persone interessate mediante il loro raggruppamento con almeno "k" altre persone. A tale scopo, i valori degli attributi sono sottoposti a una generalizzazione tale da attribuire a ciascuna persona il medesimo valore. Ad esempio, riducendo il grado di dettaglio di una località da città a Stato, si include un numero più elevato di persone interessate. Le date di nascita individuali possono essere generalizzate in una serie di date o raggruppate per mese o anno. Altri attributi numerici (ad esempio, retribuzioni, peso, altezza o il dosaggio di un farmaco) possono essere generalizzati mediante il ricorso a intervalli di valori (ad esempio, retribuzione 20.000€ – 30.000 euro).

9. Cfr. Gruppo di lavoro articolo 29, *Parere 05/2014 sulle tecniche di anonimizzazione*, *cit.*, in cui si legge che "*se un intruso riesce a determinare (ad esempio mediante un'analisi della correlazione) che due dati sono assegnati allo stesso gruppo di persone, ma non è in grado di identificare alcuna persona del gruppo, la tecnica fornisce una protezione contro l'individuazione, ma non contro la correlabilità*".

10. In tal senso, si rimanda a RUBINSTEIN I.S. AND GOOD N., *The trouble with Article 25 (and how to fix it): the future of data protection by design and default*, in International Data Privacy Law, 2020, Vol. 10, No. 1, [37-56]; in argomento, si veda anche VEALE M., BINNS R. e AUSLOOS J., *When data protection by design and data subject rights clash, in International Data Privacy Law*, 2018, Vol. 8, No. 2, [105-123].

11. In tal senso SUWANNAKIT M., *Book Review: Aurelia Tamo`- Larrieux, Designing for Privacy and its Legal Framework: Data Protection by Design and Default for the Internet of Things*, in *International Data Privacy Law*, 2019, Vol. 9, No. 4., [302-304]., secondo cui "*laws alone cannot lead to changes in practice because data protection stems from the design of technology*".

12. Le certificazioni possono essere rilasciate dal Garante o da Accredia (Ente unico nazionale designato dal Governo italiano in conformità alla normativa EN – ISO/IEC). Per un maggiore approfondimento, si rimanda a PEZZA F., sub. art.42, in RICCIO G. M., SCORZA G. e BELISARIO E. (a cura di), *GDPR e normativa privacy*, *Commentario*, ed. I, Milano, 2018, pg.384.

13. Per un'analisi approfondita, si rimanda a BRAVO F., *Sul bilanciamento proporzionale dei diritti e delle libertà "fondamentali", tra mercato e persona: nuovi assetti nell'ordinamento europeo?* in *Contr. e Impr*., 2018, 1, [190-216].
14. [1] RUBINSTEIN I.S. AND GOOD N., *The trouble with Article* 25, cit., pg. 38.

15. *Idem*, pg. 43.
16. Sul punto si veda OECD (2019), *Artificial Intelligence in Society, OECD Publishing, Paris*, pg. 103 e ss; si rimanda anche a RUBINSTEIN I.S. AND GOOD N., *The trouble with Article* 25, cit., pg. 45.

17. In particolare, con riguardo alla differential privacy, si rimanda a DE CRISTOFARO E., *An Overview of Privacy in Machine Learning*, *UCL & Alan Turing Institute*, 2020, pg. 9 ss.

18. Per approfondire si rimanda a SCHEIBNER J. E ALTRI, *Data protection and ethics requirements for multisite research with health data: a comparative examination of legislative governance frameworks and the role of data protection technologies*, in *Journal of Law and the Biosciences*, 1-30, pg. 27 e ss.
19. *Artificial Intelligence in Society*, cit.

20. *Data protection and ethics requirements*, cit.

21. ANDREA VENANZONI, *Smart cities* cit.

22. ANDREA VENANZONI, *Smart cities* cit.

Come ti scovo i bulli con l'intelligenza artificiale: il progetto "BullyBuster"

L'intelligenza artificiale può essere usata per individuare automaticamente i contenuti di cyberbullismo o per riconoscere comportamenti sospetti tramite l'utilizzo della biometria comportamentale e della "crowd analysis". Il progetto "acchiappabulli" di quattro atenei del Sud Italia

Di **Gian Luca Marcialis**, **Marco Micheletto** e **Giulia Orrù**, Università degli Studi di Cagliari - Dipartimento di Ingegneria Elettrica ed Elettronica

Bullismo e cyberbullismo sono fenomeni che, a causa della loro crescente diffusione, sono diventati una vera e propria emergenza sociale.

Le azioni violente che caratterizzano i due fenomeni, quali molestie verbali, aggressioni fisiche, persecuzioni, revenge-porn, ecc., sono generalmente condotte tra i giovani nell'ambiente scolastico ma l'alta diffusione di dispositivi mobili porta questa grave piaga sociale anche all'esterno delle scuole.

In questo contesto l'intelligenza artificiale può essere un'arma potente per identificare episodi di violenza e per combattere il bullismo sia nel mondo virtuale che in quello reale.

Attraverso l'apprendimento automatico, è possibile rilevare i modelli linguistici utilizzati dai bulli e dalle loro vittime e sviluppare regole per rilevare automaticamente i contenuti di cyberbullismo [1,2].

A livello di bullismo fisico, l'intelligenza artificiale può essere utilizzata per riconoscere comportamenti sospetti tramite l'utilizzo della biometria comportamentale [3] e della "crowd analysis" [4], cioè lo studio dei naturali movimenti di persone, gruppi di persone o oggetti.

Tale scenario ha portato alla creazione del progetto "BullyBuster - A framework for bullying and cyberbullying action detection by computer vision and artificial intelligence methods and algorithms", finanziato nell'ambito del bando relativo ai Progetti di Rilevante Interesse Nazionale (PRIN) del 2017[1], che vede coinvolti quattro gruppi di ricerca multidisciplinari afferenti a quattro atenei del Sud Italia (Università degli Studi di Bari Aldo Moro, Università degli Studi di Cagliari, Università degli Studi di Foggia, Università degli Studi di Napoli Federico II).

Il progetto BullyBuster

Il progetto prevede lo sviluppo di un sistema in grado di determinare se in un dato luogo fisico (scuola/piazza o altri contesti), durante la comunicazione social oppure tramite smartphone si stanno compiendo azioni prepotenti e violente, minacce o ingiurie.

Figura 1: *Logo del progetto PRIN2017 "BullyBuster - A framework for bullying and cyberbullying action detection by computer vision and artificial intelligence methods and algorithms".*

L' "acchiappabulli" sarà in grado di segnalare se è in corso un atto bullistico ovunque esso sia compiuto, permetterà di controllare i toni e le azioni degli aggressori, proteggendo le vittime ed allertando le forze dell'ordine di un pericolo imminente.

La cooperazione sinergica delle quattro università coinvolte prevede l'utilizzo di diverse fonti di dati: (1) un'analisi basata su video, tramite un processo di segmentazione e caratterizzazione della scena mediante descrittori strutturali temporali e spaziali, permetterà di rilevare specifiche azioni di bullismo sulla base dei movimenti della folla intorno alla vittima e, ove possibile, alla sua espressione facciale; (2) un'analisi testuale durante le comunicazioni digitali nei social network permetterà, mediante l'individuazione di parole e frasi tipiche di molestie, oppressione e stalking informatici, di individuare alcuni stati di disagio connessi a bullismo o cyberbullismo, come ansia e depressione; (3) un'analisi comportamentale mediante il rilevamento della dinamica della sequenza di tasti e l'analisi della pressione dei tasti o del touchscreen nei dispositivi elettronici (smartphone/tablet) sarà di supporto per evidenziare quanto più precocemente possibile il verificarsi di una condizione emotiva anomala, riconducibile ad un fenomeno di bullismo.

I modelli statistici e generativi alla base di questi strumenti sono ispirati da modelli di comportamento definiti degli psicologi del team. Il progetto è totalmente conforme con la normativa *privacy* e le modifiche introdotte con il GDPR grazie alla presenza nel team di ricercatori giuristi, che si occupano, inoltre, di indicare le conseguenze giuridiche dei comportamenti dei bulli e le possibili tutele per le vittime.

Un nuovo descrittore temporale per la rilevazione di eventi "anomali"

In questo articolo, ci focalizziamo su un prototipo per l'osservazione da videocamera di gruppi di soggetti non singolarmente identificabili, ma in grado di fornire sufficienti informazioni per

segnalare, in base a modelli comportamentali opportunamente codificati, eventi “anomali” come episodi di violenza o panico[2]. Il prototipo è stato sviluppato dallo staff dell’Università di Cagliari.

Esso nasce dalla definizione di un nuovo descrittore del comportamento dei gruppi inseriti in una certa area. L'obiettivo del descrittore, i cui dettagli sono stati pubblicati in [9], che si ispira all’algoritmo Local Binary Pattern monodimensionale (1D-LBP) [5], è quello di misurare, attraverso schemi appropriati, la velocità di formazione e dispersione di assembramenti nella folla. Tali schemi dipendono dal numero di gruppi osservati in una finestra temporale. Un'appropriata unità di misura, denominata "trit", rappresenta tre possibili stati dinamici su un determinato frame: il numero dei gruppi aumenta, diminuisce o rimane inalterato (Figura 2).

La nostra ipotesi è che le brusche variazioni nel numero di persone che compongono un gruppo siano riconducibili ad un evento anomalo, che può essere di conseguenza rilevato traducendo tali variazioni in sequenze di stringhe basate sul trit temporale, che pertanto saranno significativamente diverse da quelle relative ad una situazione senza anomalie.

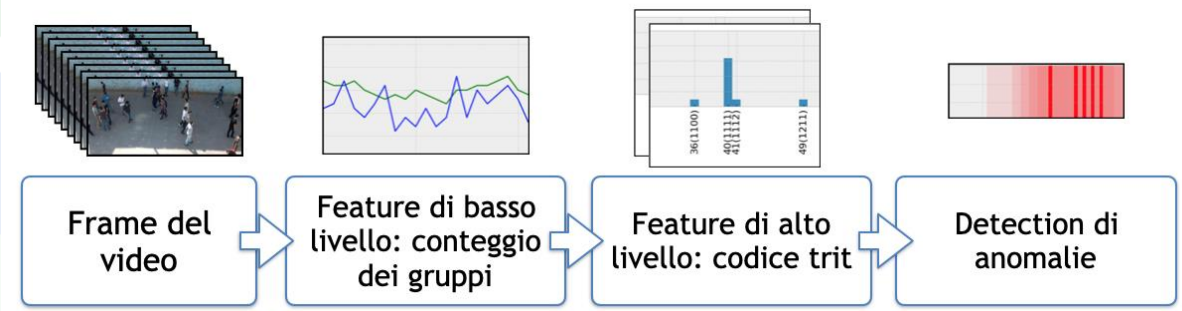

Figura 2: *Fasi di un sistema di rilevazione di anomalie: (1) si seleziona un sottoinsieme di frame dal video in esame; (2) si estraggono caratteristiche di basso livello per ottenere una stima del numero di gruppi in ogni scena; (3) si assegnano i trit corrispondenti ai vari stati dinamici; (4) rilevazione delle anomalie attraverso una specifica soglia.*

Per il conteggio dei gruppi sono stati confrontati 4 diversi metodi:

Conteggio manuale come *ground truth* (MC);

Clustering of Optical Flow (COF) [6];

Cascade Detector (CD) [7];

Blob Detector (BD).

Abbiamo quindi valutato gli istogrammi delle occorrenze dei codici trit. Al bin centrale dell'istogramma viene applicata una soglia rappresentativa dello stato di quiete, e funge da innesco per la rilevazione dell'anomalia.

Per valutare le prestazioni di un sistema di rilevamento di anomalie basato sul descrittore proposto, sono stati misurati il numero di anomalie correttamente rilevate, sulla base degli allarmi che ricadono in una finestra temporale corrispondente a circa 27 secondi e centrato sull'effettivo verificarsi dell'anomalia, e il numero di falsi allarmi.

La Tabella 1 riporta in sintesi i risultati sperimentali completi conseguiti sul dataset Motion Emotion [8], valutati in termini di *precision*, *recall* e *F1 Score*. In particolare la precision è l'abilità del sistema di non rilevare come anomalo una sequenza video senza anomalia (una precisione alta è data da pochi falsi allarmi), mentre la recall è la capacità del sistema di rilevare tutte le anomalie esattamente così come sono state definite nella modellazione del problema (è in un certo senso la "visione del mondo" del sistema intelligente).

In particolare, tali metriche sono definite dalle seguenti formule:

$$precision = \frac{TP}{TP + FP}$$

$$recall = \frac{TP}{TP + FN}$$

Dove TP (True Positive) è il numero di anomalie rilevate correttamente, FP (False Positive) è il numero di falsi allarmi, ovvero istanti in cui vi è una segnalazione di anomalia senza che essa sia avvenuta, e FN (False Negative) è il numero di anomalie non rilevate.

Precision e recall di solito sono considerati insieme per apprezzare la prestazione del sistema nella sua interezza. Tuttavia, abbiamo utilizzato anche un parametro di sintesi, lo score F1, quadrato della media armonica di precision e recall normalizzata rispetto alla loro media aritmetica, secondo la definizione:

$$F1_{\ score} = \frac{precision * recall}{precision + recall}$$

Precision, recalle e score F1 sono valori riportati in tutti i lavori allo stato dell'arte scientifico e dunque utilizzati anche nella nostra valutazione.

I parametri sono ottimizzati sulla base della metrica F1 Score. L'ottimizzazione è stata eseguita sia in modo supervisionato (F1 Score massimizzato su tutti i video) sia con una convalida incrociata *Leave-one-out* (F1 Score massimizzato su N -1 video) e test sul video escluso. La seconda ottimizzazione ci permette di apprezzare la stabilità prestazionale anche quando il sistema non ha avuto modo di estrarre informazioni da tutti i tipi di anomalia presenti nel data set.

	Supervised	Leave-one-out

	Precision	Recall	F1	Precision	Recall	F1
MC	88.89%	94.12%	91.43%	79.31%	71.87%	75.41%
COF	71.11%	88.89%	79.01%	52.50%	60.00%	56.00%
CD	75.00%	91.67%	82.50%	73.17%	83.33%	77.92%
BD	70.45%	86.11%	77.50%	56.52%	74.29%	64.20%

Tabella 1: *Risultati sperimentali ottenuti su tutti i video del ME dataset, utilizzando differenti metodi per il conteggio dei gruppi.*

È evidente dall'analisi dei risultati che il metodo scelto per il conteggio dei gruppi influisce direttamente sulle prestazioni del descrittore. Tra quelli esaminati, il più affidabile è il Cascade Detector che nel protocollo Leave-one-out raggiunge prestazioni di rilevamento migliori persino del conteggio manuale.

Inoltre, la differenza di prestazioni tra il protocollo supervisionato e quello Leave-one-out suggerisce che un'impostazione più accurata dei parametri consentirebbe un rilevamento più affidabile.

Casi particolari

Analizzeremo ora alcuni video del dataset utilizzato al fine di mostrare nel dettaglio il comportamento del descrittore proposto. Nei grafici seguenti, le linee verdi rappresentano le anomalie effettive, mentre le linee rosse le anomalie rilevate dal sistema. Se la linea rossa si trova in una zona verde chiaro, significa che l'anomalia è stata correttamente identificata.

Un primo esempio riguarda il video 009, dove è possibile osservare un caso di corretta rilevazione dell'anomalia con tutti i metodi di conteggio (Fig. 3). Il video è caratterizzato inizialmente da un flusso statico di individui, ovvero una folla strutturata, e di conseguenza un numero costante di gruppi. L'evento di panico genera un movimento della folla tale da essere identificato dai rilevatori di anomalie. Questo comportamento conferma esplicitamente le ipotesi alla base del nostro lavoro.

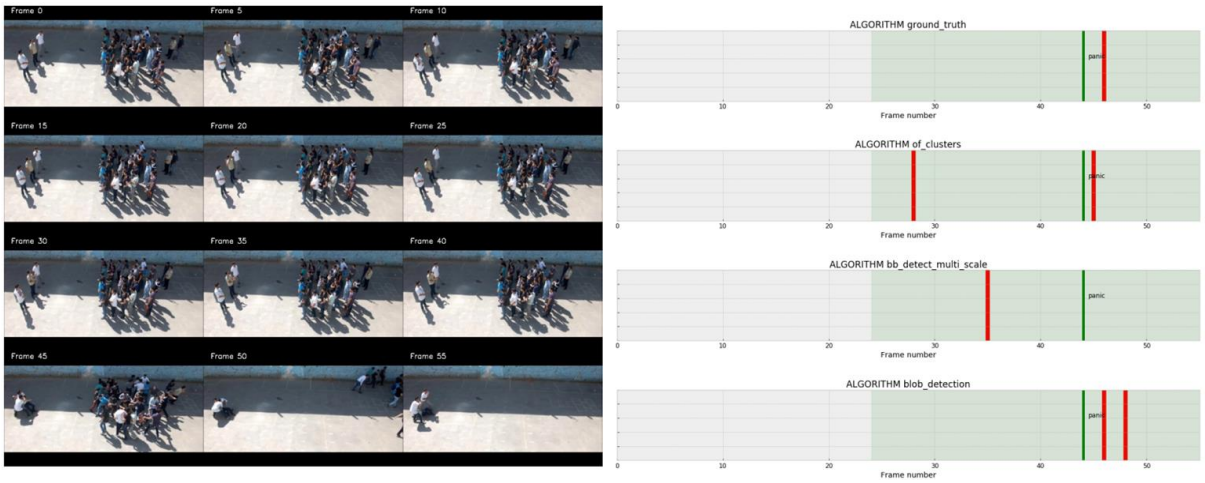

Figura 3: *Alcuni frame e descrizione grafica del rilevamento per il video 009: esso contiene una situazione di panico che è stata correttamente identificata con tutti i metodi di conteggio esaminati.*

D'altro canto, esistono delle situazioni in cui è particolarmente complicato distinguere dei veri eventi anomali da semplici cambiamenti nella composizione della folla. È il caso del video 023, riportato in Figura 4, caratterizzato da un gran numero di falsi allarmi. Ciò accade quando la folla non è ben strutturata e i piccoli cambiamenti nel numero di gruppo viene erroneamente interpretato dal sistema come molteplici anomalie. È possibile ridurre questo effetto agendo a livello di istogramma, evitando i picchi che rappresentano le piccole variazioni della folla.

Figura 4: *Alcuni frame e descrizione grafica del rilevamento per il video 023: non sono presenti anomalie all'interno del video, ma sistema rileva comunque dei falsi allarmi.*

Conclusioni

Questo lavoro rappresenta il nostro primo contributo alla lotta contro il bullismo, come primo risultato in seguito agli studi compiuti dal team dell'Università di Cagliari. Ci siamo concentrati sullo sviluppo di un rilevatore di eventi anomali su piccole o grandi folle attraverso algoritmi di *computer vision*. Abbiamo confermato la nostra ipotesi, secondo la quale è possibile identificare le rapide variazioni dei gruppi in una scena, codificandole attraverso una nuova unità di misura chiamata "trit". Sebbene siano necessari ulteriori studi per ridurre il numero di falsi allarmi in caso di cambiamenti lenti e controllati, il descrittore sviluppato è caratterizzato da un'elevata

versatilità, ed è pertanto completamente adattabile a seconda del contesto reale o del tipo di anomalia da rilevare.

Tale descrittore verrà quindi integrato nel sistema BullyBuster a partire dai modelli psicologici comportamentali individuati dal team dell'Università degli Studi di Foggia e affiancato ai sistemi di analisi di keystroke dynamics e text analysis per la detection di messaggi aggressivi legati ad attività di cyberbullismo, sviluppati dall'Università di Napoli "Federico II, e la fusione di soft biometrie come le dinamiche del tocco e di reazione alla battitura e della voce per rilevare eventuali sensazioni di panico o di aggressività nelle comunicazioni verbali attraverso messaggi vocali, sviluppata dall'Università degli Studi di Bari.

L'integrazione e il test del sistema verranno effettuate nel rispetto delle norme etiche generali e quelle oggetto delle autorizzazioni ai comitati etici, con osservanza del General Data Protection Regulation (https://www.gdpr.net/) e del Regolamento sull'Intelligenza artificiale presentato dalla Commissione Europea il 21 aprile 2021. Questi ultimi sono di fondamentale importanza per lo sviluppo dei metodi di videosorveglianza per la rilevazione di eventi anomali legati ad attività bullistica, in quanto affrontano il tema dell'identificazione biometrica real-time e ne definiscono gli ambiti applicativi in cui le forze dell'ordine possono operare. Il rispetto di questi regolamenti per tutte le fasi del progetto e la definizione di modelli psicologici e comportamentali da parte del team di psicologi è necessario e altamente rilevante in quanto il sistema BullyBuster è uno strumento pensato per operare e proteggere i minori.

Bibliografia

[1] K. Reynolds, A. Kontostathis and L. Edwards, "Using Machine Learning to Detect Cyberbullying" 2011 10th International Conference on Machine Learning and Applications and Workshops, Honolulu, HI, 2011, pp. 241-244.

[2] K. Dinakar, R. Reichart, and H. Lieberman, "Modeling the Detection of Textual Cyberbullying" in Proc. IEEE International Fifth International AAAI Conference on Weblogs and Social Media (SWM'11), Barcelona, Spain, 2011.

[3] Weiming Hu, Tieniu Tan, Liang Wang and S. Maybank, "A survey on visual surveillance of object motion and behaviors," in IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 34, no. 3, pp. 334-352, Aug. 2004.

[4] J. C. Silveira Jacques Junior, S. R. Musse and C. R. Jung, "Crowd Analysis Using Computer Vision Techniques," in IEEE Signal Processing Magazine, vol. 27, no. 5, pp. 66-77, Sept. 2010.

[5] Chatlani, N., & Soraghan, J. J. (2010). Local binary patterns for 1-D signal processing. 95-99.18th European Signal Processing Conference (EUSIPCO-2010), Aalborg, Denmark.

[6] B. K. Horn and B. G. Schunck, "Determining optical flow," Artificial intelligence, vol. 17, no. 1-3, pp. 185–203,1981.

[7] G. Bradski, "The OpenCV Library," Dr. Dobb's Journal of Software Tools, 2000.

[8] H. Rabiee, J. Haddadnia, H. Mousavi, M. Kalantarzadeh, M. Nabi, and V. Murino, “Novel dataset for fine-grained abnormal behavior understanding in crowd,” in 13th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 2016, pp. 95-101.

[9] G. Orrù, D. Ghiani, M. Pintor, G.L. Marcialis, and F. Roli, “Detecting Anomalies from Video-Sequences: a Novel Descriptor,” 2020 25th International Conference on Pattern Recognition (ICPR 2020), Milan, in press.

I quaderni di

Agenda Digitale.eu

NETWORK DIGITAL 360

Network Digital360 è il più grande network in Italia di testate e portali B2b dedicati ai temi della Trasformazione Digitale e dell'Innovazione Imprenditoriale, con oltre 50 fra portali, canali e newsletter.
Ha la missione di diffondere la cultura digitale e imprenditoriale nelle imprese e pubbliche amministrazioni italiane e di fornire a tutti i decisori che devono valutare investimenti tecnologici informazioni aggiornate e approfondite.
Il Network è parte integrante di Digital360HUB, il polo di Demand Generation di Digital360, che mette a disposizione delle tech company un'ampia gamma di servizi di comunicazione, storytelling, pr, content marketing, marketing automation, inbound marketing, lead generation, eventi e webinar.

VIA COPERNICO, 38
20125 - MILANO
TEL. 02 92852785
MAIL: MARKETING@DIGITAL4.BIZ

- Crittografia

La seconda ipotesi è costituita dalle politiche (o modalità) di attuazione della pseudonimizzazione che si dividono in

- pseudonimizzazione deterministica,
- pseudonimizzazione randomizzata dei documenti
- pseudonimizzazione completamente randomizzata.

[vi] RAMONA CAVALLI, *Anonimizzazione del dato: le tecniche possibili* cit.

[vii] Parere 05/2014 -WP 216 sulle tecniche di anonimizzazione.

[viii] Nella randomizzazione sono ricomprese le tecniche di:
1) aggiunta di rumore statistico.
2) permutazione.
3) differential privacy.

Nella generalizzazione, invece, vengono classificate le tecniche di aggregazione e di "k-anonymity". Esse sono volte a impedire l'individuazione di persone interessate mediante il loro raggruppamento con almeno "k" altre persone. A tale scopo, i valori degli attributi sono sottoposti a una generalizzazione tale da attribuire a ciascuna persona il medesimo valore. Ad esempio, riducendo il grado di dettaglio di una località da città a Stato, si include un numero più elevato di persone interessate. Le date di nascita individuali possono essere generalizzate in una serie di date o raggruppate per mese o anno. Altri attributi numerici (ad esempio, retribuzioni, peso, altezza o il dosaggio di un farmaco) possono essere generalizzati mediante il ricorso a intervalli di valori (ad esempio, retribuzione 20.000€ – 30.000 euro).

[ix] Cfr. Gruppo di lavoro articolo 29, *Parere 05/2014 sulle tecniche di anonimizzazione*, *cit.*, in cui si legge che "*se un intruso riesce a determinare (ad esempio mediante un'analisi della correlazione) che due dati sono assegnati allo stesso gruppo di persone, ma non è in grado di identificare alcuna persona del gruppo, la tecnica fornisce una protezione contro l'individuazione, ma non contro la correlabilità*".

[x] In tal senso, si rimanda a RUBINSTEIN I.S. AND GOOD N., *The trouble with Article 25 (and how to fix it): the future of data protection by design and default*, in International Data Privacy Law, 2020, Vol. 10, No. 1, [37-56]; in argomento, si veda anche VEALE M., BINNS R. e AUSLOOS J., *When data protection by design and data subject rights clash, in International Data Privacy Law*, 2018, Vol. 8, No. 2, [105-123].

[xi] In tal senso SUWANNAKIT M., *Book Review: Aurelia Tamo `- Larrieux, Designing for Privacy and its Legal Framework: Data Protection by Design and Default for the Internet of Things*, in *International Data Privacy Law*, 2019, Vol. 9, No. 4., [302-304]., secondo cui "*laws alone cannot lead to changes in practice because data protection stems from the design of technology*".

[xii] Le certificazioni possono essere rilasciate dal Garante o da Accredia (Ente unico nazionale designato dal Governo italiano in conformità alla normativa EN – ISO/IEC). Per un maggiore approfondimento, si rimanda a PEZZA F., sub. art.42, in RICCIO G. M., SCORZA G. e BELISARIO E. (a cura di), *GDPR e normativa privacy*, *Commentario*, ed. I, Milano, 2018, pg.384.

[xiii] Per un'analisi approfondita, si rimanda a BRAVO F., *Sul bilanciamento proporzionale dei diritti e delle libertà "fondamentali", tra mercato e persona: nuovi assetti nell'ordinamento europeo?* in *Contr. e Impr.*, 2018, 1, [190-216].

[xiv] RUBINSTEIN I.S. AND GOOD N., *The trouble with Article* 25, cit., pg. 38.

[xv] *Idem*, pg. 43.

[xvi] Sul punto si veda OECD (2019), *Artificial Intelligence in Society, OECD Publishing, Paris*, pg. 103 e ss; si rimanda anche a RUBINSTEIN I.S. AND GOOD N., *The trouble with Article* 25, cit., pg. 45.

[xvii] In particolare, con riguardo alla differential privacy, si rimanda a DE CRISTOFARO E., *An Overview of Privacy in Machine Learning*, *UCL & Alan Turing Institute*, 2020, pg. 9 ss.

[xviii] Per approfondire si rimanda a SCHEIBNER J. E ALTRI, *Data protection and ethics requirements for multisite research with health data: a comparative examination of legislative governance frameworks and the role of data protection technologies*, in *Journal of Law and the Biosciences*, 1-30, pg. 27 e ss.

[xix] *Artificial Intelligence in Society*, cit.

[xx] *Data protection and ethics requirements*, cit.

[xxi] ANDREA VENANZONI, *Smart cities* cit.

[xxii] ANDREA VENANZONI, *Smart cities* cit.