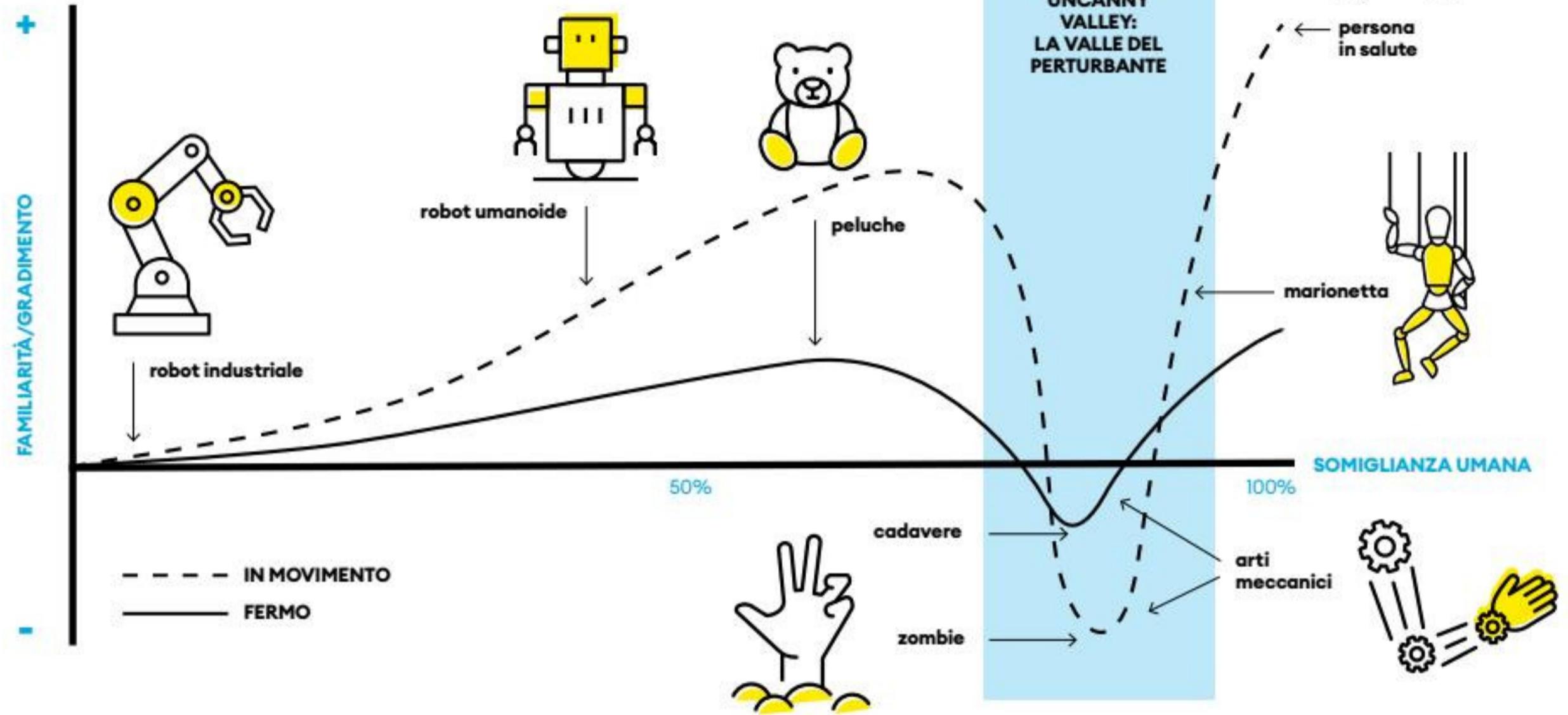


L'APPROFONDIMENTO

DI MARIELLA DAL FARRA

Presenze a confronto

Più le macchine assomigliano all'uomo meno ci piacciono



Il futuro sarà sempre più robotizzato e le macchine 'intelligenti' occuperanno spazi ancora maggiori nella nostra vita. Eppure i robot, soprattutto quelli con sembianze umane, possono inquietarci. Perché? Venite con noi, nella valle del perturbante...



Test d'intelligenza per assistenti virtuali

Ogni anno la LoupVentures, un'azienda che opera nell'ambito delle nuove tecnologie, sottopone gli assistenti virtuali presenti sul mercato a 800 domande ciascuno per testarne l'intelligenza in termini di: localizzazione ("Dov'è la caffetteria più vicina?"); commercio ("Ordinami un pacco di fazzoletti"); navigazione ("Quale autobus prendo per andare in centro?"); informazione ("Con chi gioca il Lugano domani?"); comando ("Ricordami di chiamare Antonio alle sei"). Ha "vinto" Google, con il 93% di risposte esatte, seguito da Siri (83%) e Alexa (80%). Il dato più rilevante, però, è che tutti e tre gli assistenti sono risultati, seppure in percentuali variabili, più "intelligenti" dell'anno scorso. In altre parole, stanno evolvendo...

Qual è il primo robot che ricordate di avere visto? O meglio, il primo androide? Il mio è, presumo, il cowboy assassino interpretato da Yul Brynner in *Il mondo dei robot* (1973): film intercettato per sbaglio alla tv in un noioso pomeriggio estivo che guardai fino alla fine con crescente apprensione e coinvolgimento. In particolare, ricordo la sensazione di straniamento trasmessami dal personaggio: la determinazione meccanica dei gesti, la rigidità della mimica facciale, lo sguardo lucente mai stemperato da un battito di ciglia... A proposito, è interessante notare come ancora oggi, in tempi di *deepfake* video prodotti da algoritmi *deep learning* che creano sequenze animate di fatto indistinguibili da quelle reali, **l'unico modo per non essere tratti in inganno sia proprio prestare attenzione alla frequenza con cui l'ipotetica "persona" batte le palpebre** (si veda anche il contributo di Siwei Lyu, "Detecting 'deepfake' videos in the blink of an eye", *The Conversation*, 29/8/2018). In un essere umano adulto e in buona salute, il cosiddetto "riflesso di ammiccamento" si verifica a intervalli di 2-10 secondi, e ogni battito di ciglia dura da un decimo a quattro decimi di secondo; quindi, se state guardando qualcuno e notate che ciò non avviene, fatevi una domanda...

Macchine evolute

Ma torniamo a *Il mondo dei robot*, che di recente ha anche ispirato una serie televisiva prodotta da HBO (*Westworld*) ormai giunta alla terza stagione. Quella prima, indelebile impressione di "alterità" prodotta dal Brynner-robot l'avrei poi ritrovata nei replicanti di *Blade Runner*, in *Terminator* e nei successivi film di fantascienza. Fino ad arrivare a *Neon*, l'assistente virtuale che Samsung ha presentato all'inizio di quest'anno al Consumer Electronics Show, la principale fiera dell'innovazione tecnologica di consumo. Ancora all'inizio del suo sviluppo, *Neon* è un chatbot ma, a differenza di Siri o di OK Google, possiede un aspetto fisico (un avatar) ed è in grado di conversare in maniera via via più personalizzata e pertinente con il proprio utente. Per esempio, uno dei *Neon* è un istruttore di yoga e può aiutarvi a perfezionare la pratica; un altro è un presentatore del telegiornale che ti riporta le notizie a cui sei maggiormente interessato; un'altra ancora è una consulente finanziaria che ti dà consigli per ottimizzare

il piano pensionistico. I video promozionali diffusi da Samsung attraverso un'oculata campagna di marketing suscitano sensazioni discretamente inquietanti, anche se sembra che, al momento, il prodotto "reale" sia ancora ben lontano dalle promesse dei suoi ideatori.

Area di 'disagio'

Si tratta di un'inquietudine di tipo particolare, forse lontana parente di quello che Freud chiamava "il perturbante" (*Das Unheimliche*, 1919; tradotto in inglese come *The Uncanny*). Lo strano disagio suscitato da qualcosa che sembra umano, ma che vi differisce quel tanto da lasciare intuire la sua natura "artificiale". Tre anni prima che *Il mondo dei robot* facesse la sua comparsa nelle sale cinematografiche, l'ingegnere giapponese Masahiro Mori, noto per il suo lavoro pionieristico nell'ambito dell'automazione, aveva pubblicato un importante articolo sulle reazioni emotive degli esseri umani alle entità non-umane. Mori (1970) ipotizzava che il gradimento di una persona nei confronti di una macchina tenda ad aumentare in misura proporzionale alla sua capacità di simulare la "qualità umana", ma solo fino a quando rimane facilmente distinguibile come "non-umana"; nel momento in cui la rassomiglianza diventa eccessiva, il gradimento espresso dalle persone precipita bruscamente, per poi risalire man mano che le discrepanze divengono più impercettibili.

Questo avallamento nel diagramma (vedi sopra, ndr) che incrocia l'asse della "qualità umana" con quello del "senso di familiarità e gradimento", è stato appunto definito *Uncanny Valley* (traducibile come "la valle del perturbante"): un effetto psicologico fra i più studiati negli ultimi anni, anche in ragione della spinta prodotta dalla graduale diffusione dell'Intelligenza Artificiale. Come affermano alcuni esperti del settore, "molti ricercatori hanno iniziato a spostare l'attenzione dall'aspetto esteriore della tecnologia contemporanea alle sue complesse abilità mentali, che includono l'esperienza emotiva (Gray & Wegner, 2012) e la simulazione di empatia (Liu & Sundar, 2018; Stein & Ohler, 2017). Ora che all'inquietudine generata dall'aspetto perturbante si è aggiunta quella derivata dall'intelligenza perturbante, diventa ancora più urgente districare le variabili che fanno sentire gli esseri umani a disagio in presenza

di una tecnologia sofisticata" (Stein et al., 2019). "Stay back, clever thing! Linking situational control and human uniqueness concerns to the aversion against autonomous technology", *Computers In Human Behavior*.

Riflessi atavici

La ricerca ha evidenziato principalmente due ordini di fattori. Il primo scaturisce dalla psicologia evoluzionista e si è concentrato sul ruolo giocato da alcuni indizi percettivi (per esempio, il non battere le palpebre) che potrebbero innescare una reazione avversativa automatica, inconscia, perché "interpretati" come indicatori di entità patogene (Ho et al., 2008), partner riproduttivi inadatti (Green et al., 2008) o individui psicopatici (Tinwell et al. 2013). Questa ipotesi sembra corroborata da un esperimento svolto da due ricercatori dell'Università di Princeton nel quale i soggetti coinvolti erano scimmie macaco (Ghazanfar & Shepherd, "Monkeys at the Movies: What Evolutionary Cinematics Tells Us About Film", *Projections*, vol. 5, 2011). Confrontate con immagini realistiche di macachi che però avevano un aspetto "sintetico", queste scimmie, che di solito tubano e schioccano le labbra per stabilire relazioni con i propri simili, distoglievano rapidamente lo sguardo e manifestavano segni di paura. Se poste di fronte a immagini meno realistiche o, al contrario, reali tout court, il timore scompariva. Questi nostri antichi cugini cadono dunque come noi nella "valle del perturbante": si tratta dunque di un riflesso atavico? Un po' contro-intuitivo, considerato che gli androidi sono un'acquisizione tutto sommato recente.

Il fattore normalità

L'altra spiegazione chiama in causa la violazione delle leggi fisiche. Secondo quest'ipotesi, un'entità che disattenda le aspettative acquisite sulla base dell'esperienza precedente (per esempio, un essere umano che si muove come un robot o un computer che manifesta reazioni emotive) determina una dissonanza cognitiva, che a sua volta genera l'impulso a evitare il contatto con l'oggetto che l'ha suscitata. In questa prospettiva, la sensazione di fascino/repulsione/straniamento associata agli automi sarebbe semplicemente "l'abito emotivo" dell'incongruenza rilevata a livello cognitivo inconscio.

Questa dissonanza è stata "fotografata" con la risonanza magnetica funzionale (fMRI) in un esperimento che ha coinvolto venti soggetti adulti confrontati con videoclip i cui protagonisti erano, rispettivamente: un robot con i circuiti in bellavista, un androide realistico ma imperfetto e un essere umano, tutti e tre impegnati in un'identica sequenza motoria. Le aree corticali "accese" mentre i soggetti osservano l'androide hanno una distribuzione simile a quella di quando guardano l'essere umano, ma ne coinvolgono di supplementari, evidenziando un aumento significativo dell'attività corticale (Saygin et al. 2012. "The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions", *Social cognitive and affective neuroscience*, aprile 2012).

Umano (almeno un po')

Gli autori interpretano questa maggiore attivazione come indice dell'aumentata richiesta di analisi a fronte di uno stimolo imprevedibile, in quanto l'aspetto non corrisponde al tipo di movimento. In altri termini, l'androide trasgredisce una regola codificata e aumenta quindi il margine d'errore percepito: l'incertezza che ne deriva innescherebbe la leva emotiva della repulsione, che è il modo più efficace per indurci a "tenerci alla larga". "Quando il Sistema Nervoso viene confrontato con la cosa che non dovrebbe esistere - per dirla alla Lovecraft - ecco che un'onda di errore predittivo dilaga nelle aree cerebrali in cui sono contenuti gli schemi motori": il nostro sistema ci sta segnalando che potremmo facilmente "sbagliarci". Nel progettare i dispositivi tecnologici materiali (i robot) o immateriali (gli assistenti virtuali) che verosimilmente saranno sempre più presenti nelle nostre vite, sarebbe dunque opportuno considerare con attenzione il grado di realismo che verrà loro conferito: parafrasando Nietzsche, "umano, ma non troppo umano" per favore.

**Sai chi sono?
Esempi di volti
generati da
un algoritmo**



La persona che stai guardando non esiste

Un dispositivo particolarmente efficace nel suscitare l'effetto "uncanny valley" è quello costituito dalla pagina web "This person does not exist" (thispersondoesnotexist.com): andateci e troverete il viso di una persona qualsiasi, che potrebbe appartenere a un uomo o a una donna; a un giovane, a un anziano, a un bambino; una faccia normale, forse asiatica o magari caucasica. Ricaricate la pagina e vedrete un altro volto, e poi un altro, e un altro ancora all'infinito. Facce di persone diverse, con espressioni più o meno decifrabili, alcune delle quali potrebbero sembrarvi familiari. Peccato che, appunto, non esistano: questi volti vengono generati da un algoritmo *deep learning* ogni volta che si clicca sull'icona del refresh, "esistono" per il tempo in cui l'utente tiene aperta la pagina e poi si dissolvono



nello stesso nulla da cui sono emerse, essendo costituite unicamente di stringhe alfanumeriche aggregate da una macchina intelligente e non, come ci era sembrato, di carne e sangue. Va bene, la trilogia di *Matrix* ci ha spiegato tutto in proposito, ma vederlo accadere "dal vivo" è un'altra cosa...

"This person does not exist" è opera dell'ingegnere di Uber Philip Wang che l'ha messa online il 12 febbraio, 2019 per sensibilizzare l'opinione pubblica sulle potenzialità del codice StyleGAN, la rete neurale che genera i volti. Il codice, introdotto dai ricercatori di Nvidia nel dicembre del 2018 e reso open source nel febbraio dell'anno seguente, può infatti essere impiegato per scopi illeciti come, per esempio, la falsificazione di documenti o la produzione di *deepfake* video.