23.Ottobre

Il perchè della sfiducia e della diffidenza verso l' AI

Tasti mai usati del mio telecomando
che guardo con grande diffidenza e paura.

Siamo sull'orlo di un momento cruciale nella storia della nostra specie, un momento in cui una creazione della nostra inventiva ha il potenziale per cambiare tutto. Per alcuni sarà la salvezza dell'umanità, mentre per altri potrebbe essere la nostra rovina. Stiamo entrando nell'era *dell'intelligenza artificiale (AI)* : una porta aperta in una stanza senza muri.

Sebbene *l'intelligenza artificiale* sia ancora lontana dalle macchine senzienti ritratte nella fantascienza, indubbiamente la creazione di algoritmi in grado di apprendere, comprendere il linguaggio e imitare alcuni aspetti della mente umana ha portato a enormi progressi.
Oggi *l'intelligenza artificiale* **viene già utilizzata** in centinaia di settori diversi.

Le macchine che utilizziamo quotidianamente stanno diventando più intelligenti, il che significa che l'IA non è più una tecnologia futuristica ma è sempre più integrata in ogni ambito della nostra vita. Dal suggerire quali libri potremmo voler acquistare online al potenziamento degli assistenti virtuali che abitano i nostri telefoni e altoparlanti intelligenti, alcune delle applicazioni sono più visibili di altre. In verità, l'intelligenza artificiale sta toccando e modificando le nostre vite molto più di quanto molti di noi si rendano conto.

-Le **banche** lo utilizzano per rilevare le frodi e prevedere i cambiamenti nei mercati azionari

-Le **compagnie di assicurazione** stanno impiegando l'intelligenza artificiale per aiutarle a produrre preventivi di polizza e valutare i sinistri

-Sta aiutando le **forze di polizia** a identificare i sospetti dalle immagini sgranate delle telecamere a circuito chiuso

-Nelle **aule dei tribunali** offre consigli ai giudici sull'opportunità di concedere condizioni di libertà su cauzione ai sospetti criminali

-Le macchine con la capacità di identificare le immagini **stanno aiutando i medici** a individuare le malattie.

-Gli algoritmi che utilizzano l'apprendimento automatico, uno dei rami principali dell'intelligenza artificiale, stanno aiutando **le auto a guida autonoma** a percorrere le nostre strade complesse

-Stanno aiutando **i linguisti** a decifrare le lingue perdute

-Anche sui voli, l'intelligenza artificiale viene utilizzata dai **controllori del traffico aereo** per aiutarci a tenerci al sicuro sia in volo che a terra

Sono molti i paesi che utilizzano ***"valutazioni del rischio criminale"***per stimare la probabilità che qualcuno condannato per un crimine recidivi in futuro.
Diverse aziende multinazionali stanno attualmente utilizzando sistemi basati sull'intelligenza artificiale durante ***i colloqui di lavoro*** per valutare l'onestà, lo stato emotivo e la personalità complessiva dei candidati.

Lo ***screening computerizzato*** del curriculum è una pratica delle risorse umane di lunga data e comune per eliminare i candidati che non soddisfano i requisiti per un annuncio di lavoro.
E sta aiutando le aziende a prendere decisioni su chi assumere e licenziare nel "mercato del lavoro"

E' inevitabile che emergano profonde questioni etiche su come e quando l'IA dovrebbe essere utilizzata e definire in che modo questi ***cervelli di silicio*** entrino nelle nostre vite e come potrebbero plasmare le nostre relazioni, il nostro lavoro e le nostre società.

Uno dei primi pionieri dell'informatica, **Alan Perlis**

una volta ha scherzato dicendo che "*un anno trascorso nell'intelligenza artificiale è sufficiente per far credere in Dio",* per la frustrazione per i problemi di produzione di reti di computer che potrebbero imitare il cervello umano.

Ad oggi le domande che aspettano una risposta sono sempre le stesse:
La prospettiva di creare macchine che possano sperare di competere con i nostri cervelli è ancora fuori portata?
Sono già in grado di fare le cose che gli umani possono solo sognare?
Di cosa è veramente capace l'IA?
Come può aiutarci a risolvere i nostri problemi?
E come dovremmo sentirci riguardo a questo futuro?

Gli ***algoritmi di apprendimento automatico*** con accesso a grandi set di dati stanno già prendendo miriadi di decisioni critiche, come diagnosi mediche, idoneità al welfare e assunzione di lavoro, tradizionalmente prese dagli esseri umani.

Tuttavia, episodi di pregiudizi e discriminazioni di alto profilo perpetuati da tali algoritmi stanno si sono (inconsapevolmente) introdotti durante il tuning del deep-learning erodendo la fiducia delle persone in queste decisioni, nonostante i responsabili politici e legislativi stanno proponendo regolamenti rigorosi su algoritmi complessi per proteggere le persone interessate dalle decisioni.

Anticipare e prepararsi ai rischi normativi è una preoccupazione significativa per i dirigenti ; tuttavia, la necessità più immediata è quella di affrontare **la sfiducia** pubblica che motiva la diffusa richiesta di regolamentare gli algoritmi.

Questa crescente ***mancanza di fiducia*** non solo può portare a politiche più severe che ostacolano l' innovazione , ma può anche tradursi in notevoli potenziali perdite economiche per le organizzazioni .

Sostanzialmente quattro fattori guidano la sfiducia del pubblico nei confronti delle **decisioni algoritmiche.**

Amplificazione dei pregiudizi :
gli algoritmi di apprendimento automatico amplificano i pregiudizi, sistemici o non intenzionali, nei dati di addestramento.

Opacità degli algoritmi :
gli algoritmi di apprendimento automatico sono scatole nere per gli utenti finali. Questa mancanza di trasparenza – indipendentemente dal fatto che sia intenzionale o intrinseca accresce le preoccupazioni circa la base su cui vengono prese le decisioni.

Disumanizzazione dei processi :
gli algoritmi di apprendimento automatico richiedono sempre più un intervento umano minimo o nullo per prendere decisioni. L'idea di macchine autonome che prendono decisioni critiche e che cambiano la vita evoca emozioni altamente polarizzate.

Responsabilità delle decisioni :
la maggior parte delle organizzazioni fa fatica a segnalare e giustificare le decisioni prodotte dagli algoritmi e non riesce a fornire misure di mitigazione per affrontare l'ingiustizia o altri esiti negativi. Di conseguenza, gli utenti finali non sono in grado di migliorare la loro probabilità di successo in futuro.

Cosa dovrebbero fare i Chief Data e/o gli Analytics Officer
Queste sono sfide difficili che non hanno soluzioni chiare o facili. Tuttavia, le organizzazioni devono agire ora per migliorare ***l'equità, la trasparenza e la responsabilità*** dei loro algoritmi e anticipare le normative. Ad esempio

Aumentare la consapevolezza dell'IA :
Istruire leader aziendali, data scientist, dipendenti e clienti sulle opportunità, i limiti e le preoccupazioni etiche dell'IA.
Formare i dipendenti per identificare i pregiudizi nei set di dati e nei modelli e incoraggiare discussioni aperte.
Guidare i dirigenti su quando l'IA fa davvero la differenza e quando lo faranno gli algoritmi decisionali tradizionali.

Creare un ecosistema per l'autoregolamentazione :
Istituire *team interdisciplinari per esaminare potenziali pregiudizi e problemi etici nei modelli algoritmici.*
Prevedere controlli multi livello con interventi umani per decisioni algoritmiche.
Delegare *la revisione e la certificazione da entità esterne per gli algoritmi critici. da entità esterne per*
Incorporare *la trasparenza nei modelli di dati e fornisci ricorso agli utenti finali per presentare petizioni sui risultati dell'algoritmo.*

Influenza le normative globali :
Collaborare con i governi, enti pubblici e privati, gruppi di riflessione e associazioni di settore per creare politiche che bilanciano la regolamentazione con l'innovazione.

Sebbene le normative siano necessarie, le organizzazioni che migliorano in modo proattivo la fiducia delle persone nelle decisioni algoritmiche possono evitare la perdita di entrate, definire normative e politiche eque e investimenti IA a prova di futuro contro l'impatto normativo negativo.

In conclusione:
Possiamo risolvere il "problema di fiducia" dell'IA?
Per affrontare e provare a risolvere la diffidenza degli utenti, i produttori di applicazioni AI dovrebbero **smettere di *promettere troppo*, *diventare più trasparenti* e prendere in considerazione la *certificazione di terze parti***.

In un sondaggio sui consumatori statunitensi, quando è stato presentato loro un elenco di servizi IA popolari (ad esempio, assistenti domestici, pianificazione finanziaria, diagnosi mediche e assunzioni), il 41,5% degli intervistati ha dichiarato di non fidarsi di nessuno di questi servizi. Solo il 9% degli intervistati ha affermato di fidarsi dell'IA per i propri dati finanziari e solo il 4% si fida dell'IA nel processo di assunzione dei dipendenti.

In un altro sondaggio, a 2.000 consumatori statunitensi è stato chiesto**: "Quando pensi all'intelligenza artificiale, quali sentimenti descrivono meglio le tue emozioni?"**
"Interessato" è stata la risposta più comune (45%), ma è stata seguita da vicino da "preoccupato" (40,5%), "scettico" (40,1%), "incerto" (39,1%) e "sospettoso" (29,8%).

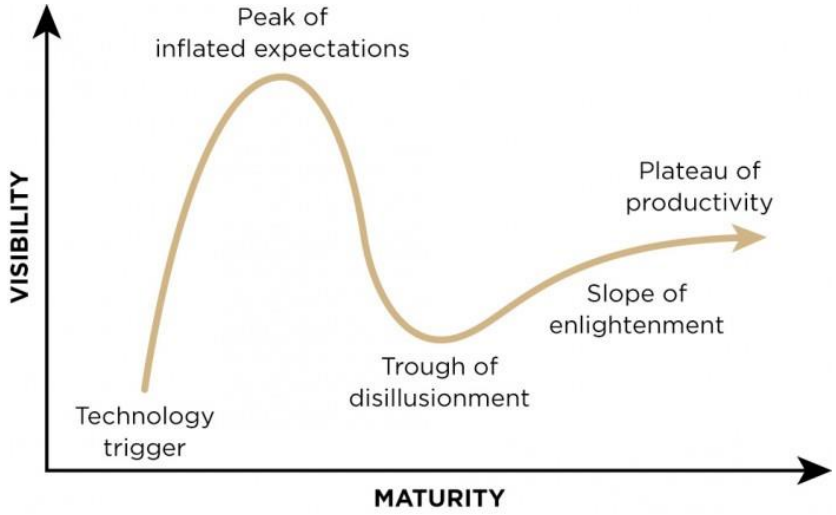
La società di ricerca IT Gartner suggerisce che tecnologie come il cognitive computing, l'apprendimento automatico, il deep learning e i consulenti esperti cognitivi sono al culmine del loro **ciclo di hype** e si stanno dirigendo verso il **"trogolo della disillusione".** (vedi allegato)

Allegato

Che cos'è l'Hype Cycle?

L'Hype Cycle è la rappresentazione grafica di come una **tecnologia evolverà nel tempo**. Include le fasi di adozione e di applicazione della tecnologia, ossia come essa potrebbe essere potenzialmente rilevante nella risoluzione di problemi delle aziende di vari settori. Ogni anno la curva viene aggiornata in base a diversi fattori come, ad esempio, il progresso scientifico, l'avanzamento tecnologico e l'estensione dell'adozione della tecnologia da parte delle aziende o dei consumatori.

Ogni Hype Cycle è suddiviso in **cinque fasi** principali relative al **ciclo di esistenza** della tecnologia in questione. Sull'**asse orizzontale** abbiamo la variabile del **tempo** che trascorre. Si passa dal momento in cui essa è soltanto un'idea o una scoperta senza possibili applicazioni, al momento in cui la tecnologia diventa parte integrante del sistema produttivo, delle vite dei consumatori e in generale dell'economia. È molto importante notare che la velocità con cui le tecnologie compiono questo percorso non è costante: alcune già in pochi anni percorrono l'intero ciclo, altre rimangono addirittura più di dieci anni nella medesima fase. Sull'**asse verticale** viene misurata la **visibilità della tecnologia** in termini di popolarità, se è nota o sconosciuta, se suscita interesse nella società e nelle imprese o meno.



La prima fase viene chiamata **"***Innovation Trigger",* letteralmente *"Grilletto dell'innovazione"*, o meglio – A*ttivatore di innovazione*. Una nuova tecnologia viene scoperta, ma ancora non si conoscono le sue possibili applicazioni e spesso non esiste un prodotto tangibile, una versione beta di un programma o un servizio provvisorio da poter provare. In questa fase intervengono massivamente i media, la notizia si diffonde in tutto il mondo, gli imprenditori cominciano a sognare che cosa potrebbero fare se quella tecnologia (spesso si tratta di innovazioni di tipo dirompente) diventasse disponibile al grande pubblico. Il problema principale di questa fase è: questa innovazione sarà qualcosa di commercializzabile oppure no?

La seconda fase prende il nome di *"Peak of Inflated Expectations",* che significa *"Picco delle aspettative esagerate"*. Vengono raggiunti alcuni traguardi nel miglioramento della tecnologia, ma spesso vi sono molti **fallimenti** correlati ai vari esperimenti, dunque la maggior parte delle aziende resta soltanto a guardare.

Nella terza fase, ***"Trough of Disillusionment"*** (ossia *"Trogolo della disillusione"*) c'è una sostanziale **diminuzione dell'interesse** da parte del mercato nella nuova tecnologia. Questo avviene poiché la sperimentazione non conduce a risultati concreti, dunque gli investitori scarseggiano e molte aziende che offrono la nuova tecnologia tendono a fallire. Soltanto alcune sopravvivono, in quanto sono riuscite ad ottenere dei risultati o lavorano direttamente per soddisfare le **aziende early adopters**.

La quarta fase, *"Slope of Enlightenment"* (*"Salita della consapevolezza"*) rappresenta una **svolta**. La tecnologia inizia a **diffondersi** e le imprese cominciano a prendere coscienza dei vantaggi che questa potrebbe apportar loro. Gli sviluppatori della tecnologia iniziano a produrre nuove generazioni, correggendo eventuali difetti nei prodotti, disorganizzazioni nei servizi o bug nei programmi. I finanziamenti e gli **investitori crescono** in numero, visti i risultati, ma alcune imprese rimangono ancora caute, in disparte, ad osservare quale piega prendono gli eventi.

La quinta ed ultima fase prende nome di *"Plateau of Productivity",* l'A*ltopiano della Produttività*. La tecnologia si rivela **ampiamente applicabile** e dunque comincia ad essere adottata sempre da più imprese. Ora esiste un mercato più competitivo tra i vari fornitori, quindi ci sono dei criteri specifici per valutare se un fornitore è affidabile o meno.
In breve, il **ciclo di Gartner** serve alle aziende, agli imprenditori e agli investitori per **capire se** una **tecnologia emergente è promettente**. Gli investimenti nelle nuove tecnologie portano con sé dei rischi evidenti. Un investitore che punta sulla tecnologia molto presto potrebbe non ricavarne guadagno. In molti preferiscono aspettare delle prove concrete sui possibili modi di utilizzo della tecnologia prima di investire. Infine ci sono alcuni investitori che sono più avversi al rischio e dunque preferiscono investire quando il mercato è più maturo e sono sicuri di ricevere del valore aggiunto dal loro investimento.

Allegato

Atteggiamenti pubblici nei confronti degli algoritmi informatici

Gli americani esprimono ampie preoccupazioni per l'equità e l'efficacia dei programmi per computer che prendono decisioni importanti nella vita delle persone

AARON SMITH

Gli algoritmi sono ovunque intorno a noi, utilizzano enormi archivi di dati e analisi complesse per prendere decisioni con impatti spesso significativi sugli esseri umani. Ci consigliano libri e film da leggere e guardare, fanno emergere notizie che pensano potremmo trovare rilevanti, stimano la probabilità che un tumore sia canceroso e prevedono se qualcuno potrebbe essere un criminale o un rischio di credito meritevole. Ma nonostante la crescente presenza di algoritmi in molti aspetti della vita quotidiana, un'indagine del **Pew Research Center** sugli adulti statunitensi rileva che il pubblico è spesso scettico nei confronti di questi strumenti quando vengono utilizzati in varie situazioni della vita reale.

Questo scetticismo abbraccia diverse dimensioni. A livello generale, il 58% degli americani ritiene che i programmi per computer rifletteranno sempre un certo livello di pregiudizi umani, sebbene il 40% ritenga che questi programmi possano essere progettati in modo privo di pregiudizi. E in vari contesti, il pubblico teme che questi strumenti possano violare la privacy, non cogliere le sfumature di situazioni complesse o semplicemente mettere le persone che stanno valutando in una situazione ingiusta.
Anche le percezioni pubbliche del processo decisionale algoritmico sono spesso altamente contestuali. Il sondaggio mostra che tecnologie altrimenti simili possono essere viste con supporto o sospetto a seconda delle circostanze o dei compiti a cui sono assegnate.

Per misurare le opinioni degli americani comuni su questo argomento relativamente complesso e tecnico, il sondaggio ha presentato agli intervistati quattro diversi scenari in cui i computer prendono decisioni raccogliendo e analizzando grandi quantità di dati pubblici e privati. Ciascuno di questi scenari era basato su esempi del mondo reale di processi decisionali algoritmici (vedere la barra laterale allegata) e includeva: un punteggio di finanza personale utilizzato per offrire ai consumatori offerte o sconti; una valutazione del rischio penale delle persone in libertà vigilata; un programma di screening automatizzato del curriculum per i candidati al lavoro; e un'analisi computerizzata dei colloqui di lavoro. Il sondaggio includeva anche domande sui contenuti a cui gli utenti sono esposti sulle piattaforme dei social media come un modo per valutare le opinioni di algoritmi più rivolti ai consumatori.
I seguenti sono tra i principali risultati.

Il pubblico esprime ampie preoccupazioni circa l'equità e l'accettabilità dell'utilizzo dei computer per il processo decisionale in situazioni con importanti conseguenze nel mondo reale

de questi esempi di processo decisionale algoritmico come ingiusti nei confronti delle persone che i sistemi basati su computer stanno valutando. In particolare, solo circa un terzo degli americani pensa che il colloquio di lavoro video e gli algoritmi del punteggio di finanza personale sarebbero equi per i candidati e i consumatori. Alla domanda diretta se ritengono accettabile l'uso di questi algoritmi, la maggioranza del pubblico afferma che non sono accettabili.
Due terzi degli americani (68%) trovano inaccettabile l'algoritmo del punteggio di finanza personale e il 67% afferma che l'algoritmo di analisi del lavoro video assistito da computer è inaccettabile.
Ci sono diversi temi che suscitano preoccupazione tra coloro che trovano questi programmi inaccettabili.
Alcune delle preoccupazioni più importanti menzionate in risposta alle domande a risposta aperta includono quanto segue:

Violano la privacy.
Questa è la principale preoccupazione di coloro che trovano inaccettabile il punteggio di finanza personale, citata dal 26% di tali intervistati.
Sono ingiusti.
Coloro che si preoccupano dello scenario del punteggio di finanza personale, della vignetta del colloquio di lavoro e dello screening automatico dei candidati al lavoro hanno spesso citato preoccupazioni sull'equità di tali processi nell'esprimere le proprie preoccupazioni.
Rimuovono l'elemento umano dalle decisioni importanti.
Questa è la principale preoccupazione di coloro che trovano inaccettabile il concetto di screening automatico del curriculum (il 36% lo menziona), ed è una preoccupazione importante tra coloro che sono preoccupati per l'uso dell'analisi del colloquio di lavoro video (16%).
Gli esseri umani sono complessi e questi sistemi non sono in grado di catturare le sfumature.
Questo è un tema relativamente coerente, menzionato in molti di questi concetti come qualcosa di cui le persone si preoccupano quando considerano questi scenari.
Questa preoccupazione è particolarmente evidente tra coloro che trovano inaccettabile l'uso dei punteggi di rischio criminale. Circa la metà di questi intervistati menziona preoccupazioni legate al fatto che tutti gli individui sono diversi o che un sistema come questo non lascia spazio alla crescita o allo sviluppo personale.

Gli atteggiamenti nei confronti del processo decisionale algoritmico possono dipendere fortemente dal contesto

Nonostante la coerenza di alcune di queste risposte, il sondaggio evidenzia anche i modi in cui l'atteggiamento degli americani nei confronti del processo decisionale algoritmico può dipendere fortemente dal contesto di tali decisioni e dalle caratteristiche delle persone che potrebbero essere interessate.
Questa dipendenza dal contesto è particolarmente evidente negli atteggiamenti contrastanti del pubblico nei confronti dei concetti di punteggio di rischio criminale e punteggio di finanza personale. Percentuali simili della popolazione pensano che questi programmi sarebbero efficaci nel fare il lavoro che dovrebbero fare, con il 54% che pensa che l'algoritmo del punteggio di finanza personale farebbe un buon lavoro nell'identificare le persone che sarebbero buoni clienti e il 49% pensa che il rischio criminale il punteggio sarebbe efficace nell'identificare le persone che

meritano la libertà condizionale. Ma una quota maggiore di americani pensa che il punteggio di rischio criminale sarebbe giusto per coloro che sta analizzando. La metà (50%) pensa che questo tipo di algoritmo sarebbe giusto per le persone che sono in attesa di libertà vigilata, ma solo il 32% pensa che il concetto di punteggio di finanza personale sarebbe giusto per i consumatori.
Quando si tratta degli algoritmi che sono alla base dell'ambiente dei social media, il livello di comfort degli utenti nel condividere le proprie informazioni personali dipende anche in larga misura da come e perché i loro dati vengono utilizzati. Una maggioranza del 75% degli utenti dei social media afferma che sarebbe a suo agio nel condividere i propri dati con quei siti se fossero utilizzati per consigliare eventi a cui potrebbero voler partecipare. Ma quella quota scende a solo il 37% se i loro dati vengono utilizzati per inviare messaggi da campagne politiche.

In altri casi, diversi tipi di utenti offrono opinioni divergenti sulla raccolta e l'utilizzo dei propri dati personali. Ad esempio, circa due terzi degli utenti di social media di età inferiore ai 50 anni ritengono accettabile che le piattaforme di social media utilizzino i propri dati personali per consigliare di entrare in contatto con persone che potrebbero voler conoscere. Ma questa opinione è condivisa da meno della metà degli utenti di età pari o superiore a 65 anni.

Gli utenti dei social media sono esposti a un mix di contenuti positivi e negativi su questi siti

Gli algoritmi danno forma al moderno panorama dei social media in modi profondi e onnipresenti. Determinando i tipi specifici di contenuto che potrebbero essere più attraenti per ogni singolo utente in base ai suoi comportamenti, influenzano la dieta mediatica di milioni di americani. Ciò ha portato a temere che questi siti stiano indirizzando un numero enorme di persone verso contenuti "coinvolgenti" semplicemente perché li fanno arrabbiare, infiammano le loro emozioni o servono altrimenti come cibo spazzatura intellettuale.

Su questo fronte, il sondaggio fornisce ampie prove del fatto che gli utenti dei social media sono regolarmente esposti a contenuti potenzialmente problematici o preoccupanti su questi siti. In particolare, il 71% degli utenti dei social media afferma di aver mai visto contenuti che li fanno arrabbiare, con il 25% che afferma di vedere spesso questo tipo di contenuti. Allo stesso modo, circa sei utenti su dieci affermano di incontrare spesso post eccessivamente esagerati (58%) o post in cui le persone fanno accuse o iniziano discussioni senza aspettare di avere tutti i fatti (59%).
Ma come spesso accade per le esperienze degli utenti sui social media in senso più ampio, questi incontri negativi sono accompagnati da interazioni più positive. Sebbene il 25% di questi utenti affermi di incontrare spesso contenuti che li fanno arrabbiare, una quota comparabile (21%) afferma di incontrare spesso contenuti che li fanno sentire in contatto con gli altri. E una quota ancora maggiore (44%) riferisce di vedere spesso contenuti che li rendono divertenti.

Allo stesso modo, gli utenti dei social media tendono ad essere esposti a un mix di comportamenti positivi e negativi da parte di altri utenti su questi siti. Circa la metà degli utenti (54%) afferma di vedere un mix uguale di persone cattive o prepotenti e persone gentili e di supporto. I restanti utenti si dividono tra quelli che vedono più meschinità (21%) e gentilezza (24%) su questi siti. E la maggioranza degli utenti (63%) afferma di vedere un mix uguale di persone che cercano di essere ingannevoli e di persone che cercano di segnalare informazioni inesatte, mentre il resto è equamente diviso tra coloro che vedono più persone che diffondono inesattezze (18%) e più persone che cercano di correggere quel comportamento (17%).
Altri risultati chiave di questo sondaggio su 4.594 adulti statunitensi condotto dal 29 maggio all'11 giugno 2018 includono:

L'atteggiamento del pubblico nei confronti del processo decisionale algoritmico può variare in base a fattori legati alla razza e all'etnia.

Solo il 25% dei bianchi pensa che il concetto di punteggio di finanza personale sarebbe equo per i consumatori, ma tale quota sale al 45% tra i neri. Allo stesso modo, il 61% dei neri pensa che il concetto di punteggio di rischio criminale *non* sia giusto per le persone in libertà vigilata, ma tale quota scende al 49% tra i bianchi.

Circa tre quarti del pubblico (74%) pensa che i contenuti che le persone pubblicano sui social media non riflettano il modo in cui la società si sente più in generale riguardo a questioni importanti, anche se il 25% pensa che i social media dipingano un ritratto accurato della società.

Gli adulti più giovani hanno il doppio delle probabilità di affermare di vedere spesso contenuti sui social media che li fanno sentire divertiti (54%) in quanto sono contenuti che li fanno arrabbiare (27%). Ma gli utenti di età pari o superiore a 65 anni incontrano questi due tipi di contenuti con una frequenza più comparabile. Il sondaggio rileva che il 30% degli utenti più anziani vede spesso contenuti sui social media che li fanno sentire divertito, mentre il 24% vede spesso contenuti che li fanno arrabbiare.