

Intervista a

Sneha Revanur

“la Greta Thunberg dell’IA”



Sneha Revanur 18 anni appartiene a una generazione che è stata immersa nella tecnologia "fin dal primo giorno". All'età di 15 anni fondò un'organizzazione, **Encode Justice**, per affrontare le implicazioni dannose dell'intelligenza artificiale. Attualmente studentessa del secondo anno di college, è stata inserita nella lista delle 100 persone più influenti **della rivista Time Magazine a settembre** ed è stata descritta da **Politico** come la "**Greta Thunberg dell'intelligenza artificiale**".



In questa intervista con il direttore esecutivo *del Bulletin* **Dan Drollette Jr., Revanur** descrive come si è interessata al problema e cosa l'ha portata a fondare un gruppo della società civile guidato dai giovani e incentrato sull'intelligenza artificiale. Racconta di come i suoi amici siano passati dal pensare "Sneha fa questa cosa dell'intelligenza artificiale e, tipo, salta le lezioni e va a Washington a volte" all'esprimere sincera preoccupazione per alcuni dei problemi associati all'intelligenza artificiale, che includono rendere obsoleti i lavori dei loro sogni, sorvegliarli 24 ore su 24 e inserendo falsi profondi che inquinano i loro social media. E tutto ciò in aggiunta alle vere e proprie catastrofi potenziate dall'intelligenza artificiale.

Encode Justice conta ora 900 giovani membri in 30 paesi in tutto il mondo e ha fatto paragoni con i precedenti movimenti per il clima e il controllo delle armi guidati dai giovani. **Revanur** e i suoi colleghi furono invitati alla Casa Bianca; ha partecipato agli sforzi volti a legiferare per un futuro dell'IA migliore e più sicuro; ha scritto editoriali per *The Hill*; e ha contribuito a sconfiggere con successo un'iniziativa elettorale statale che avrebbe inserito algoritmi distorti generati dall'intelligenza artificiale nel processo giudiziario, dimostrando quanto una sola persona può realizzare a livello di base.

Dan Drollette Jr.: Da dove vieni?

Sneha Revanur: Sono originario della California, San Jose, proprio nel cuore della Silicon Valley.

Drollette: Ciò ha influenzato il tuo interesse per l'intelligenza artificiale.

Revanur: Lo direi sicuramente. Mia sorella maggiore lavora nel settore tecnologico ed entrambi i miei genitori sono ingegneri del software. Quindi la tecnologia è sempre stata davanti ai miei occhi e mi ha portato a pensare a come guidarla nella giusta direzione. I miei genitori scherzano sempre su come intendo regolarli.

Ma sul serio, tutto questo ha significato che sono stato esposto presto a una cultura secondo cui ogni problema nella società può essere risolto con una sorta di soluzione computazionale, che si tratti di un'app mobile, di un modello di apprendimento automatico o di qualche altro meccanismo per rispondere a qualcosa. **Penso che ci sia sempre stata questa visione secondo cui l'innovazione era una sorta di soluzione miracolosa.**

E penso che questa visione sia esplosa negli ultimi anni, con l'ascesa dell'intelligenza artificiale. Dico spesso che, se fossi nato altrove, **Encode Justice** non esisterebbe. Penso davvero che crescere nella Silicon Valley, nel tipo di famiglia in cui ho vissuto, sia stato davvero fondamentale nel plasmarmi e nel plasmare la formazione di questa organizzazione.

Drollette: Come mai l'organizzazione è stata chiamata "Encode Justice"?

Revanur: Ho inventato il nome. E ho pensato molto alle sue connotazioni: tutto era molto intenzionale. Voglio dire, avrei potuto scegliere un nome per l'organizzazione che contenesse una visione molto negativa della tecnologia.

Ma ciò che penso sia così potente nel nome "**Encode Justice**" è che cattura la sensazione che l'obiettivo della nostra organizzazione non è fermare tutta la tecnologia, né porre fine all'innovazione. Stiamo invece cercando di reimmaginare ciò che abbiamo e di inserire la giustizia nel quadro di questi sistemi fin dall'inizio. In realtà è un invito all'azione, invece che un invito a chiudere le cose o ad arrendersi.

E penso che sia un approccio davvero potente. Anche se siamo sull'orlo di minacce potenzialmente catastrofiche, rimango fermamente convinto che **se agiamo rapidamente, possiamo farcela; dobbiamo solo muoverci per stabilire alcune regole della strada.** Se lo facciamo, l'intelligenza artificiale può ancora essere una forza positiva; può aprire regni di possibilità.

Quindi, credo che il messaggio racchiuso nel nome Encode Justice sia ancora vero e abbia lo stesso significato dopo tutti questi anni.

Drollette: Quando è nata l'organizzazione?

Revanur: Nell'estate del 2020.

Drollette: Quindi, se sei al secondo anno di college in questo momento, allora a quel tempo saresti stato...

Revanur: Quindici anni.

Drollette: Mi racconti come è nato?

Revanur: E' una storia interessante. Qualche anno fa, mi sono imbattuto in un'indagine condotta da ProPublica che ha scoperto tassi piuttosto sconcertanti di pregiudizi razziali in un algoritmo utilizzato dai sistemi giudiziari a livello nazionale per valutare il rischio che una determinata persona infranga nuovamente la legge. [1]

Il problema con l'algoritmo era che **aveva il doppio delle probabilità di prevedere erroneamente** che gli imputati neri rispetto agli imputati bianchi sarebbero recidivi, cioè recidiverebbero.

È stato un brusco risveglio per me, perché come la maggior parte delle persone, tendevo a considerare gli algoritmi come perfettamente oggettivi, perfettamente scientifici e perfettamente neutrali. È difficile per noi concettualizzare che queste formule matematiche apparentemente impenetrabili potrebbero effettivamente, sai, fornire ingiustizie o trasmettere alcuni degli aspetti peggiori della società umana.

Mi è diventato chiaro che questo era un argomento che mi avrebbe interessato.

Un paio di anni dopo, mi sono imbattuto in un provvedimento elettorale nel mio stato natale, la California, chiamato **California Proposition 25**, che cercava di sancire l'uso di algoritmi molto simili in tutto lo stato nel nostro sistema legale. Se fosse passato, avrebbe sostituito il già ingiusto sistema della cauzione in contanti con uno strumento algoritmico molto **simile a quello che era stato sostanzialmente incriminato** in questa indagine di ProPublica, e sarebbe stato utilizzato praticamente in tutte le cause istruttorie. .

Mi sono reso conto che c'era pochissima consapevolezza pubblica sui potenziali pericoli derivanti dall'introduzione di integratori algoritmici nel processo. [2] E c'era ancora meno partecipazione giovanile al dibattito su questa tecnologia, e penso che la partecipazione giovanile sia fondamentale, perché **qualunque sia il sistema, è il sistema che ereditiamo entrando nell'età adulta.**

Così ho iniziato a radunare i miei coetanei.

Insieme, abbiamo contattato gli elettori, creato contenuti informativi, collaborato con organizzazioni comunitarie in tutta la California, gestito banche telefoniche e alla fine siamo riusciti a sconfiggere la misura con un margine di circa il 13%.

A quel punto eravamo solo una campagna incentrata su un unico provvedimento elettorale. Ma dopo la nostra vittoria iniziale mi sono reso conto che abbiamo questa incredibile rete di giovani, non solo dalla California, ma da tutto il mondo – circa **900 studenti delle scuole superiori e universitari, provenienti da 30 paesi diversi** – che sono entusiasti e pensano in modo più critico sulle implicazioni dell'intelligenza artificiale.



Ho iniziato a pensare che avremmo potuto prendere questo e fare davvero qualcosa.

Questo è il punto in cui siamo diventati un'organizzazione più formale e a tutti gli effetti, in grado di affrontare altri progetti, come la tecnologia di riconoscimento facciale, i problemi di sorveglianza e privacy, l'erosione democratica e tutti i tipi di altri rischi derivanti dall'intelligenza artificiale, dalla disinformazione allo spostamento della manodopera. .

E penso che nell'ultimo anno sia diventato evidente che **ci sono molti rischi nuovi e imprevedibili di danni catastrofici derivanti dall'IA**; ad esempio, ChatGPT 4 è già stato jailbroken per generare

istruzioni per creare bombe. **[3]** *E l'intelligenza artificiale destinata alla scoperta di farmaci è già stata riproposta per progettare decine di migliaia di armi chimiche letali in poche ore.* **[4]**

Stiamo assistendo a sistemi che sfuggono al nostro controllo; ci stiamo muovendo verso sistemi sempre più potenti e sofisticati che potrebbero potenzialmente avere gravi danni esistenziali per l'umanità.

*Di conseguenza, organizzazioni come **Encode Justice** si stanno orientando verso la garanzia che possiamo prepararci a tali minacce, senza perdere allo stesso tempo di vista le realtà con cui siamo già faccia a faccia.*

Drollette: Avevi la minima idea che Encode Justice avrebbe avuto così tanto successo? Quante persone ne facevano parte quando hai iniziato?

Revanur: *Penso che all'inizio fossero solo 20 o 30 ragazzi. E molti di loro provenivano dalla mia scuola o da scuole vicine, tutte in California.*

E a quel punto, era piuttosto sbilanciato verso la California, perché stavamo lavorando su una questione californiana.

Ad essere sincero, onestamente non avrei mai immaginato che sarebbe cresciuto così tanto.

*E penso che il motivo per cui è stato accolto così bene è che è davvero un prodotto dei tempi. **Nell'ultimo anno abbiamo assistito a un'esplosione assoluta di interesse, quasi isterico, per l'intelligenza artificiale.** Se ciò non fosse avvenuto, non ci sarebbe stata tutta questa attenzione e visibilità sul lavoro che stiamo svolgendo. Siamo semplicemente saltati nello spazio esattamente nel momento giusto. È piuttosto sorprendente.*

Drollette: Quali sono le prospettive per Encode Justice? Credo che voi ragazzi stiate lavorando a qualcosa chiamato "Progetto per una Carta dei diritti dell'intelligenza artificiale"?

Revanur: *In realtà si tratta di un progetto che l'Ufficio per le politiche scientifiche e tecnologiche della Casa Bianca (OSTP) ha pubblicato l'anno scorso nel 2022. Il mio coinvolgimento è stato nel consigliare l'OSTP sulla creazione di tali principi, assicurandomi che riflettessero le priorità dei giovani.*

*In realtà siamo entrati in contatto per la prima volta con OSTP all'inizio del 2022, quando stavano iniziando a rimuginare su quel progetto. C'è stata una breve pausa quando il progetto è caduto nel dimenticatoio mentre ci sono stati alcuni cambiamenti nella leadership dell'agenzia, ma durante l'estate del 2022 abbiamo fatto molto sostegno, contattando i senatori e assicurandoci che fosse riportato in cima alle priorità elenco. Abbiamo scritto un editoriale su The Hill, in collaborazione con Marc Rotenberg del Center for AI and Digital Policy, chiedendo al nuovo direttore nominato dell'OSTP di **ridefinire le priorità della Carta dei diritti dell'IA.***

*E alla fine il quadro è stato rilasciato. È sicuramente un ottimo punto di partenza, ma al momento la struttura e i principi della Carta dei diritti dell'IA non sono applicabili, sono semplicemente un progetto. Quindi, ovviamente, c'è molto lavoro da fare per garantire che stiamo dando seguito a quel lavoro critico approvando effettivamente le normative con efficacia. **[5]***

*Penso che il fatto che Washington stia finalmente facendo appello alla volontà politica di intraprendere azioni significative sia indice dell'urgenza della questione. E quindi spero davvero che riusciremo a tradurre alcune delle idee e dei principi molto promettenti contenuti nel progetto per una **Carta dei diritti dell'intelligenza artificiale** in regolamenti realmente applicabili. **[6]***

Drollette: Come giornalista, penso di aver notato un cambiamento nel modo in cui la stampa di solito tratta le Big Tech. Circa 10 anni fa, tutto ciò che ho letto sulla stampa popolare

sull'informatica tendeva ad essere sulla falsariga di un servile tipo di copertura "rah, rah, tutto è meraviglioso, la tecnologia non può mai fare nulla di sbagliato". Hai l'impressione tra i tuoi colleghi che ci sia più consapevolezza che la tecnologia può portare problemi oltre che benefici?

Revanur: *Penso che le cose stiano cambiando, ma ad essere sincera, non credo che siamo ancora arrivati a quel punto. Penso che ci sia ancora questo tipo di senso residuo prevalente di ottimismo quasi incontrollato e incondizionato.*

*Per essere chiari, anch'io **condivido parte di questo ottimismo, nel senso che credo che la tecnologia abbia il potenziale per essere una forza per la trasformazione positiva della società. Penso che ci sia così tanto che la tecnologia potrebbe fare per l'umanità.***

Ma penso che abbiamo visto in prima persona che i problemi sorgono.

Quindi, anche se penso che la situazione stia cambiando, ci vorrà ancora un po' di tempo per completarlo.

*E penso che se la marea sta cambiando, lo sta facendo in modo non uniforme. Sta accadendo in particolare nella mia generazione, mentre le generazioni più anziane sono più lontane dagli impatti in prima linea della tecnologia, e quindi sono meno scettiche al riguardo. Quindi sì, penso che si stia lentamente ma inesorabilmente tendendo nella direzione di un **ottimismo più qualificato piuttosto che incondizionato**, ma non in modo uniforme.*

Drollette: Quindi forse si potrebbe dire che la tecnologia può essere una forza positiva ma deve essere attivamente indirizzata in quella direzione, soprattutto da parte delle persone colpite?

Revanur: *Penso che inquadrare questo come una questione di governo sia appropriato, perché penso che implichi che ci sia un dovere umano e una responsabilità morale da parte nostra nel fare questo governo. E penso che ciò avverrà sotto forma di regole di strada davvero significative per l'intelligenza artificiale, che arriveranno sotto forma di un'azione urgente per garantire che stiamo affrontando sia i rischi immediati che i rischi a lungo termine, ancora non realizzati, derivanti dall'intelligenza artificiale. . E quindi dipenderà davvero dalla nostra azione rapida.*

*E penso che proprio ora ci troviamo in un momento **in cui potrebbe benissimo essere possibile perdere un significativo controllo umano sui sistemi di intelligenza artificiale come lo conosciamo.** E penso che questo sia il motivo per cui è incredibilmente importante agire rapidamente, perché altrimenti i costi potrebbero essere irrecuperabili.*

Drollette: Una delle cose che ho sentito è che generalmente i giovani di oggi sono un po' apatici, che non sentono di poter davvero fare molta differenza. Con questo intendo dire che i giovani possono vedere quali sono i problemi, ma non pensano di poter contribuire a una soluzione. È questo il genere di cose che hai osservato? Oppure sei l'eccezione che conferma la regola, l'anomalia nel campus, che è andata a Washington per spingere per il cambiamento?

Revanur: *Sicuramente non direi affatto che sono anomalo. Penso che sia davvero importante sottolineare che mentre guido questa organizzazione, non lo faccio solo come singola persona: c'è un movimento di 900 studenti delle scuole superiori e universitari intorno a me.*

*Quindi penso che **non sarebbe giusto descrivere la mia generazione come apatica** quando ovviamente tutto il lavoro che sto facendo è supportato e rafforzato da questa incredibilmente grande coalizione di giovani provenienti letteralmente da tutto il mondo.*

*Anche se penso che un paio di anni fa eravamo decisamente in una posizione di apatia, perché penso che le persone non fossero in grado di concettualizzare l'impatto che l'intelligenza artificiale stava avendo sulla loro vita quotidiana. Penso che **in passato le persone tendessero a vedere***

L'intelligenza artificiale come un fenomeno tecnico completamente astratto e completamente distaccato dalla realtà vissuta.

*Ma penso che nell'ultimo anno, soprattutto, le persone di tutto il mondo, soprattutto i giovani, abbiano iniziato a riconoscere gli impatti che l'intelligenza artificiale potrebbe avere su di noi: il rischio di allucinazioni dovute a grandi modelli linguistici, l'impatto degli algoritmi dei social media. Voglio dire, **l'intelligenza artificiale sta diventando parte di ogni aspetto della nostra vita quotidiana: se fai domanda per un lavoro, ci sono algoritmi che curano elenchi di lavori per te e ti dicono effettivamente per cosa candidarti. Esistono algoritmi che esaminano il tuo curriculum, una volta che hai effettivamente inviato la domanda. Se vieni processato, ci sono algoritmi che ti valutano come imputato. C'è sorveglianza ovunque, utilizzando telecamere di sorveglianza abilitate al riconoscimento facciale che hanno la capacità di tracciare la tua posizione e la tua identità. [7] Quindi penso davvero che le persone stiano diventando sempre più consapevoli dell'ubiquità dell'intelligenza artificiale.***



*E con questo riconoscimento, penso che sempre più persone stiano diventando ansiose per **un futuro distopico**, e che l'ansia si stia traducendo in azioni politiche concrete. Non è più una questione astratta di cui si occupano solo accademici e dottorandi.*

*E questo è uno dei motivi per cui, credo, organizzazioni guidate da giovani come **Encode Justice** stanno prendendo piede.*

Drollette: E fai tutto questo mentre sei ancora uno studente universitario a tempo pieno: ti destreggi tra questo attivismo e allo stesso tempo ti occupi di lezioni, esami, stage e tutto il resto?

Revanur: È decisamente impegnativo. In questo momento, ti sto parlando tramite Zoom dalla biblioteca del Williams College, e poi ho un problema a cui devo rispondere immediatamente dopo questa chiamata. Quindi, sai, c'è molto da fare ed è decisamente difficile.

*Ma penso che ciò che mi ha aiutato a superare tutto questo è il fatto che sto facendo questo lavoro insieme a tanti miei colleghi che si trovano in una posizione molto simile: ogni membro di **Encode Justice** frequenta la scuola superiore o l'università. Ciò significa che tutti noi stiamo affrontando una serie di responsabilità molto simili in termini di avere un corso completo, stage, lavoro, responsabilità familiari e obblighi al di fuori dei nostri compiti scolastici.*

*Ci assumiamo questo compito perché **siamo consapevoli che il nostro futuro collettivo dipende da questo**. Affrontiamo questo tema da percorsi di vita diversi, background diversi, punti di vista diversi, perché il tempo stringe e la nostra generazione deve agire. Quindi, ci stiamo prendendo un*

po' di tempo della nostra vita per mettere davvero insieme le nostre teste e lavorare su questo. E penso che sia davvero potente e davvero bello.



Riferimenti

[1] Vedi “Machine Bias: esistono software utilizzati in tutto il paese per prevedere i futuri criminali. Ed è prevenuto nei confronti dei neri”, di Julia Angwin, Jeff Larson, Surya Mattu e Lauren Kirchner. ProPublica, 23 maggio 2016. Disponibile su <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

[2] Per ulteriori informazioni, vedere "I californiani dovrebbero respingere la Proposizione 25: riformare il sistema preliminare al processo senza utilizzare valutazioni del rischio di parte dal punto di vista razziale". Human Rights Watch, 29 luglio 2020. Disponibile su <https://www.hrw.org/news/2020/07/29/us-californians-should-reject-proposition-25>

[3] Il cosiddetto “jailbreaking” è un metodo per rimuovere restrizioni e limitazioni in modo che sia possibile apportare modifiche vietate e accedere a contenuti dannosi. Vedere "Ricerca: GPT-4 Jailbreak Facilmente sconfigge le barriere di sicurezza" di Roger Montti, *Search Engine Journal* , 23 ottobre 2023. Accessibile all'indirizzo <https://www.searchenginejournal.com/research-gpt-4-jailbreak-easily-defeats-parapetti-di-sicurezza/498386/>

[4] Vedere "I sistemi di scoperta di farmaci basati sull'intelligenza artificiale potrebbero essere riproposti per creare armi chimiche, avvertono i ricercatori", di Rebecca Sohn, *Scientific American* , 21 aprile 2022. Accessibile all'indirizzo <https://www.scientificamerican.com/article/ai-drug -i-sistemi-di-scoperta-potrebbero-essere-riutilizzati-per-produrre-armi-chimiche-ricercatori-avvertono/>

[5] Cfr. “A first take on the White House executive order on AI: A Great Start that might doesn't go far abbastanza” pubblicato sul sito web *del Bulletin of the Atomic Scientists* , 30 ottobre 2023, all'indirizzo <https://thebulletin .org/2023/10/a-first-take-on-the-white-house-executive-order-on-ai-a-great-start-that-forhaps-doesnt-go-long-enough/>

[6] Il progetto per una Carta dei diritti dell'intelligenza artificiale si basa essenzialmente su cinque principi . Dice che “**dovresti essere protetto da sistemi non sicuri o inefficaci**”; “**Non dovresti essere discriminato da algoritmi e i sistemi dovrebbero essere utilizzati e progettati in modo equo**”; “**Dovresti essere protetto da pratiche abusive sui dati tramite protezioni integrate e dovresti avere autorità su come vengono utilizzati i dati che ti riguardano**”; “**Dovresti sapere che viene utilizzato un sistema automatizzato e capire come e perché contribuisce a ottenere risultati che hanno un impatto su di te**”; e “**Dovresti essere in grado di rinunciare, se appropriato, e avere accesso a una persona che possa considerare e risolvere rapidamente i problemi che incontri**”.

Il testo completo della Carta dei diritti dell'AI è disponibile all'indirizzo <https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf>

[7] Vedi “Lo stato di sorveglianza high-tech non è limitato alla Cina: intervista con Maya Wang di Human Rights Watch” pubblicato nel *Bulletin of the Atomic Scientists* l'8 settembre 2022, all'indirizzo <https://thebulletin.org/premium/2022-09/lo-stato-di-sorveglianza-high-tech-non-è-limitato-alla-cina-intervista-a-maya-wang-di-human-rights-watch/>

Domani 21 gennaio:

Cos'è la Malattia X e perché è nelle notizie?

L'Organizzazione Mondiale della Sanità utilizza il termine "Malattia X" per riferirsi a un'infezione che potrebbe causare la prossima epidemia o una nuova pandemia globale.

Il divario nell'aspettativa di vita tra uomini e donne si sta riducendo in tutto il mondo

L'aspettativa di vita degli uomini in tutto il mondo sta raggiungendo quella delle donne, probabilmente a causa del calo dei decessi legati al fumo e all'alcol.