

# L'algoritmo etico?

AKA Cosa può essere effettivamente  
automatizzato?

Teresa Numerico

[teresa.numerico@uniroma3.it](mailto:teresa.numerico@uniroma3.it)

# Outline

- Il progetto di un sistema digitale senza attriti apre questioni etiche, ma soprattutto politiche
- Due diverse modalità di accedere al sapere: conoscenza relazionale e situata o conoscenza gerarchica e astratta
- Seguire le regole è sempre il frutto di scelte e pratiche
- Che fare?

Dall'etica alla politica: come definire le regole per entrare nel mondo digitalizzato? Chi ha il diritto di stabilirle?

## **LA DATIFICAZIONE DEI FENOMENI: UN MONDO SENZA ATTRITI**

# Libraries of the future



- It is both our hypothesis and our conviction that people can handle the major part of their **interaction with the fund of knowledge better by controlling and monitoring the processing of information than by handling all the detail directly themselves**



Licklider 1965, p. 28

# L'obiettivo dei procognitive systems

- A basic part of the over-all aim for procognitive systems is to get the user of the fund of knowledge into something more nearly like an **executive's or commander's position**. He will still read and think and, hopefully, have insights and make discoveries, but **he will not have to do all the searching** [...] all the **transforming**, nor all the **testing** for matching or compatibility that is **involved in creative use of knowledge**

Licklider 1965, p. 32

# I dati sono più di semplici rappresentazioni

- I dati offrono una visione dei fenomeni che contribuisce alla loro produzione, perché il loro ruolo come "rappresentazione" dipende da come li interpretiamo
- L'infrastruttura digitale mostra una sovrapposizione: il software è fatto della stessa 'materia' dei dati che interpreta. Questo crea sia opportunità che rischi nell'analizzare i fenomeni
- Da un lato, i dati possono rappresentare i fenomeni, Dall'altro, sono parzialmente indistinguibili dagli algoritmi: software e regole per interpretarli
- I dati creano una nuova componente sintetica del sapere: il collegamento mancante tra la situazione concreta e il sistema algoritmico digitale necessario per interpretarla senza frizioni

# AI è il prodotto di sistemi socio- tecnici

- Spesso dimentichiamo che per poter ottenere i dati digitali in una forma direttamente utilizzabile dai sistemi algoritmici è necessario il lavoro di tante persone
- Si stima che più di cento milioni di persone lavorino a vario titolo nell'ambito dell'intelligenza artificiale
- Svolgono ruoli diversi:
  - produrre i dati,
  - etichettarli,
  - eliminare le parti troppo oscure che non si vogliono usare nell'addestramento
  - Programmare i sistemi algoritmici
  - Svolgere il ruolo di reinforcement learning by human feedback (RLHF)
  - Definire gli obiettivi dei sistemi e controllare che producano profitti, o almeno che rendano le azioni delle aziende che se ne occupano sempre appetibili

# Che succede se esternalizziamo la ragione?

- Se esternalizziamo la ragione tramite i sistemi algoritmici e socio-tecnici stiamo esercitando una nuova esternalizzazione industriale delle capacità di memoria e di discernimento umane
- Una volta che il processo dovesse essere compiuto non ci saranno più spazi per il dubbio, per l'errore per la confusione perché non avremo più a che fare con il mondo, ma solo con una sua esteriorizzazione digitale, priva perfino di una eccedenza intangibile
- Non ci sarà spazio tra rappresentazioni e i relativi fenomeni, considerato che non sarà possibile esperire nessun accesso extra verso di loro
- La scelta di effettuare questa esternalizzazione prima che epistemica sarà politica



# Potere e progresso

- Acemoglu e Johnson (che hanno ricevuto il Nobel 2024 per l'economia) sostengono che non tutte le tecnologie costituiscono un progresso per la collettività
- Perché lo siano è necessario che siano messe al servizio di tutti e non solo di quelli che le hanno messe a punto
- Fare questo passaggio è compito delle istituzioni politiche



Come evitare di allucinare il mondo riconoscendo invece il carattere sociale e relazionale dell'apparenza?

**ALCUNI MITI CIRCA IL POTERE DELL'IA  
NELL'INTERPRETARE I FENOMENI**

## L'errore epistemico, etico, ontologico e politico di Narciso

---

- Narciso non ha riconosciuto l'immagine per quello che è, ne è rimasto catturato dal riflesso fidandosi della falsa ipotesi che indicasse la presenza di un altro in carne ed ossa, rimanendo paralizzato nell'equivoco

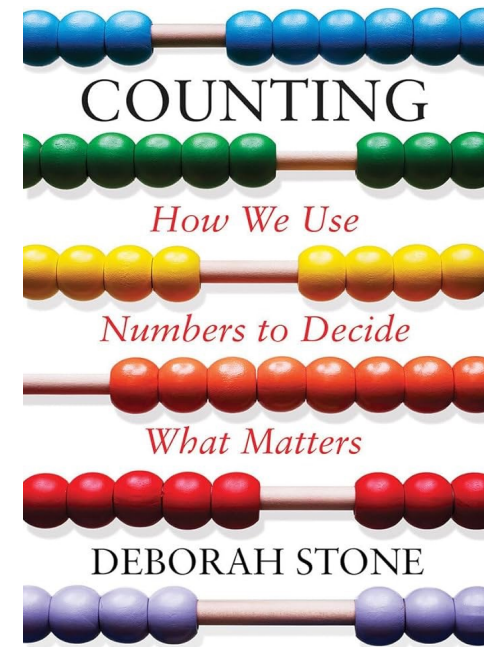


# Il mito che più è meglio

- Automated systems often operate under the hypothesis that *more data equals better insights* when interpreting data
- Big Data Rhetoric promotes a focus on volume, often at the expense of relevance, prioritizing quantity over the quality and contextual value of fewer, well-qualified data points
- While according to Varoquaux G. Luccioni A.S. Whittaker M. (2024, 21 sept.) *Hype sustainability and the price of the Bigger-is-better paradigm in AI* (Arxiv.org) it is not true that the improved performance is always the product of increased scale, or that all interesting problems addressed by AI require large-scale models
- According to Gigerenzer (2023), in uncertain situations – where critical understanding is inherently ambiguous – complex assessments and reliance on extensive data can hinder good judgment, simplified approaches are often more effective in navigating uncertainty

# Numeri come interpretazioni

- If every number begins with a judgment, and if we allow numbers to determine people's fates, we should hold numbers to the same ethical standards we hold our judges to. We should expect those who count to disclose and justify their judgments.
- The more we outsource decisions to automated counting systems, the more important it becomes to authorize humans to make exceptions. Above all, we shouldn't use numbers to cover over our deep conflicts and ethical dilemmas. That's what the Founding Fathers did. They reconciled the North and South by counting slaves as three-fifths of a person in the federal census. Some questions can't be answered with a number. To count well, we need humility to know what can't or shouldn't be counted.



## che cos'è l'astrazione?

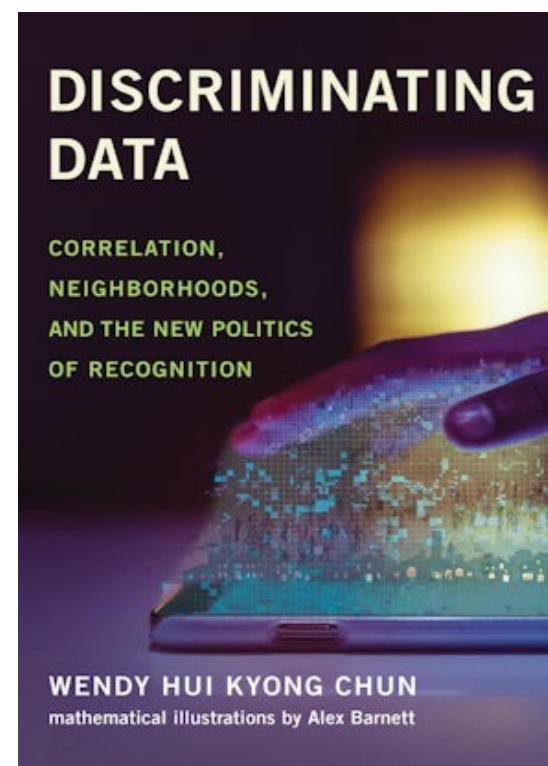
- If we conceive of abstraction simply as a process of isolating certain properties, we express only a very unimportant aspect of such a problem solution. To understand abstraction essentially means understanding what has to be abstracted rather than merely knowing how it has to take place. To understand the abstraction leading to an elegant solution of the problem means understanding how the solution can really be found.

Damerow 1996 Abstraction and representation

# Il caso dell'omofilia

- Wendy Chun showed very clearly the risks of generalization and abstraction in analysing the judgements about homophily to explain the creation of groups that seem to share similar behaviors and preferences

Chun (2018 e 2021)



# Il contenuto sintetico intossica gli algoritmi di apprendimento

- When most of the content will come from generative AI tools the system could implode because the result of using the generated content as model will render more probable the more common solutions, excluding the others

Shumailov, I., Shumaylov, Z., Zhao, Y. *et al.* AI models collapse when trained on recursively generated data. *Nature* **631**, 755–759 (2024).  
<https://doi.org/10.1038/s41586-024-07566-y>





# L'interpretazione è relazionale

- Algorithmic approaches often assume that context can either be disregarded or anticipated through statistical probability distributions of words or token distances.
- However, these statistics rely on the assumption that past texts are relevant for interpreting and interacting with future texts, based on the premise of a stable world scenario.
- Yet, the most valuable insights often emerge from exceptions—those unique, unpredictable phenomena.
- Unfortunately, for algorithms to function effectively, these exceptions are downplayed or excluded.
- Interpretation, as Emily Bender suggests, is not universal or uniform but a relational and situated process, shaped by context and interaction

## On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?



Emily M. Bender<sup>1\*</sup>, Timnit Gebru<sup>2\*</sup>,  
Angelina McMillan-Major<sup>1</sup>, Shmargaret Shmitchell<sup>3</sup>

<sup>1</sup> University of Washington <sup>2</sup> Black in AI <sup>3</sup> The Aether  
\*These authors contributed equally.

**LA DISCREZIONALITÀ IMPLICITA DI  
SEGUIRE UNA REGOLA**

# What is automation?

- **La questione**

- Come possiamo esercitare l'automazione in modo responsabile?
- Quali sono i rischi e i benefici del suo esercizio

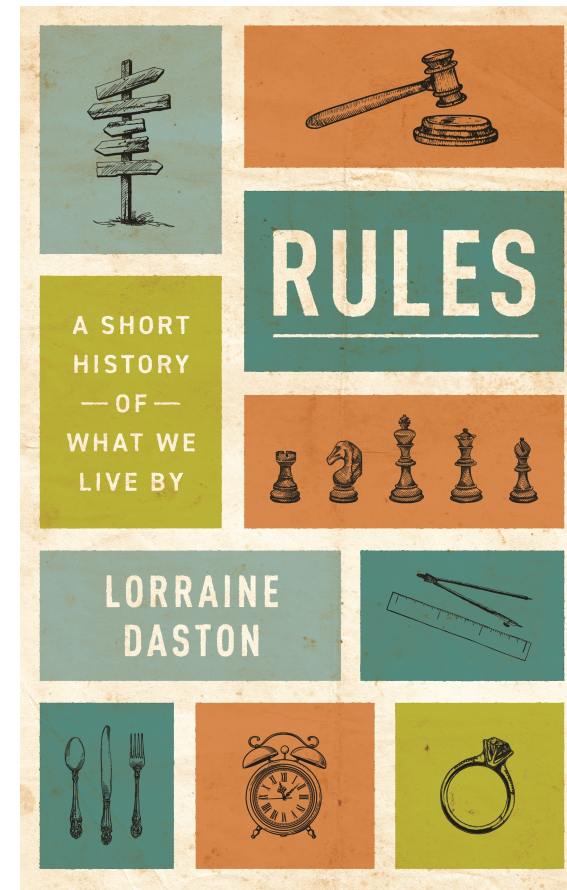
- **Metafore equivocate** While usiamo termini come "learning" e "intelligence" per i sistemi socio-tecnici, ma i loro processi sono differenti da quelli umani, perché orientati da obiettivi strettamente economici

- **La sfida della validazione**

- L'automazione spesso distrugge lo status quo senza un chiaro benchmark su come validare le sue decisioni per misurarne la prestazione rispetto a questioni come la giustizia sociale e la non discriminazione
- Nei contesti nei quali manca la chiarezza sulla funzione obiettivo i processi decisionali automatizzati sono ambigui e incerti

# Discrezione e regole

- Under what circumstances do explicit rules no longer need the support of implicit ones? One answer – Wittgenstein's answer – is never
- Even the algorithmic rules [...] cannot evade interpretation
- Hardly a rule can be applied without discretion, judgement and analogy



**CHE FARE?**

# Abbatere la barriera tra le due culture

- Le discipline umanistiche e sociali (humanities and social sciences) hanno un compito particolarmente arduo, ma prezioso non sui contenuti ma sui metodi: coltivare la singolarità dell'interpretazione e la costruzione di senso come valore, come processo dinamico, originale e molteplice
- Lo possono fare perché sono meno colonizzate di altre dalle procedure quantificatorie, forse per la loro condizione di marginalità e scarsa profittabilità
- Ma questa posizione laterale potrebbe essere un valore capace di riattivare pratiche di consapevolezza e resistenza rispetto non alle macchine, ma alla retorica che spinge per l'autorevolezza oracolare delle procedure interpretative algoritmiche

## Interrompere la connessione tra astrazione ed estrazione

- La capacità astrattiva di questi sistemi è messa nelle piattaforme a servizio di una logica estrattiva, espropriativa e appropriativa
- Dobbiamo trovare metodi che scindano il percorso dell'astrazione da quello dell'espropriazione e appropriazione, costruendo un sistema di sapere basato sulla condivisione e la collaborazione e non sul potere dei forti di mettere a valore risorse comuni
- L'importanza delle situazioni locali, situate e plurali non può essere sottovalutata ai fini di preservare la variazione e le differenze

# Tutelare il pluralismo della molteplicità della differenza

- La molteplicità e il pluralismo, la lotta alla standardizzazione, la tutela della variabilità sono centrali
- I sistemi politici e giuridici dovrebbero battersi per istituire un nuovo patto sociale che definisca limiti, confini e applicazione della delega decisionale e definire le sfere in cui esercitare questa delega solo dove offrano davvero un vantaggio per la collettività



# Il diritto è in prima linea in questo scontro per la regolazione della società

- Le tecnologie propongono un sistema di regolazione dei rapporti sociali e con l'ambiente
- Questa regolazione è il frutto dei risultati ottenuti in pochi laboratori di ricerca delle grandi aziende tecnologiche
- Perché siano davvero utili alla collettività seguendo il suggerimento di Acemoglu e Johnson abbiamo bisogno che la politica e il diritto negozino le condizioni per cui questi strumenti possano davvero essere a beneficio della collettività e non solo di coloro che li hanno costruiti

# Alcune fonti bibliografiche

- Berry D., Fagerjord A. (2017) *Digital Humanities*, Polity, London.
- Chun, Wendy Hui Kyong: Queering Homophily. In: Clemens Apprich, Wendy Hui Kyong Chun, Florian Cramer u.a. (Hg.): Pattern Discrimination. Lüneburg: meson press 2018, S. 59–97. DOI: <https://doi.org/10.25969/mediarep/12350>.
- Chun, W. H. K. (2021) Discriminating data: Correlation, neighborhoods, and the new politics of recognition. MIT press, Cambridge (Mass.).
- Daston L. (2022) *Rules*, Princeton University Press.
- Fiormonte D. , Numerico T. (2015) *The Digital Humanist*, Punctum Books, New York
- Gigerenzer G. (2023) How to stay smart in a smart world, Penguin Publishing Group, London
- Hildebrandt, M. (2020). *Law for computer scientists and other folk*. Oxford University Press.
- Hui, Y. (2021). *Art and cosmotronics*. U of Minnesota Press.
- Macmillan, F. (2020). *Intellectual and cultural property: between market and community*. Routledge.
- Malabou, C. (2022). *Plasticity: The Promise of Explosion*. Edinburgh University Press.
- Negroponte, N. (1995). *Being Digital*. Vintage Books.
- Numerico, T. (2019). Politics and Epistemology of Big Data: A Critical Assessment. *On the Cognitive, Ethical, and Scientific Dimensions of Artificial Intelligence: Themes from IACAP 2016*, 147-166.
- Numerico T. (2021) *Big Data e algoritmi*, Carocci, Roma.
- Rouvroy, A., Berns, T., & Carey-Libbrecht, L. (2013). Algorithmic governmentality and prospects of emancipation. *Réseaux*, 177(1), 163-196.
- Rushkoff, D. (2022). *Survival of the richest: Escape fantasies of the tech billionaires*. WW Norton & Company.
- Stiegler, B. (2018). *Automatic society, volume 1: The future of work*. John Wiley & Sons.
- Supiot, A. (2017). *Homo juridicus: On the anthropological function of the law*. Verso Books.
- Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power: Barack Obama's books of 2019*. Profile books.