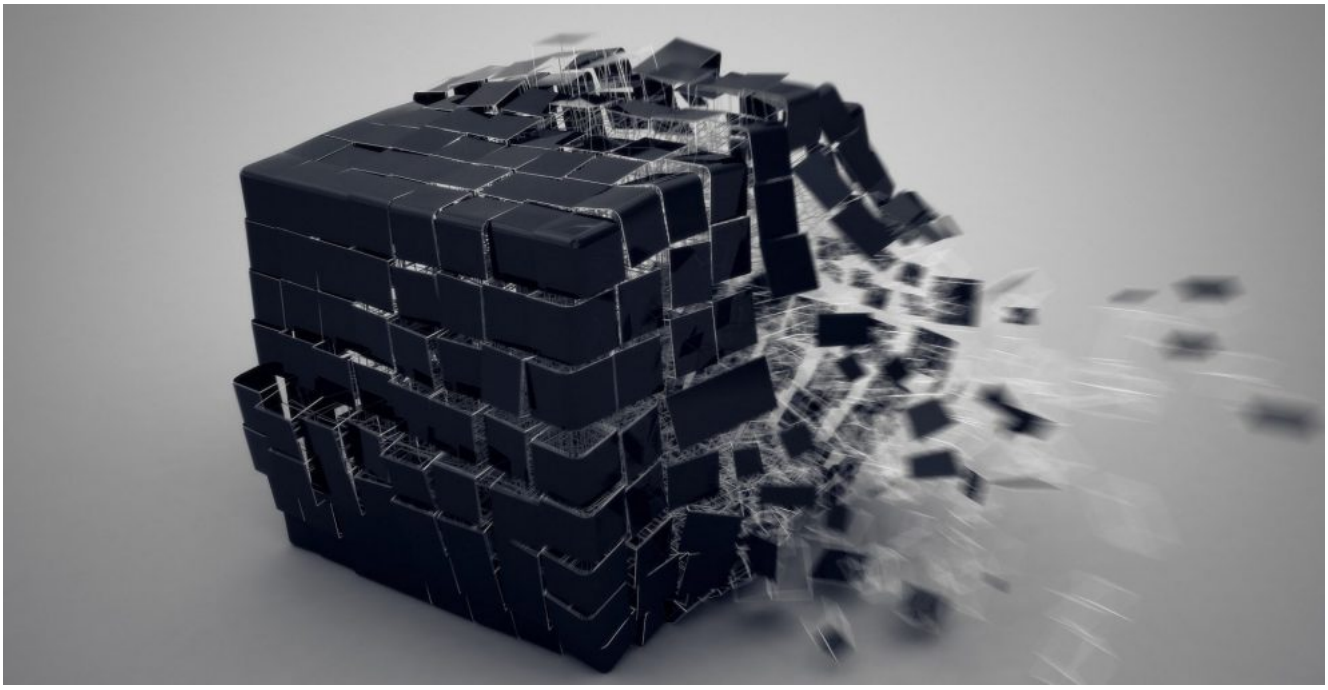
Gli algoritmi e i “mostri” della tecnologia



Ho visto lo sventurato, il miserabile mostro che avevo creato (…) demoniaco cadavere a cui avevo così miseramente dato la vita (Mary Wollstonecraft Godwin)

Nel 2007 l’allora sindaco di Washington, Adrian Fenty, era determinato a risolvere un problema: gli studenti di molte scuole non raggiungevano buoni risultati. Si partì dall’idea che la colpa doveva essere degli insegnanti, che non facevano bene il loro lavoro. Per cui si affidò il compito di valutare gli insegnanti ad un software denominato IMPACT.
L’algoritmo prendeva in considerazione una serie di elementi, tra i quali dei test da somministrare agli studenti e da ripetere nel tempo. Il modello, detto a valore aggiunto (value added model) è fondamentalmente un modello statistico che cerca di distinguere l’impatto causale di un insegnante sull’apprendimento dei suoi studenti da altri elementi, quali abilità degli studenti e fattori extrascolastici.

Gli algoritmi si espandono

I modelli statistici per la valutazione degli insegnanti negli Usa risalgono agli anni ‘70. Man mano che la potenza di calcolo degli elaboratori aumentava, questi sistemi automatizzati si sono diffusi e perfezionati. Ma il primo settore nel quale sono stati applicati gli algoritmi è probabilmente quello finanziario. I sistemi automatizzati hanno rivoluzionato l’intero settore, finendo per sostituire gli esseri umani nelle “decisioni” su vendite e acquisti. Poi, nel 2008, il crollo della borsa fece comprendere che non solo gli algoritmi erano sempre più connessi al mondo degli uomini, ma che potevano anche alimentare ed amplificare i problemi della società.

Negli anni seguenti, come se nulla fosse accaduto, sorsero algoritmi sempre più sofisticati, intelligenze artificiali che imparavano da sole (machine learning) e prendevano decisioni in totale autonomia. Il loro uso si è progressivamente espanso a tutti i settori della società. La statistica e la matematica iniziarono ad occuparsi attivamente di comportamenti, desideri, movimenti, acquisti, crimine e terrorismo.

La presenza di algoritmi nelle nostre vite ormai è pervasiva. Un algoritmo può essere usato (Nicholas Diakopoulos, Algorithmic accountability reporting: on the investigation of black boxes) a fini di prioritizzazione, per stabilire come devono essere distribuite le risorse dei servizi. Ad esempio, per le ispezioni agli edifici per verificare se sono adeguati i servizi antincendio, oppure se reggono in caso di terremoto. La polizia americana utilizza spesso algoritmi di prioritizzazione (es. Predpol) per stabilire in quali zone della città deve inviare le pattuglie disponibili.

Un altro utilizzo comune degli algoritmi è a fini di classificazione, ad esempio per distinguere tra contenuti leciti o illeciti online. Un algoritmo di associazione, invece, crea relazione tra elementi, come se fossero hyperlink. Serve a raggruppare elementi, ad esempio i pregiudicati che hanno avuto problemi con droghe. Infine

abbiamo gli algoritmi di filtraggio, che includono o escludono informazioni in base a specifiche regole o criteri. Ovviamente ogni algoritmo può avere anche più scopi.
Un algoritmo di filtraggio può, quindi, essere anche un algoritmo di classificazione. Come ContentID di Youtube che classifica i contenuti presunti illeciti online e li filtra. Oppure come una App di news personalizzate che classifica le notizie, le filtra in base a criteri prefissati dagli utenti (utilizzando i criteri inseriti dal programmatore) e quindi le associa agli utenti, prioritizzandole.
Il passo dalla valutazione alla predizione è stato relativamente breve. Oggi gli algoritmi predicono la nostra attendibilità, il nostro potenziale, il nostro futuro. C'è un algoritmo per stabilire se siamo (saremo?) bravi insegnanti, un altro per prevedere se saremo buoni studenti, e poi lavoratori, mariti o mogli, amanti, criminali, perfino terroristi.
Sarah Wysocki, insegnante della MCFarland Middle School, era unanimemente considerata tra le migliori. La descrivevano entusiasta, creativa, visionaria, flessibile, motivante e incoraggiante. Eppure venne licenziata, insieme ad oltre 200 insegnanti, a causa di un pessimo punteggio calcolato da Impact. I bassi punteggi ottenuti dai suoi studenti delle elementari forse erano dovuti alle problematiche del quartiere popolare, forse alla povertà delle famiglie, forse anche a problemi di salute dei ragazzi. Ma per i programmatori di Impact non erano motivi validi. La fine della sua carriera è stata decisa da un algoritmo.

Black Box

Un algoritmo consta di tre elementi: i dati di input, il processo algoritmo vero e proprio, il risultato, cioè l'output, si tratti di una previsione o un punteggio. Di questi tre elementi molto spesso si conosce solo l'output. L'algoritmo vero e proprio (il codice) non è conoscibile perché è una "proprietà intellettuale", ed è protetta anche in

base alla recente direttiva Trade Secrets dell'Unione europea.
Le nostre vite sono sempre più esposte e trasparenti, in base a leggi introdotte dai nostri governi e logiche di mercato. Il regime di "sorveglianza" al quale siamo sottoposti pervasivamente è giustificato da ragioni di "sicurezza", o semplicemente per motivi di "efficienza economica" o per "semplificare la navigazione online".
Nel contempo i metodi operativi delle aziende e delle stesse istituzioni sono sempre più opachi, gli algoritmi non sono trasparenti ma vere e proprie black box (Frank Pasquale, The black box society). La segretezza degli algoritmi è tale da impedirci di comprenderne le logiche e quindi di distinguerne il buon funzionamento dall'abuso. Eppure, sono gli algoritmi che ormai dettano le regole di tantissime attività umane, come ad esempio nella selezione del personale.
Il problema è che gli algoritmi, prima di prendere decisioni autonomamente, devono essere addestrati. Il training prevede che siano forniti al sistema degli input, opportunamente identificati (taggati), in modo che l'algoritmo impari cosa riconoscere.
Nel 2009 (HP Computers are racist) un software di rilevamento facciale non riesce a riconoscere una persona di colore. Secondo l'azienda non era un problema "razziale", ma una questione di insufficiente illuminazione dello sfondo che determinava problemi di contrasto. Il software aveva difficoltà a "vedere" persone di pelle scura. In realtà un software non "vede", ma umanizzandolo l'azienda nascondeva il vero problema, e cioè che il software di fatto determinava, ovviamente non intenzionalmente, uno svantaggio per le persone di pelle scura. In sostanza non era "neutrale". Quello che per l'azienda era un problema meramente "tecnico", poteva però avere delle implicazioni nella vita reale sotto forma di discriminazione sociale. La causa probabilmente stava nel fatto che nel laboratorio la maggior parte dei dipendenti erano bianchi.
Altro caso interessante riguarda un sistema algoritmico che produceva risultati "sessisti", associando alle donne immagini

di cucina e così via. Analizzando gli input forniti alla macchina, si vide che due collezioni di immagini, tra cui una supportata da Microsoft e Facebook, presentavano una distorsione di genere nella raffigurazione di attività come la cucina e lo sport: mentre le immagini di cucina, shopping e lavaggio erano associate a donne, quelle di sport erano legate ad uomini. Il software di apprendimento automatico in fondo non faceva altro che il suo lavoro: apprendeva.
L'esperimento più interessante è quello del chatbot Tay della Microsoft. Doveva essere un esperimento intelligente di apprendimento automatico, imparando direttamente dalle persone con le quali interagiva su Twitter. L'idea era che diventasse indistinguibile dai millennial, ma ci sono volute meno di 24 ore perché Tay si trasformasse in un razzista xenofobo. Secondo alcuni, in realtà, l'esperimento è stato un successo. Il bot, preda dei troll, ha imparato fin troppo bene dagli utenti. La debacle è l'esempio di come gli esseri umani possono corrompere la tecnologia.
Gli algoritmi sono costruiti per approssimare il mondo in modo da soddisfare gli scopi del loro architetto e incorporano una serie di presupposti su come funziona e su come dovrebbe funzionare il mondo. Gli algoritmi, quindi, possono riflettere i pregiudizi dei programmatori incorporati nel codice, nel momento in cui interpretano e leggono i dati. Perché un programmatore generalmente non riconosce come pregiudizievoli i suoi criteri e quindi può inavvertitamente (ma anche volontariamente, volendo) introdurre dei criteri non neutrali. Il processo decisionale dipendente da algoritmi finirà per replicare i pregiudizi strutturali su vasta scala, ampliandoli.
Un software di screening valuta i candidati in base a criteri che vengono inseriti nel codice, criteri spesso soggettivi, come il nome. Se un nome "suona" da bianco piuttosto che da nero, da italiano piuttosto che straniero, da settentrionale piuttosto che meridionale, da uomo piuttosto che donna (Amazon e l'intelligenza artificiale sessista: non assumeva donne). Perché un algoritmo sia davvero "neutrale" occorre che i dati

di training siano neutrali anch’essi, che i criteri di selezione inseriti nel codice siano neutrali.
Le forme di discriminazione ipotizzabili sono molte. Alcune polizze assicurative potrebbero essere limitate in base al luogo in cui si vive, le migliori condizioni per le carte di credito potrebbero essere offerte solo a determinate persone, un negozio online potrebbe mostrare prezzi diversi per gli stessi prodotti in base alle caratteristiche individuali, un prestito potrebbe essere rifiutato a chi ha amici poveri sui social, l’applicazione di misure di prevenzione potrebbe avvenire in base ai vicini di casa.

I bias degli algoritmi

Sui bias (pregiudizi) degli algoritmi, Carlo Blengino fa una interessante riflessione su Il Post, che conclude con una serie di domande:

> *Può una macchina, un algoritmo evoluto, aiutare, correggere e fin anche sostituire il giudizio umano? Ma soprattutto, i nostri inevitabili bias cognitivi saranno amplificati o troveranno invece mitigazione, se non soluzione, nell’uso di sistemi esperti artificiali?.*

Gli esseri umani, è noto, sono intrisi di pregiudizi, è il meccanismo di funzionamento del cervello che alimenta tali bias. Ciò che differenzia un essere umano da una scimmia è il meccanismo di compressione cognitiva col quale analizziamo le informazioni in blocchi piuttosto che singolarmente, che è alla base della creazione delle procedure complesse e automatizzate che ci governano (abitudini). In tal modo il cervello può concentrarsi su più cose contemporaneamente rendendoci molto più efficienti. Di contro, molte informazioni non sono più analizzate dalla coscienza critica.
La capacità di riconoscimento dei pattern e la capacità combinatoria, ci hanno permesso di essere la specie animale più avanzata, ma porta anche un lato oscuro. Quel processo

comporta la possibilità di saltare a conclusioni sbagliate per il semplice motivo che le informazioni iniziali non vengono sempre sottoposte ad analisi. Siamo così desiderosi di cercare dei modelli nel mondo che ci circonda, e così soddisfatti quando li abbiamo trovati, che non effettuiamo controlli sufficienti sulle nostre intuizioni apparenti.
E gli inevitabili bias degli essere umani possono essere introdotti negli algoritmi nella fase di programmazione o di addestramento. Con l'amplificazione dovuta alla scala, enormemente più vasta, di applicazione dell'algoritmo.
Quindi, un algoritmo aiuterà l'uomo a liberarsi dei suoi bias? Oppure sarà l'uomo a risolvere i bias degli algoritmi con maggiori controlli sulla neutralità dei codici?

Una "soffiata" digitale

Una fonte confidenziale chiama la Polizia, comunicando che in un certo luogo vi sono due persone intente a spacciare droghe. La Polizia si reca sul posto. Dopo solo due minuti di osservazione, un evento imprevisto (sopraggiunge qualcuno che potrebbe vederli?) decidono di intervenire. Hanno visto due persone, sono lì ferme. Nonostante vi fossero degli elementi per ritenere che una delle due persone fosse solo un acquirente, entrambe vengono arrestate per spaccio.
La "fonte confidenziale", infatti, parlava di "due" spacciatori. Giunti sul posto, i poliziotti vedono due persone, e, poiché la fonte confidenziale era stata "riscontrata" (cioè trovata attendibile) più volte, "leggono" gli elementi che si presentano ai loro occhi in base alla "soffiata".
Secondo il "principio di coerenza", infatti, si tende a scegliere o a creare l'interpretazione dei fatti più coerente con i dati disponibili. Il principio del "minimo impegno" fa sì che il soggetto tenda a compiere solo le inferenze minime indispensabili a produrre un'interpretazione coerente, per cui tra due spiegazioni ugualmente coerenti sceglierebbe quella che richiede il minor numero di supposizioni. Elementi tali da

far ritenere che una delle due persone è solo un acquirente non vengono considerati, oppure sono ridotti a coerenza con le informazioni fornite dalla “fonte confidenziale”.

Autorità e controllo

Il processo di interpretazione dei dati è un processo complesso che può portare ad errori, specialmente in casi dubbi, nei quali i dati possono essere interpretati in modi diversi. Il modo di rapportarsi dell’“interprete” ai dati è fondamentale, perché un pregiudizio dell’interprete può portare a risultati sbagliati.
Se il risultato, cioè l’interpretazione dei dati, viene da un algoritmo, l’output potrebbe essere discriminatorio, se l’algoritmo non è neutrale. Ma il problema primario è che per verificare se l’algoritmo è neutrale o meno è innanzitutto necessario essere coscienti del fatto che l’algoritmo possa sbagliare. Ma, chi utilizza i risultati dell’algoritmo spesso non è a conoscenza del codice di programmazione né dei dati di input (addestramento). Se l’algoritmo è stato utilizzato varie volte e in molti casi è risultato attendibile (è stato “riscontrato”) si tende a credere che lo sia sempre. Il poliziotto che decide di inviare una pattuglia in una determinata zona in base all’algoritmo difficilmente penserà che l’algoritmo si può sbagliare.
Una black box in fondo, non è altro che una soffiata digitale. Se riscontrata più volte la si ritiene affidabile. Diventa incontestabile, una sorta di “autorità” che non è possibile mettere in discussione.
Come ha evidenziato Milgram con i suoi esperimenti, le persone sono estremamente disponibili ad assoggettarsi a un’autorità. E l’intera società è strutturata in modo da condizionare, attraverso le istituzioni (scuola, famiglia, lavoro), l’individuo a sottostare alla regola dell’obbedienza all’autorità. Ma oggi l’autorità è sempre più espressa in forma di algoritmo.
È l’impenetrabile black box, difesa dalle leggi moderne, a

garantire l’esercizio del potere nelle società moderne. La strutturazione del sistema in black box rimuove le responsabilità dei singoli rendendo difficile nell’ambito delle gerarchie e delle competenze stabilire chi deve pagare per un eventuale errore. L’obbedienza all’autorità fa il resto.
Sulla segretezza degli algoritmi si fonda l’elemento essenziale della “sorveglianza”, cioè il “controllo”. L’individuo viene frammentato in dati e ricomposto a creare quella che non è la banale transcodifica della sua vita in informazioni gestibili da un elaboratore, quanto piuttosto l’esercizio di un vero e proprio potere, una forma di controllo (John Cheney-Lippold, We are data). Essere governati non è altro che essere controllati costantemente, osservati, indottrinati, valutati e censurati. Se un medico per svolgere il suo lavoro necessita di un badge che gli consente l’ingresso nel reparto, vuol dire che c’è un algoritmo che stabilisce quando e se quel medico può essere un medico. Basta revocargli i permessi perché non lo sia più. Così per uno studente, e per un insegnante. È un algoritmo a decidere chi o cosa siamo. Dipende dall’algoritmo se noi abbiamo dei diritti, perché in fondo dipende dall’algoritmo se siamo identificati come soggetti che hanno diritti. Questa è la “datificazione”.
Oggi una persona è datificata non tanto in base a ciò che è, ma per come è vista dal programmatore. L’esempio più classico: il “terrorista”. Un terrorista oggi non è più una persona, quanto piuttosto un modello (type) ricavato dalla datificazione di una serie di comportamenti tipici di soggetti ritenuti terroristi (signature, o firma). I dati (comportamenti) da estrarre sono selezionati dai programmatori del modello, quindi alla fine il modello di terrorista non descrive affatto un terrorista quanto piuttosto come un terrorista è visto dal programmatore (lo schema di un terrorista).
Ad esempio, se delle persone sono viste con armi in prossimità di un luogo frequentato da terroristi, oppure se sono viste comunicare con terroristi, o dare ordini a terroristi. È anche

possibile che l’algoritmo confonda e riunisca in un unico individuo il comportamento di più soggetti che abitano lo stesso appartamento. La costruzione di un modello ideale nel quale far rientrare una persona fisica non si basa sulla realtà quanto piuttosto è un’approssimazione di un fenomeno dinamico, la traduzione in una quantità di numeri trattabili da un software. In tal senso la costruzione di un modello di terrorista (ma anche di altre tipologie) non è tanto l’estrazione di dati dalla realtà (raw data), quanto piuttosto la costruzione di dati a partire dall’osservazione della realtà (cooked data). E, come tale, è soggetta a molteplici errori e fenomeni discriminatori.
La classificazione in base ad algoritmi non riscontra l’individuo effettivo come si estrinseca nella vita reale, quanto piuttosto un suo “doppio” virtuale, che può essere “maschio” ma anche “femmina”, “famoso” o “non famoso”, “giovane” o “vecchio”. La categorizzazione di “uomo” non è un uomo bensì la datificazione dello schema di attività che generalmente pone in essere un “uomo”. Per un algoritmo una donna che si “comporta” da uomo, legge riviste da “uomo”, fa acquisti da “uomo”, è un “uomo”, con un disallineamento tra la vita reale e la vita basata sui dati (vita online o virtuale). E la classificazione è del tutto indipendente dalla volontà dell’individuo. Nella sua visione più ottimistica, quella “Silicon Valley oriented”, i problemi degli algoritmi sono facilmente risolvibili. Come? Accumulando più dati. Sempre più dati.
Nella sua visione più ottimistica, quella “Silicon Valley oriented”, i problemi degli algoritmi sono facilmente risolvibili. Come? Accumulando più dati. Sempre più dati.

Cosa è normale?

I modelli algoritmici utilizzano i dati dei comportamenti passati per prevedere il futuro. In tal modo si realizza una serie di schemi comportamentali (profili) che servono a classificare le persone fisiche. I cittadini vengono

categorizzati assegnando loro un profilo di rischio. L'inclusione in categorie dipende e rafforza l'idea che certi comportamenti siano la "norma".
Porre in essere comportamenti che si discostano da tale "normalità", comporta automaticamente l'inclusione in categorie a rischio, finendo per indicare anche l'avvio di un percorso delinquenziale o terroristico. E il discrimine tra attività sovversiva e semplice attività di protesta è molto sottile agli occhi dello stato.
Così, il sistema finisce per discriminare tutti coloro che non si comportano a "norma", favorendo il conformismo. Tende a replicare possibili disuguaglianze (si pensi ai musulmani, ritenuti spesso non a norma in paesi occidentali), oltre a creare difficoltà nell'affrontare eventi del tutto nuovi od inaspettati.

Monstrum

> *"Ho visto lo sventurato, il miserabile mostro che avevo creato (…) demoniaco cadavere a cui avevo così miseramente dato la vita".*

Così Mary Wollstonecraft Godwin introduce la "creatura" nel libro "Frankenstein o del Prometeo moderno". Il suo artefice crea il mostro perché lo riverisca come un dio. Ma il mostro (dal latino "monere", avvertire) è anche una critica spietata, non tanto alla scienza e alla tecnologia, ma a ciò che con essa viene fornita (Anne k. Mellor, Mary Shelley: Her Life, Her Fiction, Her Monsters).
Il mostro, in realtà, è un'inevitabilità storica, una necessità per stabilire i confini tra la società e ciò che la minaccia, ciò che deve rimanerne fuori, ciò che le è alieno. Nel romanzo al di fuori dei confini della città (della società) stanno i demoni, fatti di parti di cadavere, di frammenti e pezzi presi qua e là e ricomposti in un tutt'uno (come i dati raccolti e ricomposti dagli algoritmi), per creare ciò su cui devono essere focalizzate le paure e i

timori della società intera (così come oggi si fa con gli islamici, i migranti, i terroristi, ecc…). Il mostro rappresenta i mali della società, e una volta cacciato, ci assicurano, i problemi saranno tutti risolti. Ma la scomoda verità è che il mostro non è affatto a noi estraneo o diverso o separato, bensì è la tecnologia (la creazione) a separarlo dal resto della società. Al punto che senza di essa il mostro non ha più motivo di esistere (nel romanzo senza il suo creatore il mostro si uccide).

"Cosa è normale in un mondo retto dagli algoritmi?", si chiede John Cheney-Lippold in We are data. Nella società capitalista, in quella moderna retta da algoritmi, il mostro è tutto ciò che non rientra nella normalità, che non rientra nelle categorie accettate nella società, tutto ciò che è diverso, coloro che non si assoggettano all'autorità costituita, quelli che sfuggono alla black box. Creare il mostro ci consente di definire i confini di cosa è normale e cosa no. Nella società moderna, dove regna l'imperativo tecnologico, mirabilmente riassunto nelle parole del nipote del professore in Frankenstein Junior di Mel Brooks ("si può fare!"), per cui tutto ciò che è potenzialmente fattibile va fatto senza tenere conto delle implicazioni etiche, il compito di creare il "mostro" è svolto sempre più dagli algoritmi.

E Sarah Wysocky, l'insegnante licenziata a causa di un pessimo punteggio calcolato da un algoritmo? L'algoritmo che la licenziò fu contestato. Un modello statistico dovrebbe essere basato su grandi quantità di dati, ma nel caso specifico il tutto si riduceva ai test di una trentina di studenti. In particolare si evidenziò che il modello non funzionava correttamente perché la distribuzione degli studenti non era casuale. Piuttosto dipendeva da fattori quali le richieste specifiche dei genitori, dalle specializzazioni degli insegnanti, e così via. Un insegnante ritenuto particolarmente portato per bambini con problemi di concentrazione otteneva inevitabilmente punteggi più bassi. Così la selezione in base all'algoritmo finiva per creare forti disincentivi per gli insegnanti ad occuparsi degli studenti più problematici.

Ciò non portò ad alcun ripensamento del sistema, che anzi fu giustificato anche di fronte a evidenze contrarie. Buoni punteggi ottenuti dagli studenti di insegnanti con basso punteggio erano ritenuti confermativi della valutazione del sistema perché in tali casi l'insegnante, a rischio di licenziamento, era portato ad alterare i punteggi degli studenti. Infine, si concluse, il modello statistico può anche sbagliare in casi singoli, ma l'efficacia si vede nel complesso. In sostanza non era il sistema che si adattava alla realtà, ma in alcuni casi si adattava la realtà per giustificare il sistema.
Sarah Wysocki fu licenziata. Ma, essendo comunque un'ottima insegnante venne immediatamente assunta da un'altra scuola. In un quartiere per ricchi, che non utilizzava quel modello di valutazione degli insegnanti.