DS-GA 3001.009: Responsible Data Science

Interpretability

Prof. Julia Stoyanovich
Center for Data Science
Computer Science and Engineering at Tandon

@stoyanoj

http://stoyanovich.org/
https://dataresponsibly.github.io/

Transparency themes

- **Explaining black-box models**
- **Online ad targeting**
- **Interpretability**

Algorithmic rankers

https://freedom-to-tinker.com/2016/08/05/revealing-algorithmic-rankers/

Input: database of items (individuals, colleges, cars, …)

Score-based ranker: computes the score of each item using a **known formula**, e.g., monotone aggregation, then sorts items on score

Output: permutation of the items (complete or top-k)

Do we have transparency?

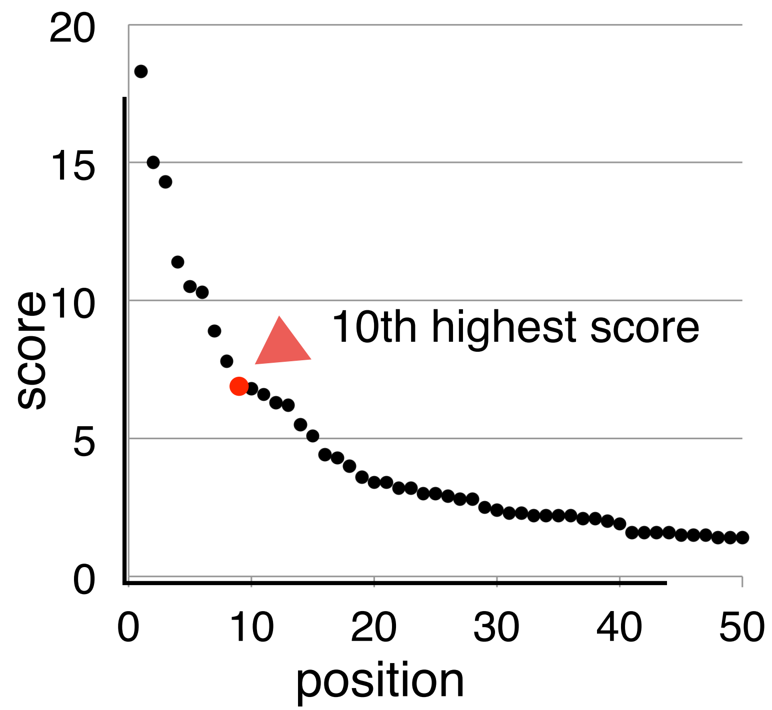
We have syntactic transparency, but lack interpretability!

Opacity in algorithmic rankers

https://freedom-to-tinker.com/2016/08/05/revealing-algorithmic-rankers/

Reason 1: The scoring formula alone does not indicate the relative rank of an item.

Scores are absolute, rankings are relative. Is 5 a good score? What about 10? 15?



Opacity in algorithmic rankers

https://freedom-to-tinker.com/2016/08/05/revealing-algorithmic-rankers/

Reason 2: A ranking may be unstable if there are tied or nearly-tied items.

Rank	Institution	Average Count	Faculty
1	▶ Carnegie Mellon University	18.4	123
2	▶ Massachusetts Institute of Technology	15.6	64
3	▶ Stanford University	14.8	56
4	▶ University of California - Berkeley	11.5	50
5	▶ University of Illinois at Urbana-Champaign	10.6	56
6	▶ University of Washington	10.3	50
7	▶ Georgia Institute of Technology	8.9	81
8	▶ University of California - San Diego	8	51
9	▶ Cornell University	7	45
10	▶ University of Michigan	6.8	63
11	▶ University of Texas - Austin	6.6	43
12	▶ University of Massachusetts - Amherst	6.4	47

Opacity in algorithmic rankers

https://freedom-to-tinker.com/2016/08/05/revealing-algorithmic-rankers/

Reason 3: A ranking methodology may be unstable: small changes in weights can trigger significant re-shuffling.

THE NEW YORKER

DEPT. OF EDUCATION FEBRUARY 14 & 21, 2011 ISSUE

THE ORDER OF THINGS

What college rankings really tell us.

By Malcolm Gladwell

1. Chevrolet Corvette 205
2. Lotus Evora 195
3. Porsche Cayman 195

1. Lotus Evora 205
2. Porsche Cayman 198
3. Chevrolet Corvette 192

1. Porsche Cayman 193
2. Chevrolet Corvette 186
3. Lotus Evora 182

Opacity in algorithmic rankers

https://freedom-to-tinker.com/2016/08/05/revealing-algorithmic-rankers/

Reason 4: The **weight of an attribute** in the scoring formula **does not determine its impact** on the outcome.

Rank	Name	Avg Count	Faculty	Pubs	GRE
1	CMU	18.3	122	2	791
2	MIT	15	64	3	772
3	Stanford	14.3	55	5	800
4	UC Berkeley	11.4	50	3	789
5	UIUC	10.5	55	3	772
6	UW	10.3	50	2	796
		….			
39	U Chicago	2	28	2	779
40	UC Irvine	1.9	28	2	787
41	BU	1.6	15	2	783
41	U Colorado Boulder	1.6	32	1	761
41	UNC Chapel Hill	1.6	22	2	794
41	Dartmouth	1.6	18	2	794

Given a score function:

$$0.2 * faculty +$$
$$0.3 * avg\ cnt +$$
$$0.5 * gre$$

Rankings are not benign!

THE NEW YORKER

DEPT. OF EDUCATION FEBRUARY 14 & 21, 2011 ISSUE

THE ORDER OF THINGS

What college rankings really tell us.

By Malcolm Gladwell

Rankings are not benign. They enshrine very particular ideologies, and, at a time when American higher education is facing a crisis of accessibility and affordability, we have adopted **a de-facto standard of college quality** that is uninterested in both of those factors. And why? Because a group of magazine analysts in an office building in Washington, D.C., decided twenty years ago to **value selectivity over efficacy**, to **use proxies** that scarcely relate to what they're meant to be proxies for, and to **pretend that they can compare** a large, diverse, low-cost land-grant university in rural Pennsylvania with a small, expensive, private Jewish university on two campuses in Manhattan.

Harms of opacity

https://freedom-to-tinker.com/2016/08/05/revealing-algorithmic-rankers/

1. **Due process / fairness.** The subjects of the ranking cannot have confidence that their ranking is meaningful or correct, or that they have been treated like similarly situated subjects - **procedural regularity**

2. **Hidden normative commitments.** What factors does the vendor encode in the scoring ranking process (syntactically)? What are the **actual** effects of the scoring / ranking process? Is it stable? How was it validated?

Harms of opacity

https://freedom-to-tinker.com/2016/08/05/revealing-algorithmic-rankers/

3. **Interpretability.** Especially where ranking algorithms are performing a public function, **political legitimacy** requires that the public be able to interpret algorithmic outcomes in a meaningful way. Avoid *algocracy*: the rule by incontestable algorithms.

4. **Meta-methodological assessment.** Is *a* ranking / *this* ranking appropriate here? Can we use a process if it cannot be explained? Probably yes, for recommending movies. Probably not for college admissions.

“Nutritional labels” for data and models

[K. Yang, J. Stoyanovich, A. Asudeh, B. Howe, HV Jagadish, G. Miklau; *SIGMOD 2018*]

http://demo.dataresponsibly.com/rankingfacts/nutrition_facts/

an (ongoing) attempt
at regulation

New York City Local Law 49

January 11, 2018

Local Law 49 of 2018 in relation to automated decision systems used by agencies

THE NEW YORK CITY COUNCIL
Corey Johnson, Speaker

Sign In

LEGISLATIVE RESEARCH CENTER

Council Home | Legislation | Calendar | City Council | Committees

RSS | Alerts

Details | Reports

File #:	Int 1696-2017 Version: A	Name:	Automated decision systems used by agencies.
Type:	Introduction	Status:	Enacted
		Committee:	Committee on Technology
On agenda:	8/24/2017		
Enactment date:	1/11/2018	Law number:	2018/049
Title:	A Local Law in relation to automated decision systems used by agencies		
Sponsors:	James Vacca, Helen K. Rosenthal, Corey D. Johnson, Rafael Salamanca, Jr., Vincent J. Gentile, Robert E. Cornegy, Jr., Jumaane D. Williams, Ben Kallos, Carlos Menchaca		
Council Member Sponsors:	9		
Summary:	This bill would require the creation of a task force that provides recommendations on how information on agency automated decision systems may be shared with the public and how agencies may address instances where people are harmed by agency automated decision systems.		
Indexes:	Oversight		
Attachments:	1. Summary of Int. No. 1696-A, 2. Summary of Int. No. 1696, 3. Int. No. 1696, 4. August 24, 2017 - Stated Meeting Agenda with Links to Files, 5. Committee Report 10/16/17, 6. Hearing Testimony 10/16/17, 7. Hearing Transcript 10/16/17, 8. Proposed Int. No. 1696-A - 12/12/17, 9. Committee Report 12/7/17, 10. Hearing Transcript 12/7/17, 11. December 11, 2017 - Stated Meeting Agenda with Links to Files, 12. Hearing Transcript - Stated Meeting 12-11-17, 13. Int. No. 1696-A (FINAL), 14. Fiscal Impact Statement, 15. Legislative Documents - Letter to the Mayor, 16. Local Law 49, 17. Minutes of the Stated Meeting - December 11, 2017		

The original draft

August 16, 2017

Int. No. 1696

By Council Member Vacca

A Local Law to amend the administrative code of the city of New York, in relation to automated processing of data for the purposes of targeting services, penalties, or policing to persons

Be it enacted by the Council as follows:

Section 1. Section 23-502 of the administrative code of the city of New York is amended to add a new subdivision g to read as follows:

g. Each agency that uses, for the purposes of targeting services to persons, imposing penalties upon persons or policing, an algorithm or any other method of automated processing system of data shall:

1. Publish on such agency's website, the source code of such system; and

2. Permit a user to (i) submit data into such system for self-testing and (ii) receive the results of having such data processed by such system.

§ 2. This local law takes effect 120 days after it becomes law.

MAJ
LS# 10948
8/16/17 2:13 PM

not what was adopted

How I got involved

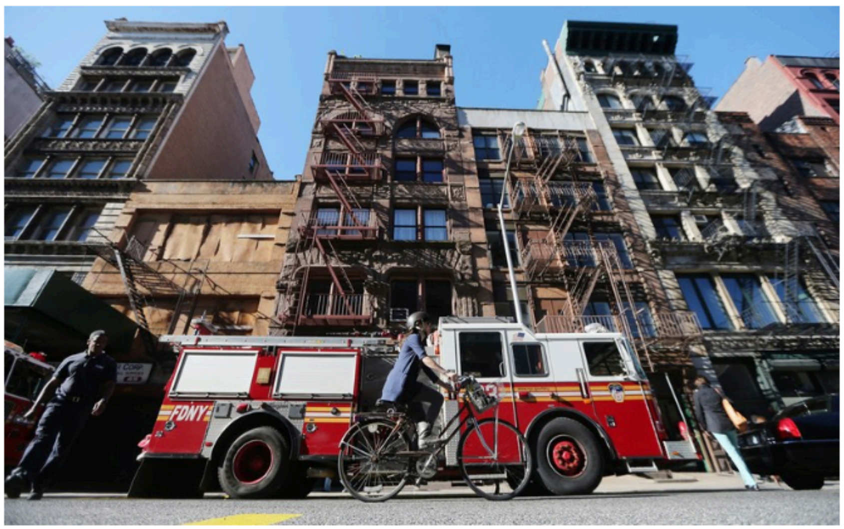
October 16, 2017

THE NEW YORKER

By Julia Powles December 20, 2017

ELEMENTS

NEW YORK CITY'S BOLD, FLAWED ATTEMPT TO MAKE ALGORITHMS ACCOUNTABLE



Automated systems guide the allocation of everything from firehouses to food stamps. So why don't we know more about them?

Photograph by Mario Tama / Getty

https://dataresponsibly.github.io/documents/Stoyanovich_VaccaBill.pdf

Summary of Local Law 49

January 11, 2018

An **Automated Decision System (ADS)** is a "computerized implementation of algorithms, including those derived from machine learning or other data processing or artificial intelligence techniques, which are used to make or assist in making decisions."

Form task force that surveys the current use of ADS in City agencies and develops procedures for:

- requesting and receiving an **explanation** of an algorithmic decision affecting an individual (3(b))
- interrogating ADS for **bias and discrimination** against members of legally-protected groups (3(c) and 3(d))
- allowing the **public** to **assess** how ADS function and are used (3(e)), and archiving ADS together with the data they use (3(f))

The ADS Task Force

May 16, 2018

Visit **alpha.nyc.gov** to help us test out new ideas for NYC's website.

The Official Website of the City of New York — NYC — 简体中文 Translate Text Size

NYC Resources | NYC311 | Office of the Mayor | Events | Connect | Jobs | Search

Mayor | First Lady | News | Officials

SHARE — Email — Print

Mayor de Blasio Announces First-In-Nation Task Force To Examine Automated Decision Systems Used By The City

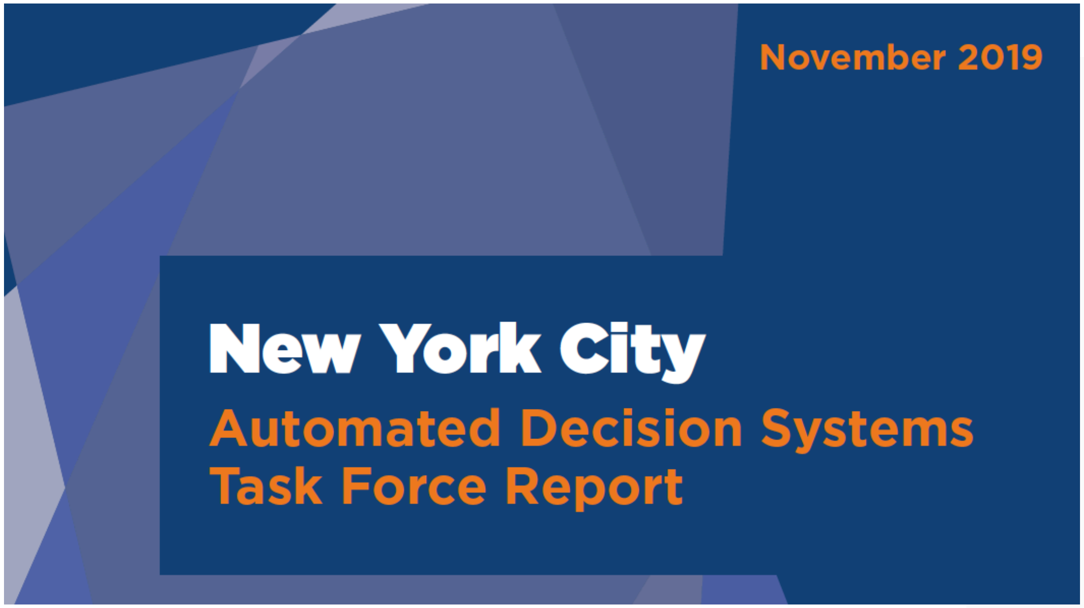
May 16, 2018

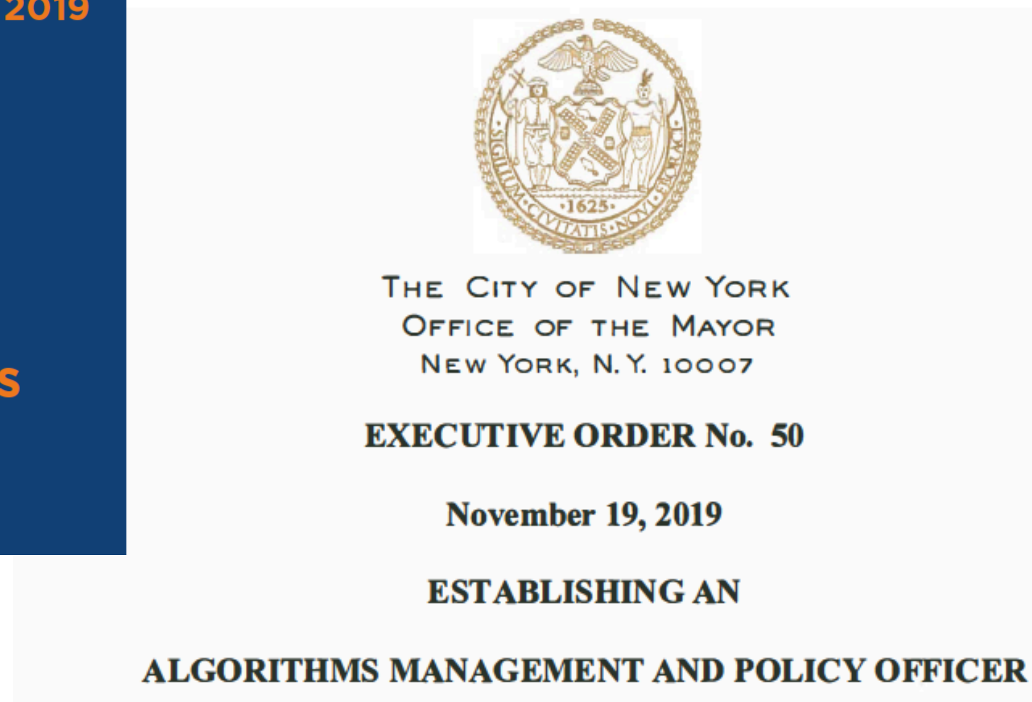
NEW YORK— Today, Mayor de Blasio announced the creation of the Automated Decision Systems Task Force which will explore how New York City uses algorithms. The task force, the first of its kind in the U.S., will work to develop a process for reviewing "automated decision systems," commonly known as algorithms, through the lens of equity, fairness and accountability.

"As data and technology become more central to the work of city government, the algorithms we use to aid decision making must be aligned with our goals and values," said **Mayor de Blasio**. "The establishment of the Automated Decision Systems Task Force is an important first step towards greater transparency and equity in our use of technology."

The outcome (so far)

November 19, 2019





The City of New York
Office of the Mayor
New York, N.Y. 10007

EXECUTIVE ORDER No. 50

November 19, 2019

ESTABLISHING AN

ALGORITHMS MANAGEMENT AND POLICY OFFICER

https://www1.nyc.gov/site/adstaskforce/index.page

https://www1.nyc.gov/assets/adstaskforce/downloads/pdf/ADS-Report-11192019.pdf

https://www1.nyc.gov/assets/home/downloads/pdf/executive-orders/2019/eo-50.pdf

from transparency to interpretability

Point 1

algorithmic transparency is not synonymous with releasing the source code

publishing source code helps, but it is sometimes unnecessary and often insufficient

Point 2

algorithmic transparency requires data transparency

data is used in training, validation, deployment

validity, accuracy, applicability can only be understood in the data context

data transparency is necessary for all ADS, not only for ML-based systems

Point 3

data transparency is not synonymous with making all data public

release data whenever possible;

also release:

data selection, collection and pre-processing methodologies; data provenance and quality information; known sources of bias; privacy-preserving statistical summaries of the data

Point 4

actionable transparency requires interpretability

explain assumptions and effects, not details of operation

engage the public - technical and non-technical

Point 5

transparency by design, not as an afterthought

provision for transparency and interpretability at every stage of the data lifecycle

useful internally during development, for communication and coordination between agencies, and for accountability to the public

interpretability: in the eye of the beholder

What are we explaining?

[J. Stoyanovich, J. Van Bavel, T. West; *NMI 2020*]

process (same for everyone? why is this the process?) vs. outcome

procedural justice aims to ensure that algorithms are perceived as fair and legitimate

data transparency is unique to algorithm-assisted decision-making, relates to the justification dimension of interpretability

To whom are we explaining and why?

[J. Stoyanovich, J. Van Bavel, T. West; *NMI 2020*]

accounting for the needs of different stakeholders

social identity - people trust their in-group members more

moral cognition - is a decision or outcome morally right or wrong?

How do we know that we explained well?

[J. Stoyanovich, J. Van Bavel, T. West; *NMI 2020*]

nutritional labels! :)

… but do they work?