

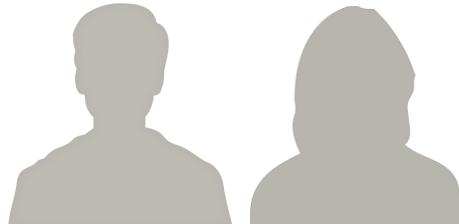
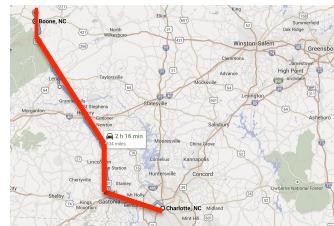
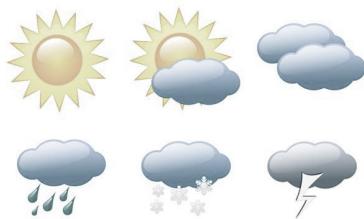
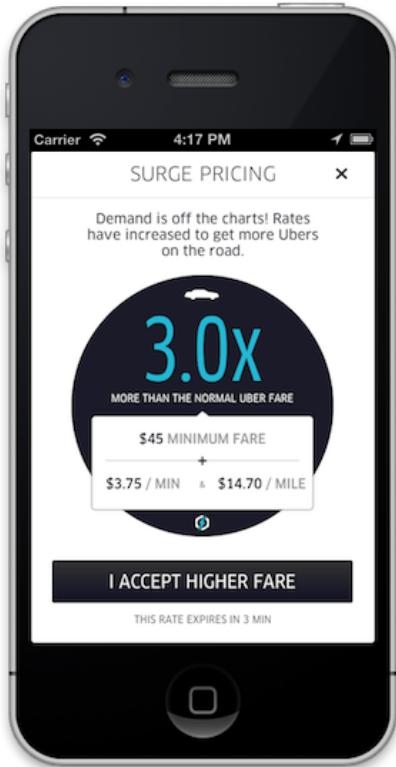
Algorithmic Accountability & Transparency in Journalism

Nicholas Diakopoulos, Ph.D.

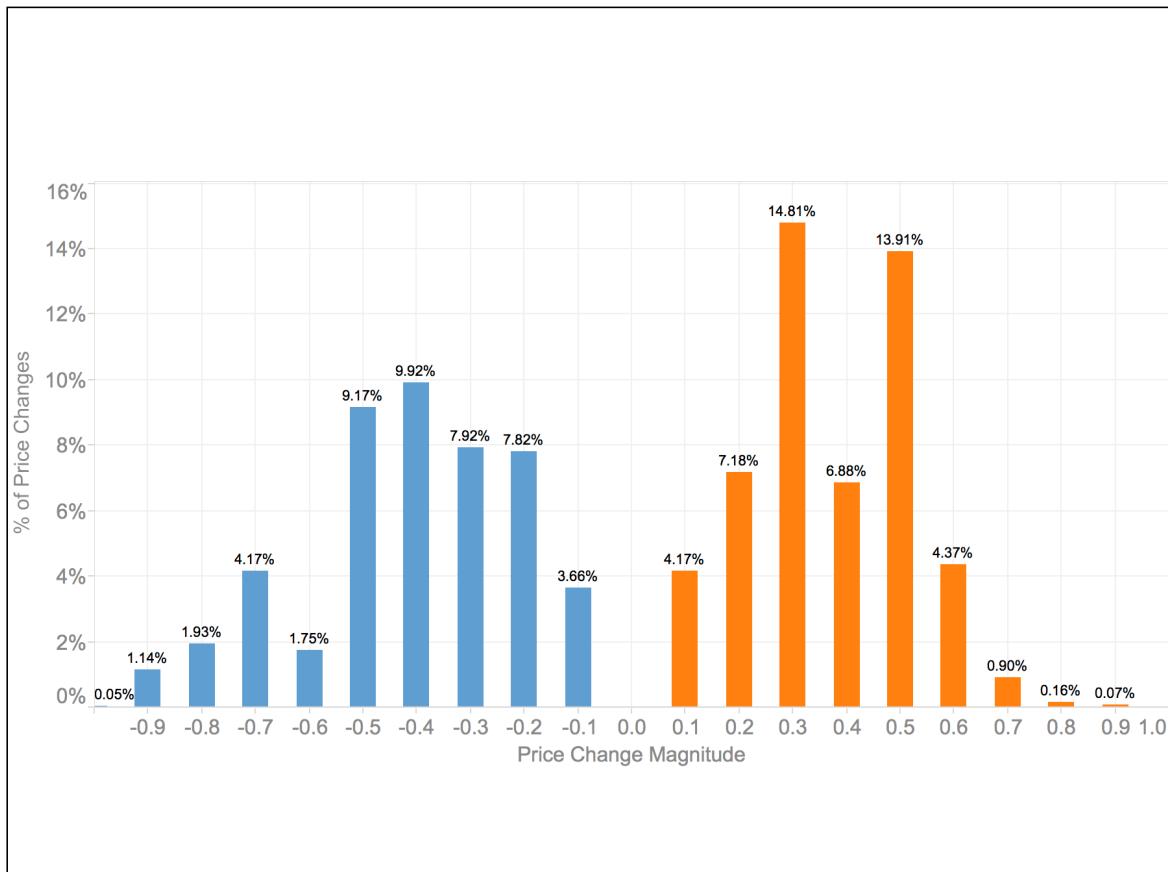
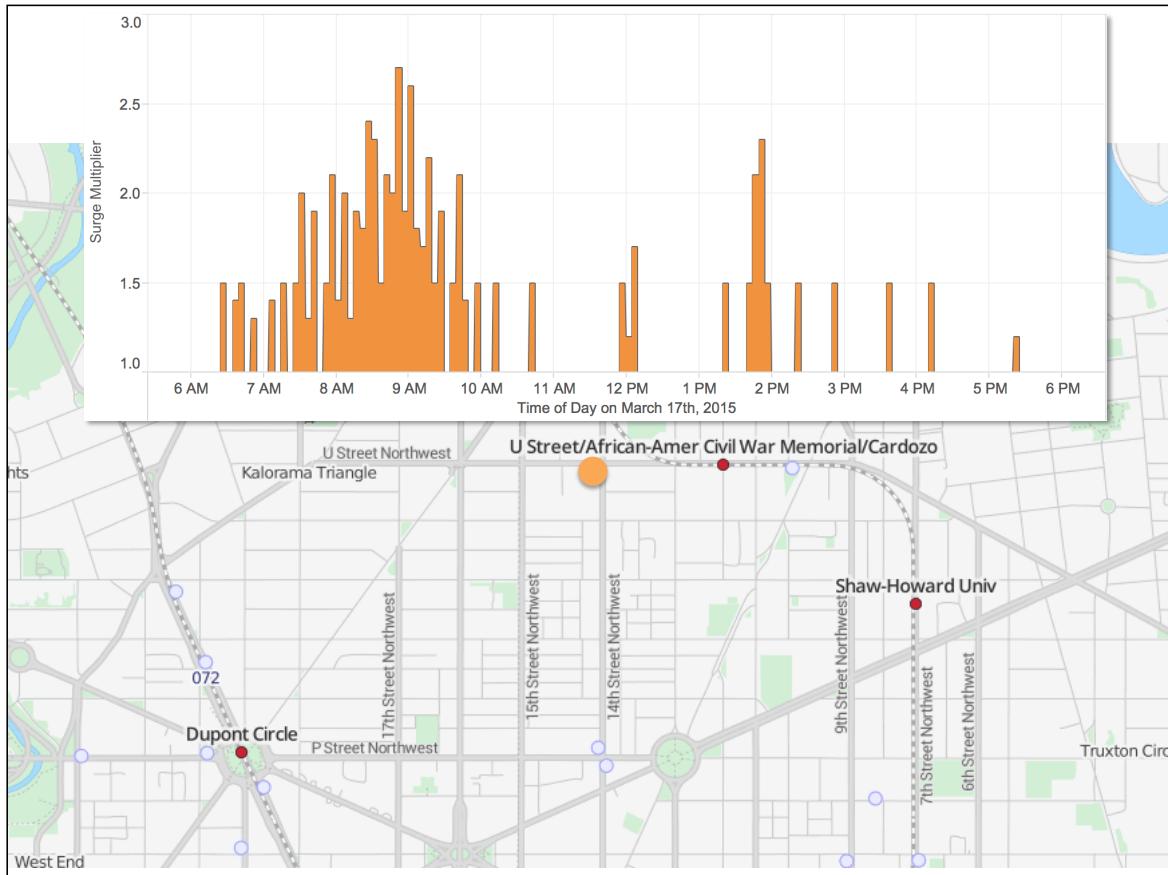
College of Journalism, University of Maryland, College Park
@ndiakopoulos | nickdiakopoulos.com | nad@umd.edu

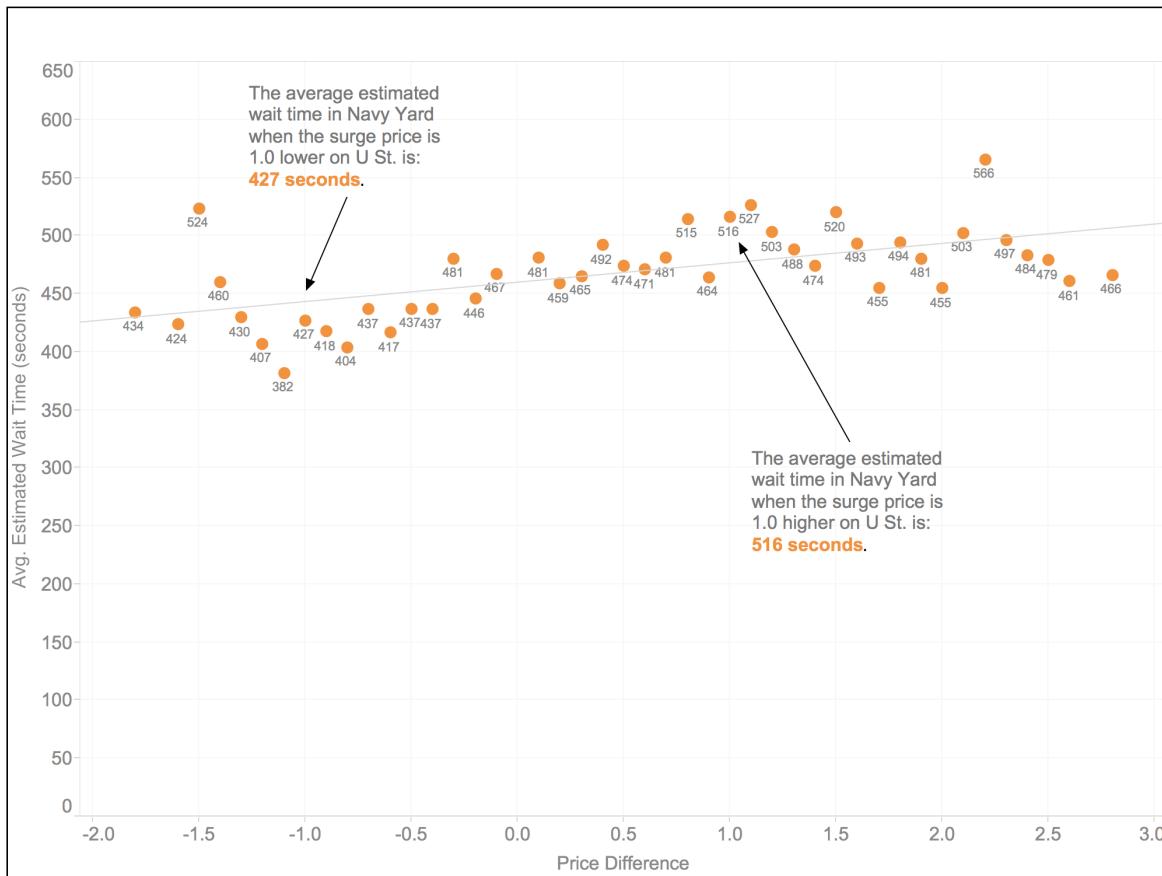
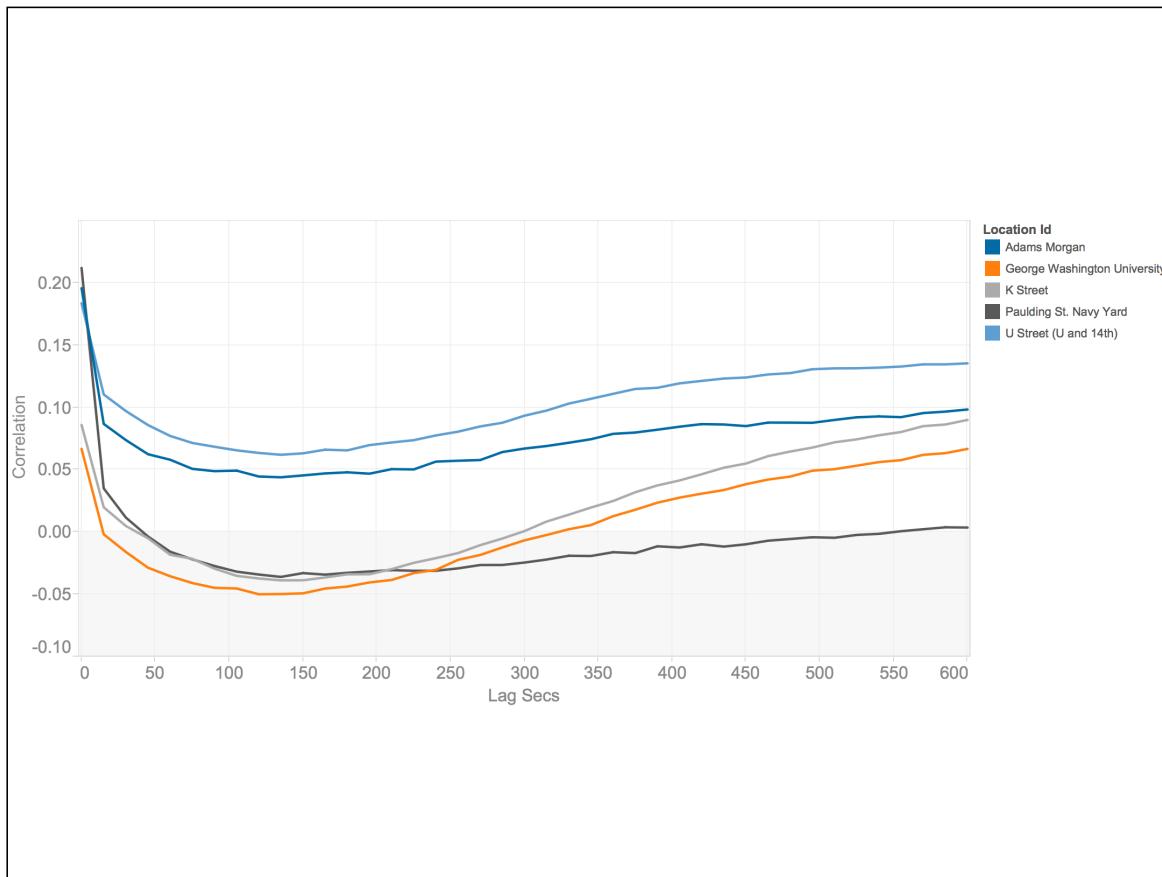


BEFORE YOU CONFIRM YOUR RIDE,
YOU'LL NEED TO ACCEPT THE RATE:



Do surge prices even work? If so, how?





IRS Audits Fraud Detection
Political Messaging Urban Planning
Credit Scores **AirBnb Pricing** Predictive Policing
Yelp Reviews Copyright Infringement Student Matching Net Neutrality
Comment Moderation College Rankings Twitter Trends Advertising
Government Censorship Welfare Management **News Personalization**
Automatically Generated Content Teacher Rankings Romance & Dating
Criminal Sentencing Search Rankings & Completions
Human Resources News Layout
Recommender Systems



"I'd like to meet the algorithm that thought we'd be a good match."

Algorithmic Accountability

How to investigate algorithms that are used in media, government, or industry?

How can we characterize the bias, power, or influence of an algorithm?

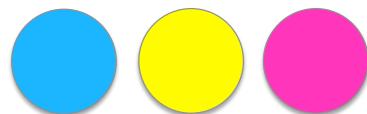
What role might journalists play in holding algorithmic power to account?

N. Diakopoulos. Algorithmic Accountability Reporting: On the Investigation of Black Boxes. Tow Center Report, Columbia University. Feb., 2014.

Prioritization



Classification



Association



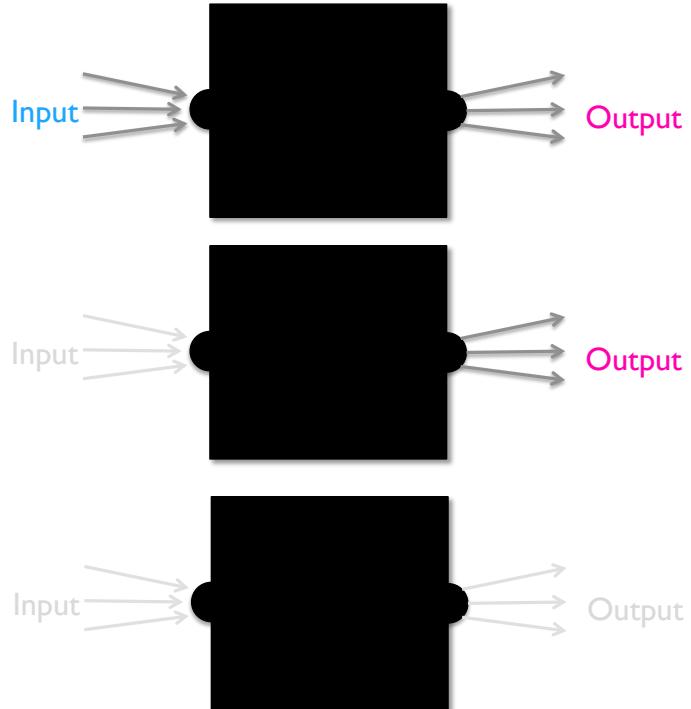
Filtering



Input-Output of an Algorithm



Black-box Scenarios



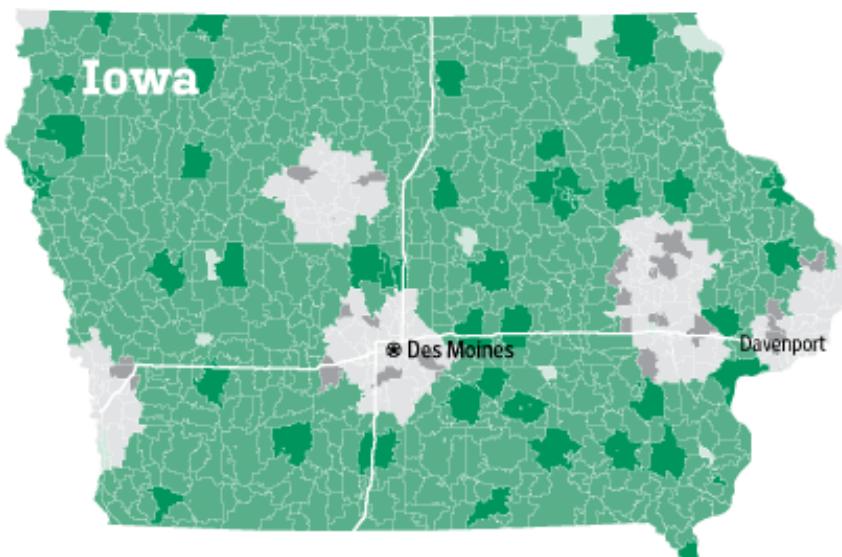
WSJ Price Discrimination

Do different people pay different prices depending on their geography or browser history?



Jennifer Valentino-DeVries, Jeremy Singer-Vine, and Ashkan Soltani. Websites Vary Prices, Deals Based on Users' Information. Wall Street Journal. Dec, 2012.

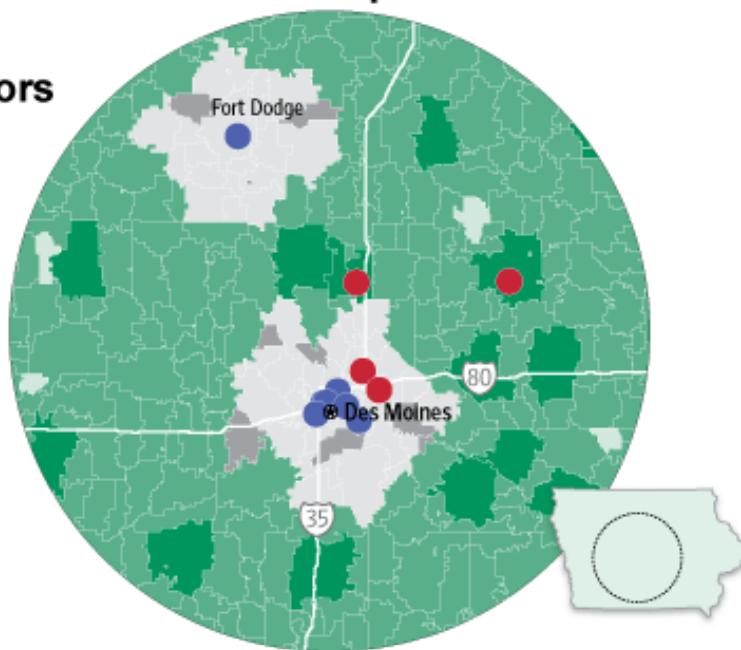
Likelihood of receiving higher prices, by ZIP code



Locations of stores relative to price zones

- Staples
- Competitors

PRICES



Newsworthiness

What are **consequences** of algorithm to public, or personal interest? Significance? Number of **people affected?**

Discriminatory / Unfair

Mistake that denies a service

Censorship

Breaks law or social norm

False Prediction

Algorithm Auditing

Code Audit

Requires specialized knowledge, versioning, system setup, FOIA access, data access

Noninvasive User Audit

User surveys about their experience. Non-experimental, sampling issues, self-report



Scraping Audit

Analyze input / output. EULA, CFAA issues

Sock Puppet

Impersonate users with programs. Ethics of deception

Crowdsourced Audit

Have real users report back data about platform.
Combine with software reporting.

Christian Sandvig, Kevin Hamilton, Karrie Karahalios, Cedric Langbort. Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms. ICA 2014.

Auditing Challenges

“Proof” is Elusive

Lack of internal knowledge of systems
Interviews are important but can be adversarial

Legal and Regulatory Issues

Terms of Use, EULAs, DMCA, Computer Fraud and Abuse Act
FOIA: trade secrets, no document creation

Ethical Implications

Publication & gaming, individual privacy, deception



A Business Case for Transparency

ATM Fraud “protection”

Me in Berlin (x2): 30 minutes on the phone, a poor customer experience, no explanation

Me: “can you explain why it happened?”

Rep: “probably the machine misread your card.”

Adjudicating False Positives

Better data quality →
more accurate prediction →
less cost



Transparency Challenges

Trade Secrets

Including FOIA exception

Not source code, but benchmarks – e.g. restaurant inspections

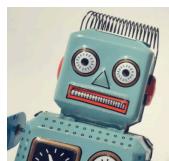
Gaming / Manipulation

But what if manipulation is aligned with positive behavior change for public?

Cognitive Complexity

Transparency information needs to be accessible and understandable

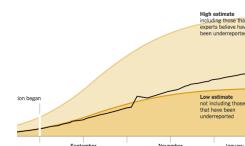
Workshop Transparency



Automated News Content



Algorithmic Curation



Simulation & Modeling

What could be made **public** about the algorithm, how it works, or how it is employed by the news organization?

Towards a Transparency Standard

1. Human Involvement

Goal / purpose / intent; Who has oversight? Who are authors?

2. Data

Data quality (accuracy, completeness, uncertainty, timeliness); representativeness of sample; collection, transformation, vetting, editing; public or private data; personal information or profiling

3. Model

What are features used; What are feature weights? What software was used?

Towards a Transparency Standard

4. Inferencing

Accuracy of classifications and predictions (FP, FN rates); benchmarks to standard datasets; Nature of errors; Confidence values

5. The Algorithmic Presence

If an algorithm in use? Personalization? A/B testing being used?



Ongoing Work Needed

User Modeling / User Studies

What are the decisions the public or other stakeholders would make based on the algorithm in question? What could a user know about an algorithm that would **change their interaction, decision, or outcome**? How would users respond to this information?

Manipulation – “Threat modeling”

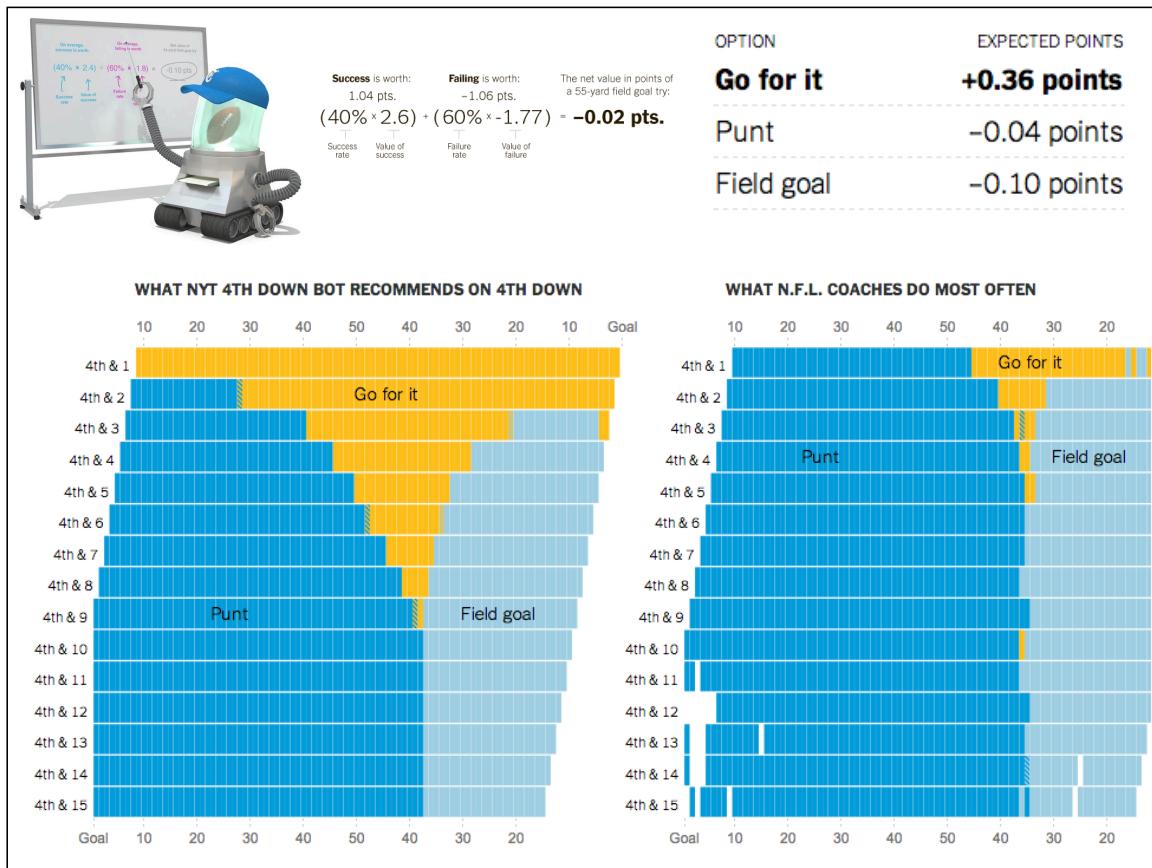
If a piece of information were made available about the algorithm, how might it be **gamed, manipulated, or circumvented**? Who would stand to gain or lose from that manipulation? What are the **consequences** of that manipulation to the public or organization? Can we build manipulation resistance?

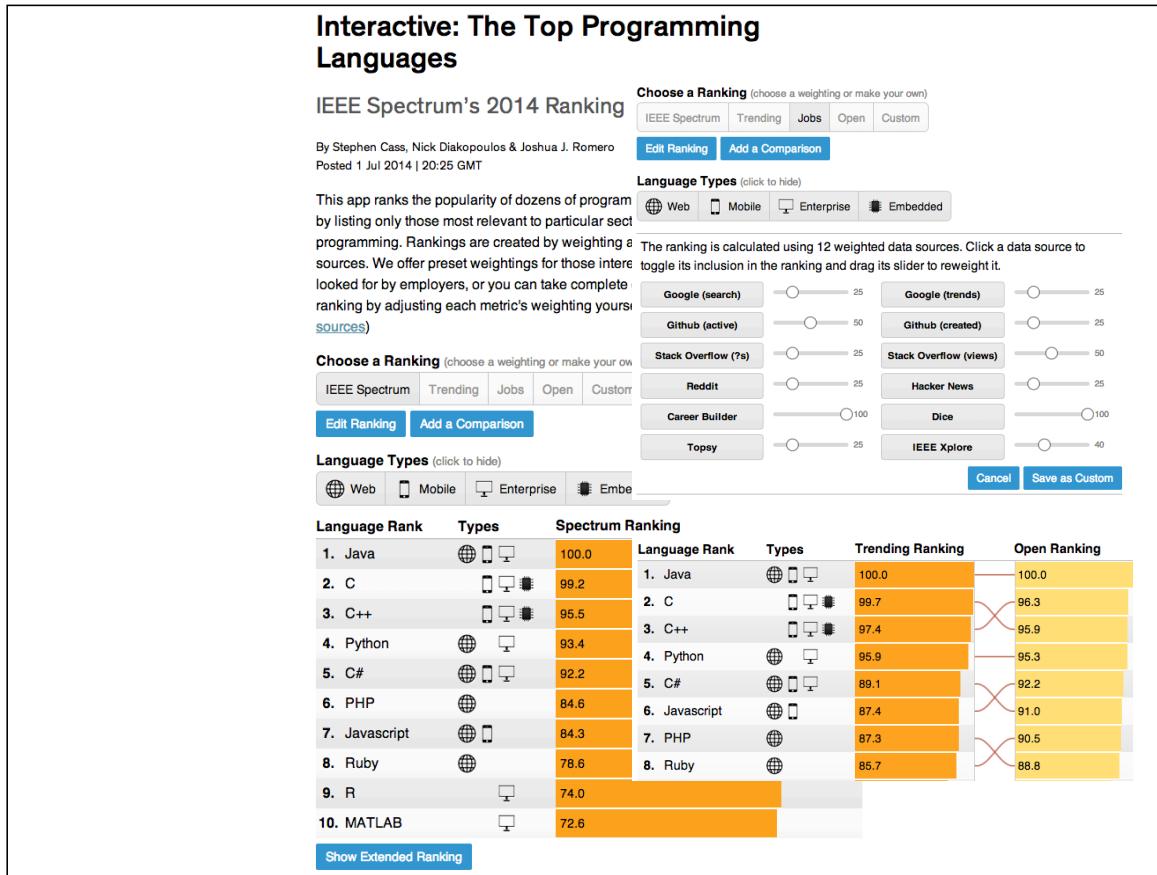
Cost

What degree of **time/money/effort** would be required by the creator to publish the relevant transparency info?

Presentation

What would be an appropriate vehicle or format to **publish** this info? How could this information be presented in a salient way to users? How **often** would this information need to be disclosed?





Thanks! Questions?

Contact

Nick Diakopoulos

Twitter: @ndiakopoulos

Email: nad@umd.edu

Web: <http://www.nickdiakopoulos.com>

More Info

N. Diakopoulos. Algorithmic Accountability Reporting: On the Investigation of Black Boxes. Tow Center, Feb., 2014.

N. Diakopoulos. Algorithmic Accountability: Journalistic Investigation of Computational Power Structures. Digital Journalism. 2015.

N. Diakopoulos, S. Cass, and J. Romero. Data-Driven Rankings: The Design and Development of the IEEE Top Programming Languages News App. Proc. Symposium on Computation + Journalism. 2014.