Alan Turing Institute: May 30, 2017

Algorithmic Accountability: Design for Safety

Ben Shneiderman **@benbendc**

Founding Director (1983-2000), Human-Computer Interaction Lab
Professor, Department of Computer Science
Member, Institute for Advanced Computer Studies
Member, National Academy of Engineering

Photo: BK Adams

UNIVERSITY OF MARYLAND

Interdisciplinary research community

- Computer Science & Info Studies
- Psych, Socio, Educ, Jour & MITH

hcil.umd.edu
vimeo.com/72440805

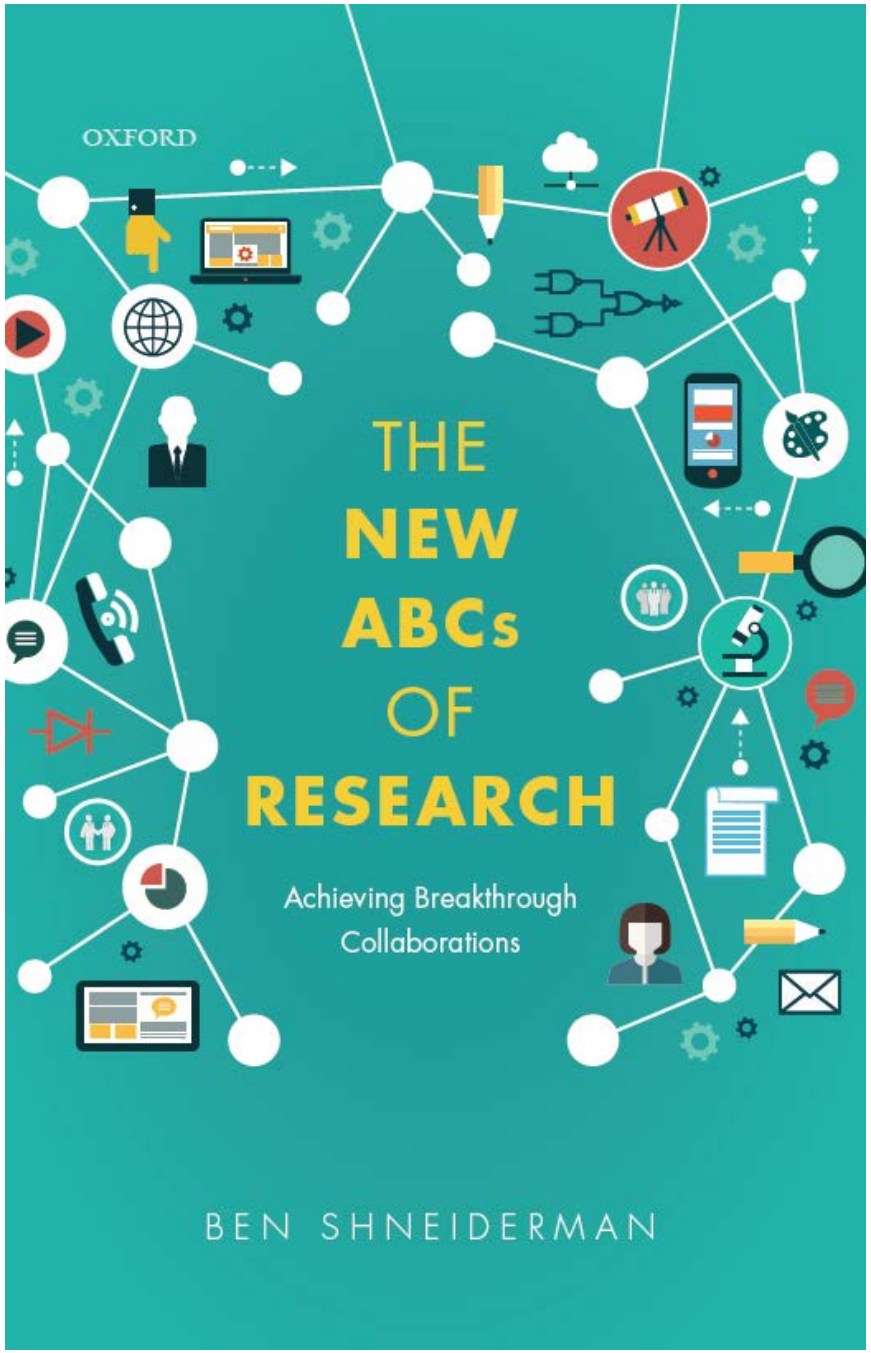
The New ABCs of Research

Guiding Principles

Applied & Basic Combined

Science, Engineering & Design

www.cs.umd.edu/hcil/newabcs



Designing the User Interface

Design Theories

Direct manipulation

Menus, speech, search

Social Media

Information Visualization

www.cs.umd.edu/hcil/DTUI6

Sixth Edition: 2016

Web links

Tiny touchscreen keyboards

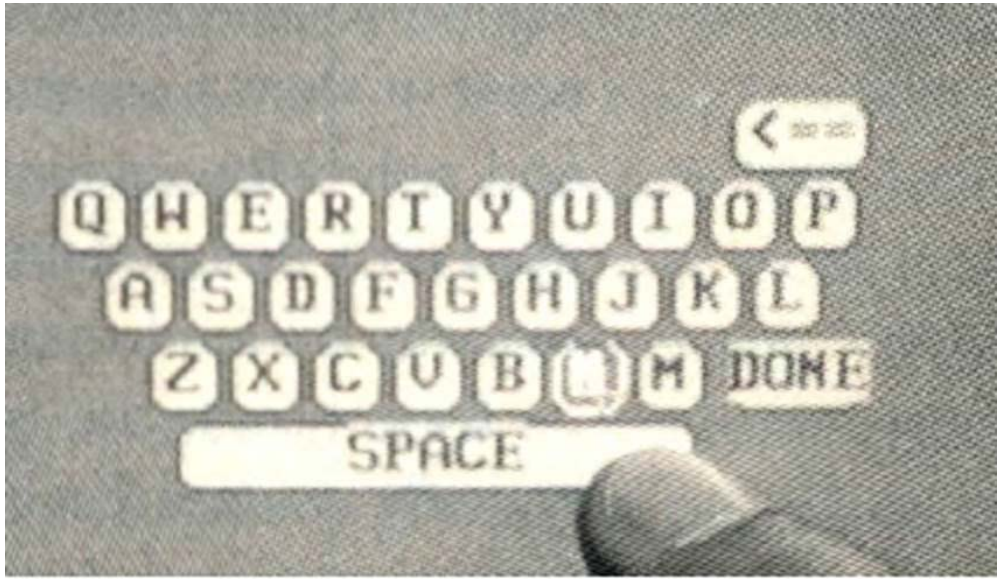
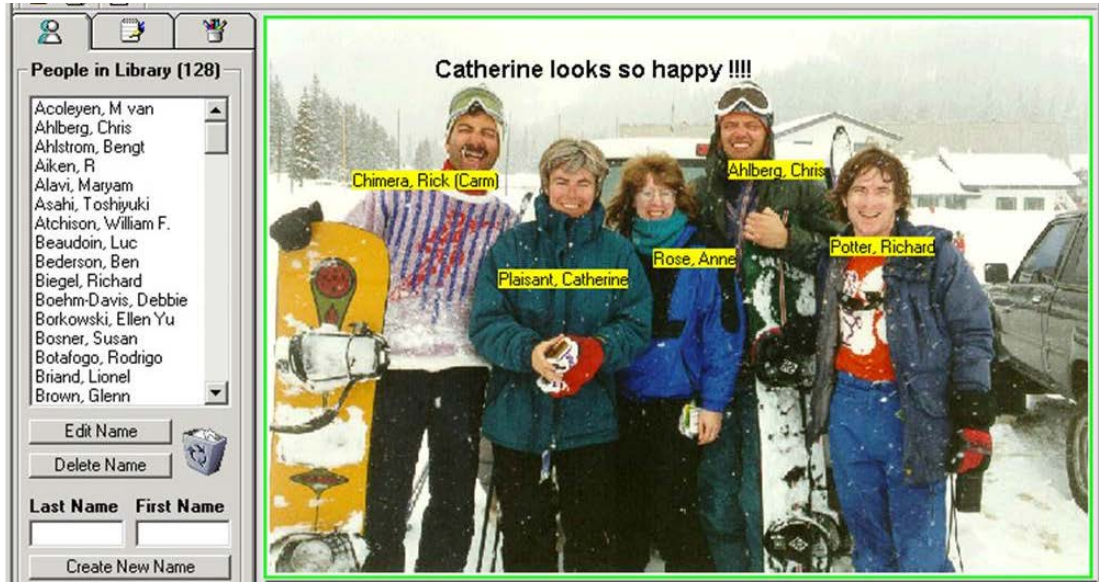
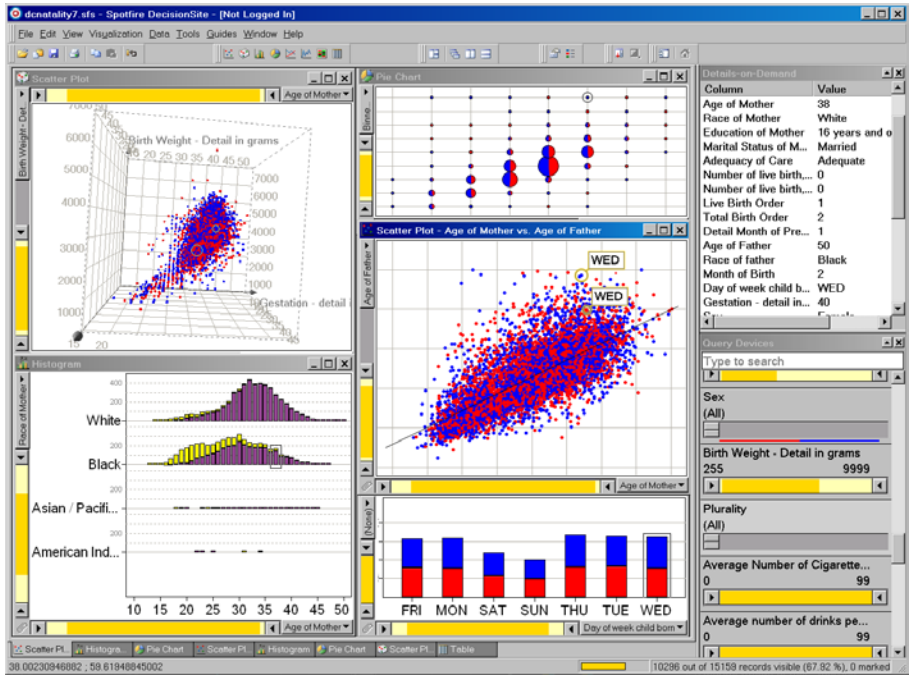


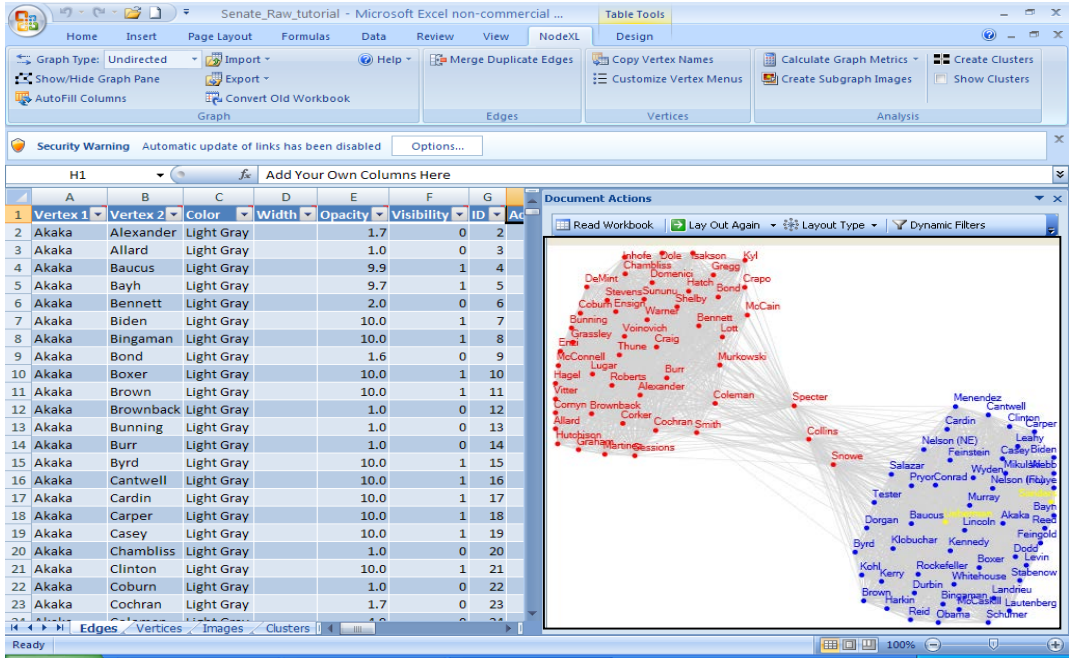
Photo tagging





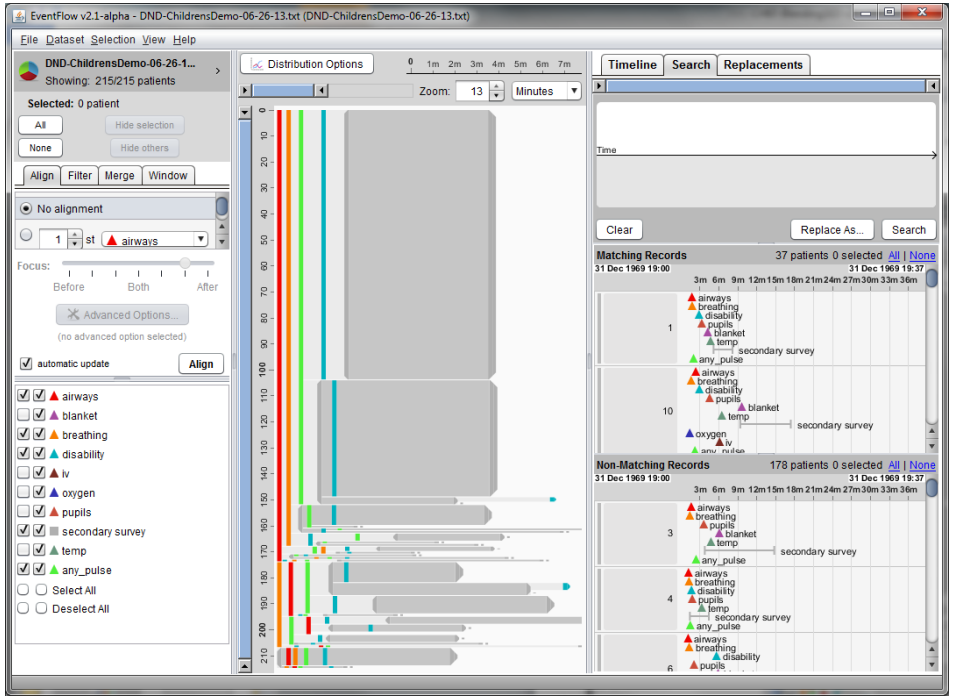
Spotfire

Treemaps
FinViz



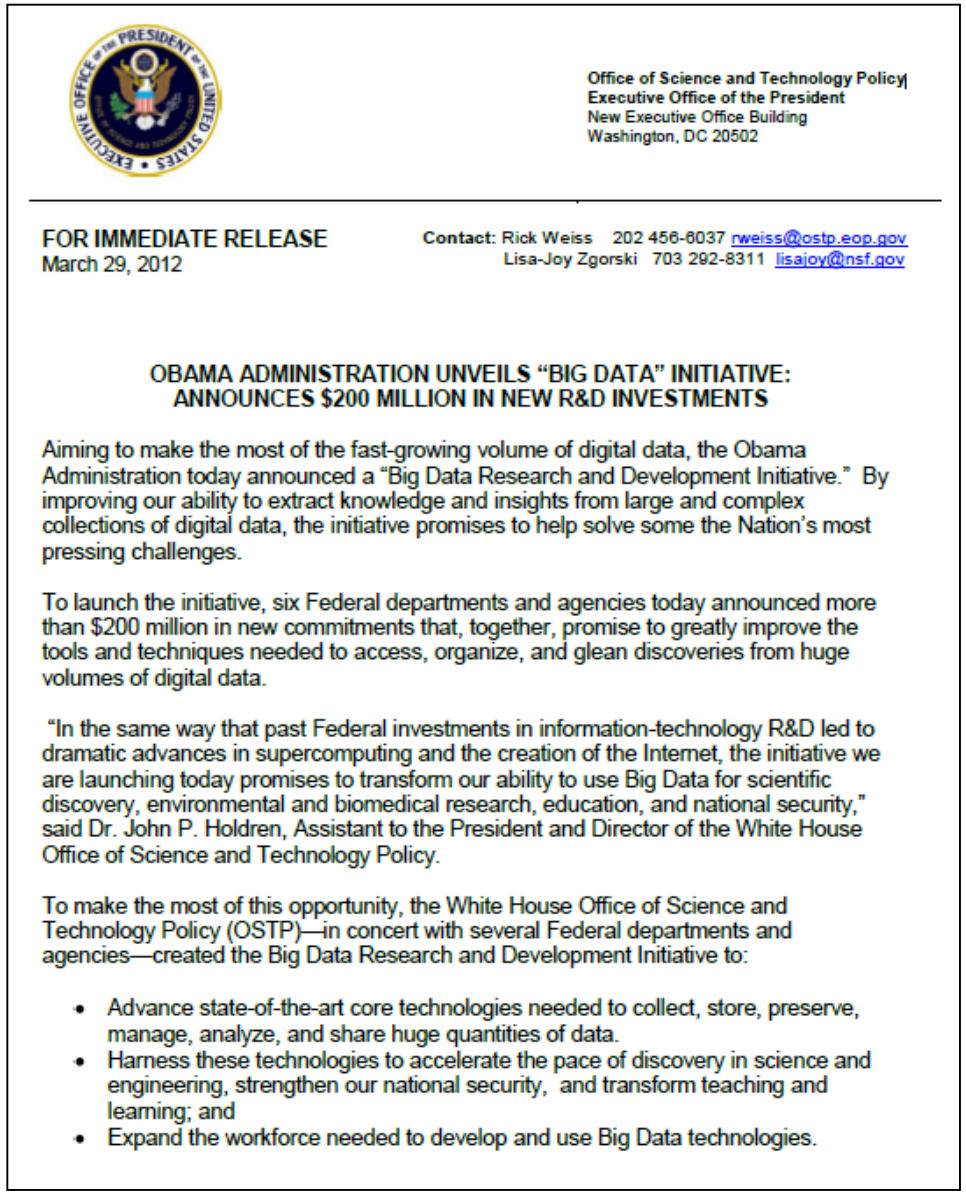
NodeXL

EventFlow




Wikimedia

Obama Unveils BIG DATA Initiative (3/2012)



Office of Science and Technology Policy
Executive Office of the President
New Executive Office Building
Washington, DC 20502

FOR IMMEDIATE RELEASE
March 29, 2012

Contact: Rick Weiss 202 456-6037 rweiss@ostp.eop.gov
Lisa-Joy Zgorski 703 292-8311 lisajoy@nsf.gov

OBAMA ADMINISTRATION UNVEILS "BIG DATA" INITIATIVE: ANNOUNCES $200 MILLION IN NEW R&D INVESTMENTS

Aiming to make the most of the fast-growing volume of digital data, the Obama Administration today announced a "Big Data Research and Development Initiative." By improving our ability to extract knowledge and insights from large and complex collections of digital data, the initiative promises to help solve some the Nation's most pressing challenges.

To launch the initiative, six Federal departments and agencies today announced more than $200 million in new commitments that, together, promise to greatly improve the tools and techniques needed to access, organize, and glean discoveries from huge volumes of digital data.

"In the same way that past Federal investments in information-technology R&D led to dramatic advances in supercomputing and the creation of the Internet, the initiative we are launching today promises to transform our ability to use Big Data for scientific discovery, environmental and biomedical research, education, and national security," said Dr. John P. Holdren, Assistant to the President and Director of the White House Office of Science and Technology Policy.

To make the most of this opportunity, the White House Office of Science and Technology Policy (OSTP)—in concert with several Federal departments and agencies—created the Big Data Research and Development Initiative to:

- Advance state-of-the-art core technologies needed to collect, store, preserve, manage, analyze, and share huge quantities of data.
- Harness these technologies to accelerate the pace of discovery in science and engineering, strengthen our national security, and transform teaching and learning; and
- Expand the workforce needed to develop and use Big Data technologies.

BIG DATA challenges:

- Developing scalable algorithms for processing imperfect data in distributed data stores

- Creating effective **human-computer interaction tools** for facilitating rapidly customizable **visual reasoning** for diverse missions.

http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release_final_2.pdf `

White House Big Data Strategies (5/2016)

- Create next-generation capabilities: foundations & technologies

- Support R&D on trustworthiness of data & resulting knowledge
  → better decisions, breakthrough discoveries & confident action

- Build cyberinfrastructure for agency missions

- Increase value of data → promote sharing & management of data

- Understand collection, sharing & use → privacy, security & ethics

- Improve the landscape for education & training

- Enhance connections in the innovation ecosystem

https://www.whitehouse.gov/blog/2016/05/23/administration-issues-strategic-plan-big-data-research-and-development

White House Big Data Strategies (5/2016)

- Create next-generation capabilities: foundations & technologies

- Support R&D on trustworthiness of data & resulting knowledge
  → better decisions, breakthrough discoveries & confident action

- Build cyberinfrastructure for agency missions

- Increase value of data → promote sharing & management of data

- Understand collection, sharing & use → privacy, security & ethics

- Improve the landscape for education & training

- Enhance connections in the innovation ecosystem

https://www.whitehouse.gov/blog/2016/05/23/administration-issues-strategic-plan-big-data-research-and-development

White House Big Data Strategies (5/2016)

- Support R&D on trustworthiness of data & resulting knowledge
  → better decisions, breakthrough discoveries & confident action

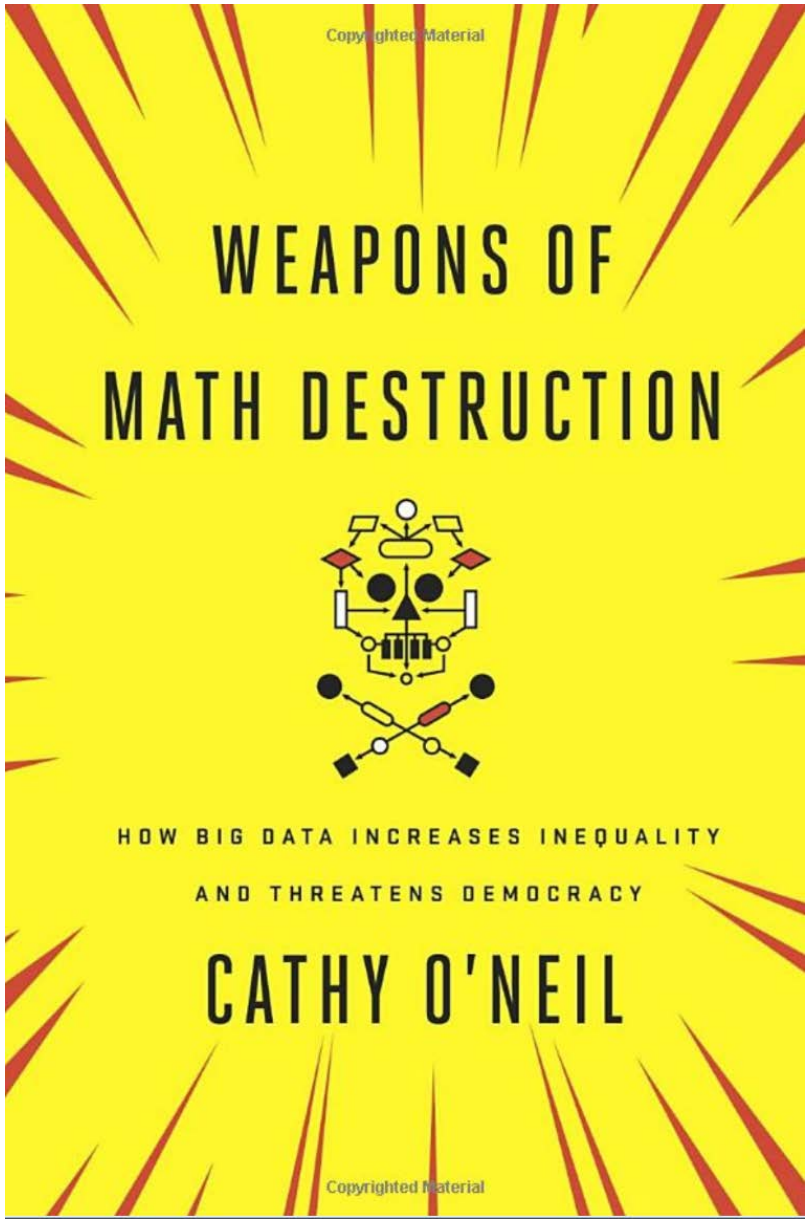
… promote transparency … including tools that provide detailed audits … to show … the steps that led to a specific action.

https://www.whitehouse.gov/blog/2016/05/23/administration-issues-strategic-plan-big-data-research-and-development

Weapons of Math Destruction
Cathy O’Neil

Opacity, scale & damage

“These … algorithms, slam doors in the face of millions of people, often for the flimsiest of reasons, and offer no appeal. They’re unfair.”



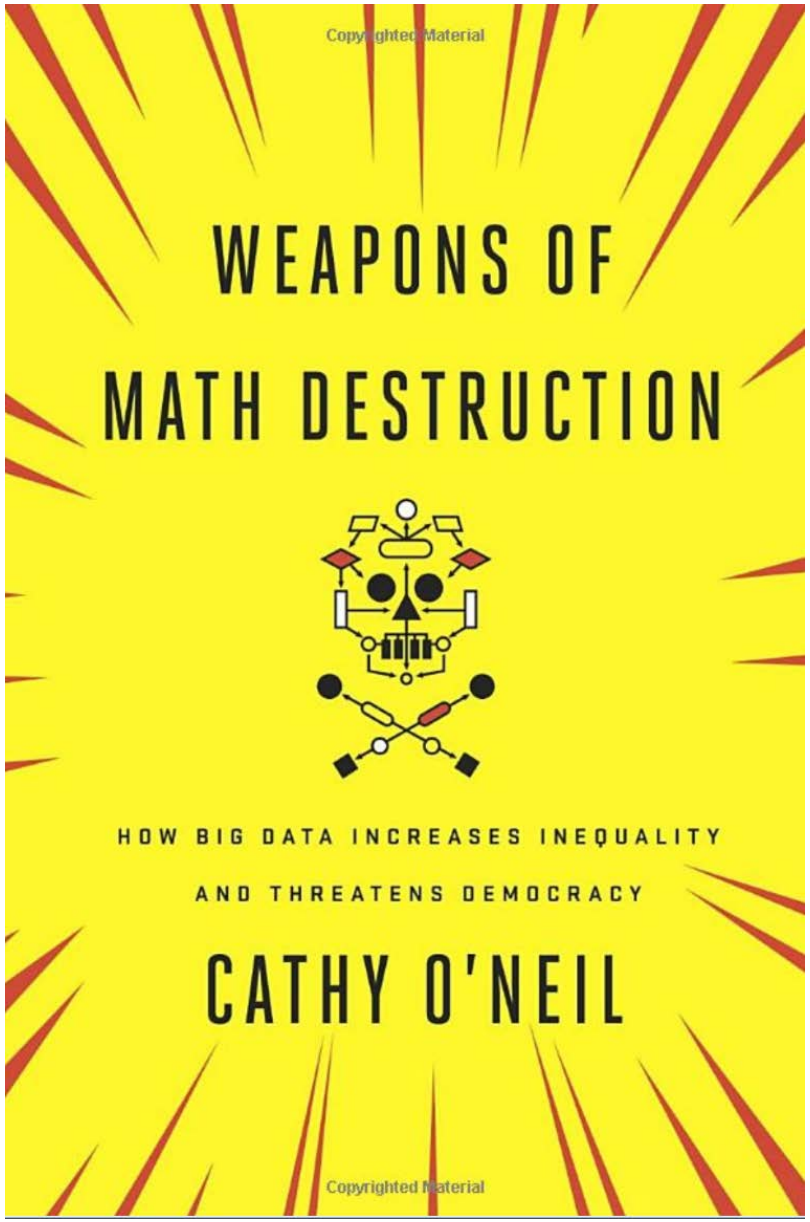
Permission: Cathy O’Neil

Weapons of Math Destruction
Cathy O'Neil

Opacity, scale & damage

"These … algorithms, slam doors in the face of millions of people, often for the flimsiest of reasons, and offer no appeal. They're unfair."

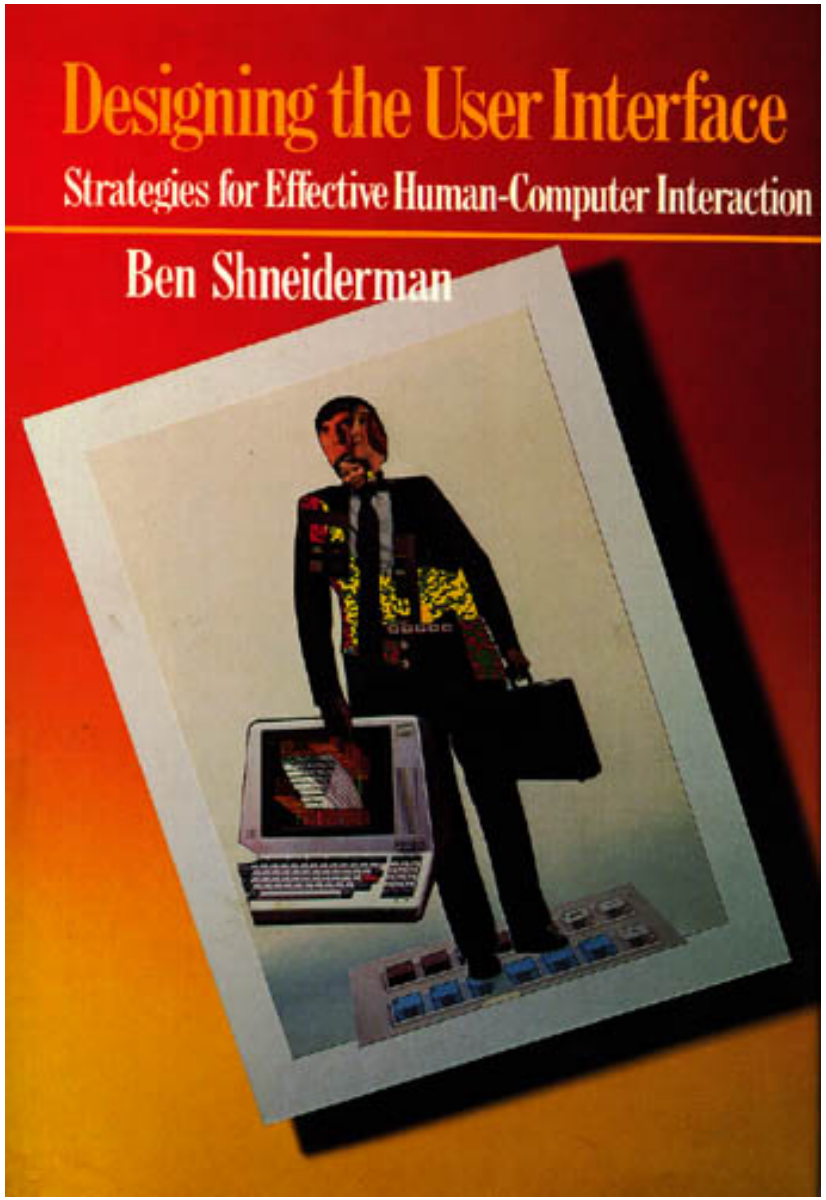
Algorithms can be biased, harmful & deadly!



Permission: Cathy O'Neil

Designing the User Interface

Balancing automation & human control



First Edition: 1986

Designing the User Interface

Ensuring human control
while increasing automation

Sixth Edition: 2016

Developer

Apple Design Guidelines

User Control

… people - not apps - are in control.

Flexibility

… users complete, fine-grained control over their work.

RESPONSIBILITY

PNAS Opinion: To mitigate the dangers of faulty, biased, or malicious algorithms requires independent oversight (November 29. 2016) http://www.pnas.org/content/113/48/13538.full

Updated May 25, 2017

Statement on Algorithmic Transparency and Accountability

by ACM U.S. Public Policy Council, approved January 12, 2017
ACM Europe Policy Committee, approved May 25, 2017

Computer algorithms are widely employed throughout our economy and society to make decisions that have far-reaching impacts, including their applications for education, access to credit, healthcare, and employment. The ubiquity of algorithms in our everyday lives is an important reason to focus on addressing challenges associated with the design and technical aspects of algorithms and preventing bias from the onset.

An algorithm is a self-contained step-by-step set of operations that computers and other 'smart' devices carry out to perform calculation, data processing, and automated reasoning tasks. Increasingly, algorithms implement institutional decision-making based on analytics, which involves the discovery, interpretation, and communication of meaningful patterns in data. Especially valuable in areas rich with recorded information, analytics relies on the simultaneous application of statistics, computer programming, and operations research to quantify performance.

There is also growing evidence that some algorithms and analytics can be opaque, making it impossible to determine when their outputs may be biased or erroneous.

Computational models can be distorted as a result of biases contained in their input data and/or their algorithms. Decisions made by predictive algorithms can be opaque because of many factors, including technical (the algorithm may not lend itself to easy explanation), economic (the cost of providing transparency may be excessive, including the compromise of trade secrets), and social (revealing input may violate privacy expectations). Even well-engineered computer systems can result in unexplained outcomes or errors, either because they contain bugs or because the conditions of their use changes, invalidating assumptions on which the original analytics were based.

The use of algorithms for automated decision-making about individuals can result in harmful discrimination. Policymakers should hold institutions using analytics to the same standards as institutions where humans have traditionally made decisions and developers should plan and architect analytical systems to adhere to those standards when algorithms are used to make automated decisions or as input to decisions made by people.

This set of principles, consistent with the ACM Code of Ethics, is intended to support the benefits of algorithmic decision-making while addressing these concerns. These principles should be addressed during every phase of system development and deployment to the extent necessary to minimize potential harms while realizing the benefits of algorithmic decision-making.

Principles for Algorithmic Transparency and Accountability

1. Awareness: Owners, designers, builders, users, and other stakeholders of analytic systems should be aware of the possible biases involved in their design, implementation, and use and the potential harm that biases can cause to individuals and society.

2. Access and redress: Regulators should encourage the adoption of mechanisms that enable questioning and redress for individuals and groups that are adversely affected by algorithmically informed decisions.

3. Accountability: Institutions should be held responsible for decisions made by the algorithms that they use, even if it is not feasible to explain in detail how the algorithms produce their results.

4. Explanation: Systems and institutions that use algorithmic decision-making are encouraged to produce explanations regarding both the procedures followed by the algorithm and the specific decisions that are made. This is particularly important in public policy contexts.

5. Data Provenance: A description of the way in which the training data was collected should be maintained by the builders of the algorithms, accompanied by an exploration of the potential biases induced by the human or algorithmic data-gathering process. Public scrutiny of the data provides maximum opportunity for corrections. However, concerns over privacy, protecting trade secrets, or revelation of analytics that might allow malicious actors to game the system can justify restricting access to qualified and authorized individuals.

6. Auditability: Models, algorithms, data, and decisions should be recorded so that they can be audited in cases where harm is suspected.

7. Validation and Testing: Institutions should use rigorous methods to validate their models and document those methods and results. In particular, they should routinely perform tests to assess and determine whether the model generates discriminatory harm. Institutions are encouraged to make the results of such tests public.

Algorithmic Transparency & Accountability

1. Awareness: Owners, designers, builders, users & other stakeholders should be aware of… possible biases and harm

Algorithmic Transparency & Accountability

1. Awareness: Owners, designers, builders, users & other stakeholders should be aware of… possible biases and harm

2. Access and redress: Regulators should adopt mechanisms that enable questioning & redress

Algorithmic Transparency & Accountability

1. Awareness: Owners, designers, builders, users & other stakeholders should be aware of… possible biases and harm

2. Access and redress: Regulators should adopt mechanisms that enable questioning & redress

3. Accountability: Institutions should be held responsible…even if it is not feasible to explain how the algorithms produce their results

Algorithmic Transparency & Accountability

1. Awareness: Owners, designers, builders, users & other stakeholders should be aware of… possible biases and harm

2. Access and redress: Regulators should adopt mechanisms that enable questioning & redress

3. Accountability: Institutions should be held responsible…even if it is not feasible to explain how the algorithms produce their results

4. Explanation: Systems & institutions that use algorithmic decision-making are encouraged to produce explanations of the procedures & decisions

5. Data Provenance: Algorithm builders should maintain a description of how training data was collected

Algorithmic Transparency & Accountability

5. Data Provenance: Algorithm builders should maintain a description of how training data was collected

6. Auditability: Models, algorithms, data & decisions should be recorded so that they can be audited

Algorithmic Transparency & Accountability

5. Data Provenance: Algorithm builders should maintain a description of how training data was collected

6. Auditability: Models, algorithms, data & decisions should be recorded so that they can be audited

7. Validation & Testing: Institutions should use rigorous methods to validate their models

Designing the User Interface

Ensuring human control
while increasing automation

Sixth Edition: 2016

INDEPENDENT OVERSIGHT

PNAS Opinion: (November 29. 2016) http://www.pnas.org/content/113/48/13538.full

To mitigate the dangers of faulty, biased, or malicious algorithms requires independent oversight

INDEPENDENT OVERSIGHT

Corporate internal audit committees & advisory boards
& External audits required by SEC

University Accreditation, NSF & EPSRC Advisory Boards

Zoning Boards, Planning Commissions, Environmental Impact Statements

NASA, FAA, FDA, DHS, Federal Reserve, etc.

Planning oversight

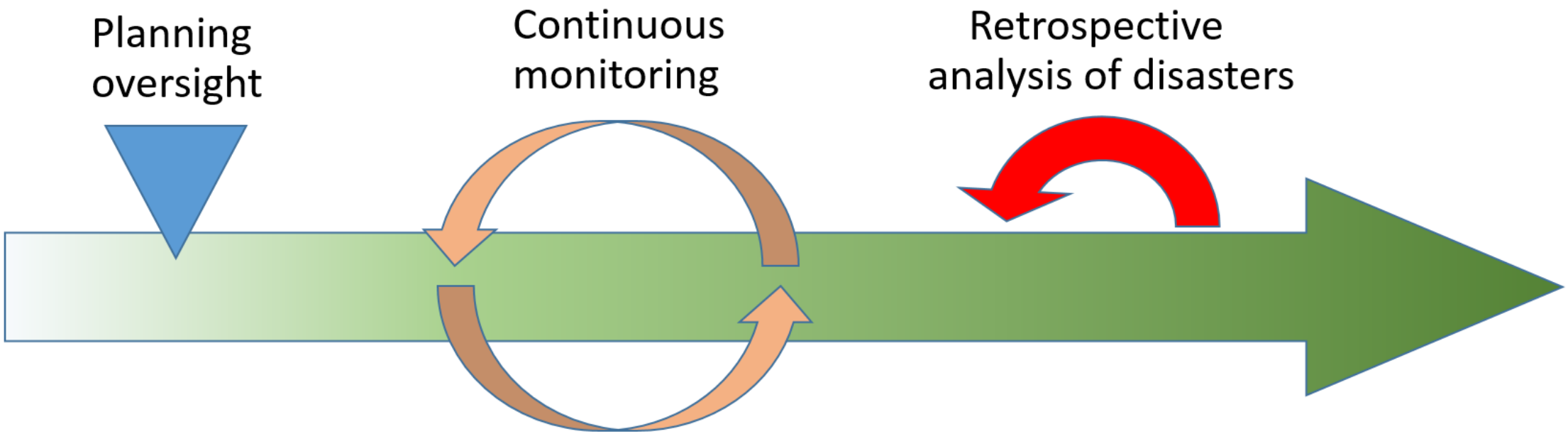
Continuous monitoring

Federal Reserve

Retrospective analysis

NTSB Tweet: 6:25 PM - 7 Jul 2013

INDEPENDENT OVERSIGHT



- Degree of Independence, subpoena power

- Powers to enforce recommendations

PNAS Opinion: (November 29. 2016) http://www.pnas.org/content/113/48/13538.full
To mitigate the dangers of faulty, biased, or malicious algorithms requires independent oversight

Designing the User Interface

Ensuring human control
while increasing automation

NATIONAL ALGORITHMS SAFETY BOARD

Sixth Edition: 2016

Clarifying responsibility accelerates quality

- Independent oversight: Open adversarial reviews
- Transparency: Open the black box
- Accountability: Open failure reporting
- Liability: No “hold harmless” contracts

RESPONSIBILITY