



Firestarter Talk Descriptions

The Social, Cultural & Ethical Dimensions of “Big Data”

March 17, 2014 - New York, NY

<http://www.datasociety.net/initiatives/2014-0317/>

As part of the conference, twelve participants gave five-minute firestarter talks. Below are the abstracts of their talks. To watch a video of the talks, visit the conference website at:

<http://www.datasociety.net/initiatives/2014-0317/>

Alessandro Acquisti: The Economics of Privacy (and Big Data): Facts, Myths, and Unknowns

A substantial amount of as of yet untested assumptions pervades the debate over the economic benefits of big data and its potential privacy trade-offs. I will use past and current findings from the economics literature to highlight a) which facts concerning the economics of privacy and personal data we have a decent grasp of; b) which myths we should try to dispel; and c) which crucial questions have not yet been solved, or even addressed, but should be.

Alex Howard: Data Journalism in the Second Machine Age: on Accountability, Algorithms and Transparency

Journalists have been adapting and adopting technology to gather information, report news and hold the powerful to account for centuries. In the 20th century, computer-assisted reporting provided investigative journalists with new ways to apply an empirical lens to creating acts of journalism. In the 21st century, data journalists are using and creating powerful new tools and platforms to understand the world, tell stories and inform the public.

Clay Shirky: Analog thumbs on digital scales: Hard-to-detect self-dealing by data-centric organizations

There are some risks of self-dealing by data-driven organizations that can't be solved through transparency. When algorithmic selection relies on random or frequently re-weighted inputs, an organization can run millions or billions of trial runs, then select the

one it prefers, allowing them to optimize for privately held preferences. Even an auditor with complete access to both the data and the algorithm couldn't detect this form of self-dealing. This points more generally to the need to consider governance alongside transparency as a check on abuse.

Corrine Yu: Civil Rights Principles for the Era of Big Data

On February 27, a broad coalition of leading civil rights and media policy organizations came together to endorse Civil Rights Principles for the Era of Big Data. For the coalition as a whole, and for many of the signatories, this is a historic first step into the national conversation that's starting now around big data. Through these principles, the signatories highlight the growing need to protect and strengthen key civil rights protections in the face of technological change. They call for an end to high-tech profiling; urge greater scrutiny of the computerized decision-making that shapes opportunities for employment, health, education, and credit; underline the continued importance of constitutional principles of privacy and free association, especially for communities of color; call for greater individual control over personal information; and emphasize the need to protect people, especially disadvantaged groups, from the documented real-world harms that follow from inaccurate data.

Deborah Estrin: You and Your Small Data

Each time you use their smartphone, social media, search engine, mobile game, or loyalty card, they implicitly generate a trail of digital breadcrumbs that together form a digital trace of your activities. We call these data that are particular to an individual, their small data. These data are used in aggregate by digital services to improve system performance, tailor service offerings, conduct research, and target ads. But these highly personalized, data can also be analyzed to draw powerful inferences about your health and everyday behaviors. In the future we are building towards at Cornell Tech and Open mHealth, and more broadly in the community, you would be able to opt in to access to your digital traces through Personal Data APIs, and to choose apps that privately process, fuse, and filter your Small Data for you.

Kevin Bankston: The Biggest Data of All

When talking about big data, it's worth stepping back to talk about the biggest data set of all, available only to ISPs and governments--the Internet and voice traffic that transits the Internet backbone. What are the current rules and practices when it comes to

accessing and using that data? What types of data are we talking about, and what kinds of distinctions between types of data make sense? For example, does the longtime distinction between communications "content" and "metadata", with the latter receiving weaker protection, make sense in an age of bulk data analysis where metadata can often be just as revealing, if not more revealing, than the actual content of your communications? And what responsibilities do ISPs and governments have when it comes to handling this vast stream of information that includes the mass of our personal and private communications and data?

Latanya Sweeney: Transparency Builds Trust

Our national security and largest hi-tech companies have found value in personal data sharing. They trust the quality of the personal data they receive to make revenue and national security decisions. The opposite is true too. Lack of trust breeds distrust. Not knowing what others do with the personal data they receive leaves the public distrustful and vulnerable. When data sharing is transparent (e.g., thedatamap.org), society can understand risks and harms and pose effective remedies and technical solutions. Approaches might be reporting arrangements recursively along data sharing chains or allowing individuals to audit their own data sharing chains (e.g., HIPAA).

Mark Latonero: Big data and human trafficking

This presentation will discuss the socio-technical complexities that arise as big data and analytics are applied to human rights interventions, using human trafficking as a specific example. The promises of big data approaches for human trafficking interventions raise important questions about data privacy, monitoring, and interpretation. What are the responsibilities of data and technology companies or computer and social science researchers (monitoring and analyzing a number of human sensor networks) when they are able to detect evidence of the most pressing human rights violations? Can policies be developed that strike the right balance between privacy, data ownership, and human rights benefits?

Natasha Schüll: Tracking for profit; tracking for protection

Casinos' data-intensive player-tracking systems, developed to gather behavioral intelligence, run predictive analytics, and more profitably market to gamblers (by mail or in the moment), are in some cases also equipped with algorithms that can detect patrons' problematic gambling behavior and prompt real-time intervention. In some

jurisdictions, player-tracking systems are set up to perform both roles simultaneously, despite their apparent conflict. This case, in which a single data-monitoring platform serves at once to persuade and to protect consumers, brings to the fore the challenging tensions at the heart of "big data" as it plays out in our everyday life transactions.

Nick Grossman: Regulation 2.0: Bringing an Internet Approach to Real-World Regulation

The last 15 years on the internet has been an experiment in generating trust among strangers. Along the way, Internet-based networks, from eBay to Airbnb, have invented increasingly sophisticated trust and safety mechanisms that in many ways mirror the intentions of government regulations, but achieve their goals through intensive use of peer review and data, rather than up-front granting of permission. This approach to regulation is in itself a major innovation. So, as more and more internet-based businesses bump into real-world regulatory regimes, the question is: should we regulate them the old way or the Internet way?

Nuala O'Connor: The Digital Self & Technology in Daily Life

The increasing ubiquity of technology in daily life leads us to take for granted the sometimes opaque decision making inherent in seemingly personalized algorithms. O'Connor poses the questions "do we know what motivates the machine" and what is the impact on individuality, on learning and distribution of knowledge, on our digital selves.

Tim Hwang: On Cognitive Security

Data is the raw material of prediction. To that end, the increasing availability of social data opens the possibility of predicting and influencing the behavior of large groups. This data is -- and increasingly will be -- used by actors to shape (and reshape) social systems at scale. This talk will briefly explore what would appear to be the logical next step: what occurs when quantitative methods of influence and the parties that use them come into competition with one another? It will discuss the concept of "cognitive security," and how big data is poised to inform the future landscape of persuasion.