RESPONSES OF DR. MICHAEL KEARNS
QUESTIONS FOR THE RECORD
HEARING ON "ALGORITHMS: HOW COMPANIES' DECISIONS ABOUT DATA AND CONTENT IMPACT CONSUMERS."
SUBCOMMITTEE ON COMMUNICATIONS AND TECHNOLOGY AND THE SUBCOMMITTEE ON DIGITAL COMMERCE AND CONSUMER PROTECTION
HOUSE COMMITTEE ON ENERGY AND COMMERCE

The Honorable Robert E. Latta

1. **How is information that companies obtain about consumers used to model an individual and make predictions about a person's behavior?**

A. Detailed information about consumers (gender, age, race, location, app usage, search queries, etc.) is analyzed using modern machine learning and statistical methods to create highly individualized predictive models. These models may discover informative combinations of the inputs that would be difficult for humans to discern, and may make predictions about consumers that are much more revealing than the raw data. Large and diverse data sets about consumers are the foundation for effective models that predict behavior.

   a. **What are the benefits and dangers of these models?**

   A. These models present a number of substantial benefits to both consumers and advertisers. For example, these models help consumers make purchasing decisions based upon their interests, and they help marketers target advertisements to consumers who are likely to be interested in their products. The potential dangers of these models involve their impact on (1) consumer privacy, (2) the incentive that online companies have to amass massive amounts of information about consumers, and (3) the potential for bias or discriminatory behavior based upon the information produced by the models.

2. **Can real or perceived bias be cured in highly complex algorithmic systems to enhance reliability or intended outcomes?**

   a. **If so, how do online platforms conduct this curing or correction process?**

   A. In the past few years, new algorithmic research has emerged that may provide practical methods to reduce bias, while still achieving good predictive accuracy. These methods include algorithms for auditing predictive models for bias, and reducing or correcting that bias. Online platforms are still in the process of considering implementing such methods into their modeling.

3. **There are conflicting reports about how accurately companies can predict what consumers are really interested in. For instance, the Wall Street Journal wrote in November 2017 a story entitled "Google Has Picked An Answer For You−Too Bad It's Often Wrong." In that article, they note "the Internet giant is promoting a**

single result over all others, and many are contentious, improbable or laughably incorrect." Are we still in the early stages of companies being able to accurately gauge their users' interests?

A. That particular article is talking about something quite different than modeling consumer behavior, preferences and desires. Instead, it is focused on more of a pure language understanding problem, such as answering questions like "Does money buy happiness?" or "Who are the worst CEOs of all time?" For these questions, the massive amounts of data that companies like Google, Amazon, and Facebook have collected on consumer behavior, at the collective and individual level, is not especially helpful. So, for example, your Amazon purchases, GPS coordinates and Google searches are incredibly valuable in predicting your future online and offline behavior, but they do not help answer the question about money buying happiness.

a. Are there any special considerations for Internet service providers?

A. As I mentioned in my testimony, in general, the large consumer-facing tech companies have amassed large and diverse data sets that are directly relevant to making detailed inferences about individuals, including search queries, shopping behavior, location data, and social interactions. ISPs generally do not have access to the same depth or breadth of data, in part because of packet encryption via the https protocol.

The Honorable Gregg Harper

1. **Based on your research, do you think consumers do things they otherwise would not because of ho[w] their data is being used? Or are they instead being presented options that they may not have known they had?**

A. Consumers are definitely presented with options they would not have had if their data was not collected and analyzed by online platforms. In general, every Amazon recommendation you receive, and the ads you see on Google, are tailored and specialized based on your particular past behavior. Sometimes, this specialization may present users with beneficial choices — as in when Amazon recommends a book I would love that I didn't even know about. And sometimes these choices may be detrimental or even discriminatory, as when ads for high-interest payday loans are targeted towards low-income individuals.

The Honorable Michael C. Burgess

1. **If a digital platform or intermediary knows all of a user's travels over the course of weeks or months, is there anything they can't deduce by correlating location with mapping? For example, can illness be inferred if a person is repeatedly going to an out-patient facility?**

A. In general, GPS and other precise location-based data is tremendously powerful, especially when combined with other public or commercial data sets mapping "points of interest" (stores, parks, medical facilities, homes, etc.) to physical location. The example you give is more than plausible — location data showing a consumer repeatedly visiting a

chemotherapy facility might indeed strongly suggest that person, or someone close to them, is receiving treatment.

The Honorable Adam Kinzinger

1. **Could a malicious actor hack and combine the collected data to create a full profile of an individual and use that profile to access their accounts?**

A. Protecting the security of consumer information must be a priority. A malicious actor could obtain detailed profiles of people by hacking into a data source or company with sufficiently rich and diverse consumer information. It's conceivable that a hacker could develop a level of knowledge about a consumer to answer basic account security questions or correctly determine a consumer's password.

2. How good of a model of an individual can be made from existing data sources?

A. The largest consumer-facing tech companies have amassed massive data sets that have enabled them to develop very accurate models of individual consumers.

   a. **Do you think it is possible to predict the wants of a consumer before that consumer knows them?**

   A. Yes, machine learning and algorithms are widely used by consumer-facing technology companies to predict consumer purchasing habits, potential social connections, and search queries.

2. **How do algorithms reflect the biases and interests of their creators? Is it possible to write non-normative code?**

A. In general, I do not believe there is a widespread problem of programmers encoding their personal beliefs and biases directly into their algorithms. But there is no shortage of other ways in which deployed algorithms and models demonstrably exhibit biases (by age, gender, race, etc.). Regarding writing non-normative code, there is indeed a recent but growing body of research providing design principles for making algorithms less biased.