**Friday Podcasts**

Name: Sanjay Jagadeesh

**Episode Title:** “The Algorithms Around Us” **Podcast:** *Science Friday* **Date:** Sept. 14, 2018

|  |  |  |
| --- | --- | --- |
| **Headings** (after)  Uses of Algorithms  Rise of AI  Trusting algorithms (what is behind AI)  Algorithms & People  When to use algorithms  Challenges/Problems of algorithms/AI  Challenges with algorithms in use  How to overcome the problems of algorithms  The need for regulation  Limitations of algorithms  Larger Issue  Unanswered Questions  Legal Steps Being Taken | **Notes** (during)  California judges are using the advice of an algorithm to decide which defendants should and should not be released.  Algorithms have been used as judgement on which inmates get released on parole.  Algorithms are used in:   * The car industry * The medical world * Social media feeds   Algorithms have found their way into our everyday lives.  Algorithms have been used for a long time (dating back to the 1920s and 1930s).  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~  In the last 5-10 years, there has been a change in the amount of data collected on us and in how that data is analyzed and used to predict behavior.  With the rise of Artificial Intelligence (AI), algorithms have been given a lot more power and are being used in more situations to make decisions.  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~  The only way to objectively judge whether an algorithm is trustworthy is by getting to the bottom of how it works so you can right the error.  Often, when you “dig behind the surface” of something that looks magical (such as AI), who see something simple behind the scenes.   * There may be reckless things there too.   You cannot look at algorithms in isolation: there are people who write the algorithms.  You must accept that when algorithms are out, they are in the world.  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~  Algorithms are being used by people, about people.   * We all have inherent flaws & subconscious biases.   We are good at both completely believing and completely disbelieving machines.   * This is happening within the people who make the algorithms.   We need to think about humans and machines together, not separately.  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~  Different situations can require different responses to algorithms:   * There are situations where you want the best prediction you can get (the most accurate). * **IF** an algorithm can prove to be better than a human it will get more power/sway (hand over some level of control).   Ex: cancer diagnosis algorithms that screen biopsy slides and look for tumors.   * Though they can make mistakes (which you need to work carefully around), they may be more sensitive than a human pathologist to tiny indications that may show a possibility of cancer.   There could also be situations (particularly in the criminal justice field) where it may not be ideal to hand over control to algorithms. Even if we do hand over control, it may only be in specific ways.  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~  Some of the challenges are that there are often laudable goals behind using algorithms. Ex:   * Reducing pretrial incarceration. * Eliminating money bail. * Eliminating the idea of punishing people because they are poor.   It could be worthwhile to see if AI can help achieve these goals.  The problems are more subtle:   * Just because an algorithm makes a recommendation does not mean you have to follow it. * It is not clear how algorithms make predictions. * It is not clear whether the data being used to train the algorithm has the right signals (the algorithm may be predicting for a slightly different scenario then the one you want it to predict for). * There are many subtleties that require extensive domain knowledge (cannot just use a generic algorithm).   Within the systems, there may be problems:   * Ex: Bias and unfairness in the material used to train the algorithms (either done intentionally or unintentionally). * The algorithm may reflect society accurately but sometimes you do not want that.   You do not want an algorithm to reflect the biases of the past but to help us move towards a better society (not just in the right direction).  There are questions completely outside of the algorithm as well (ex: what the right direction is and how should we move that way).  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~  Algorithms are used in the hiring process (who to hire and who to not).   * Can lead to biases. * There may be something that you do not know about that could be a red flag to the algorithm that stops the company from hiring you.   Once your information is entered into an algorithm, it takes on an air of authority that makes it almost impossible to argue against (could be unfair).  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~  The question may not be “How can we perfect AI?” but “How can we design them for redress/appealability?”  It is not enough to evaluate how the algorithmic system works. You need to evaluate how it affects the parts around it.   * Minor changes in data (or even one erroneous data point) may make a big difference in the decision of an algorithm.   Checks and Balances:   * There is a larger system of decision making in which algorithms should be a part of (humans are also included).   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~  Algorithms are not made with checks and balances in mind (they are made as generic processes that can replace humans).   * There are algorithms that do everything. * There are new AI patterns all the time (which then do new algorithms).   There is a lot of hype around algorithms (some are just sophisticated methods, not AI).  There is a need for a regulating body on how good an algorithm is (Ex: People are basically allowed to use anything that they’ve created on members of the public).   * A system not only to protect intellectual property but to also evaluate the benefits an algorithm offers to society compared to its costs.   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~  Algorithms are not so good (at least not yet) at open ended problems (ex: “why does my stomach hurt?”) as training it becomes much harder (there could be multiple different answers said in numerous ways).   * The more defined/precise a problem is, the more beneficial an algorithm could be.   Within algorithms, there could be unseen issues (ex: does it work equally wee for dark skinned people versus light skinned people).  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~  Larger issue: a lot of the work in AI (especially deep learning) is how we represent information (which is a complicated thing itself).   * If we could find the right representation of the data, then the inference of that data would be easy.   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~  There are questions on what an algorithm would choose in a dire situation (ex: the trolley problem).  Using algorithms can lead to questions about governance (are you still in control?)  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~  European Union’s General Data Protection Regulation (GDPR):   * Put the power slightly back into the hands of an individual (can still hide information in terms and conditions). * You are allowed to ask an algorithm why it made the decision it made regarding you (what makes a valid explanation?) * The law has now provided the opportunity to think through research, how we ask these questions, and how we should solve them. | **Academic vocabulary…**  Deployed  Objectively  Isolation  Inherent  Biases  Laudable  Subtly  Extensive  Inevitably  Resolve  Appalling  Redress  Erroneous  Biplane  Critically  Aggregating  Novelty  Assessing  Inference  Conundrum  Caveat  Efficacy  Constitutes |
| **Technical terminology…**  Algorithms  Data  Training  Artificial Intelligence (AI)  Machine Learning  Domain  Black Box Algorithm  Encoding  System  Black Boxes  AI pattern  Intellectual Property  Expert systems  Deep Learning  Source Code  IBM Watson  Babylon  European Union’s General Data Protection Regulation (GDPR) |

Preliminary Research

* Choose one of the following **topics** from the podcast and...
  1. Craft 2-3 **research questions** related to that topic.
  2. Pick one of those questions to research.
  3. Identify the **search terms** you’ll use when researching.
  4. Locate 3 quality **sources** that speak to that question.
     + NOTE: “quality” sources tend to have an author (often someone you can look up who has credentials or experience relevant to the topic), often cite their sources, are of a minimum length, often come from institutions we can recognize, tend not to be cluttered with advertisements, are often from a .edu or .gov. or .org domain, and depending on the topic, are current.
* Topics:

1. proliferation of algorithms
2. how algorithms work
3. bias and flaws in algorithms
4. sentencing, bail, and parole decisions
5. cancer diagnosis
6. general practitioner work
7. hiring and records
8. patents
9. driverless cars
10. data protection (GDPR in Germany)
11. Any other topic you notice or think about based on this podcast.

|  |  |  |  |
| --- | --- | --- | --- |
| **Topic chosen:** | **2-3 Research Questions related to this topic.**  (Highlight the one you choose to research) | **Search Terms:** | **Links to 3 sources:** |
| Bias and flaws in algorithms | How do biases and flaws find their way into algorithms? | Bias  Flaws  Algorithms  Impacts  People  Decisions | [Greenlining-Institute-Algorithmic-Bias-Explained-Report-Feb-2021.pdf](https://greenlining.org/wp-content/uploads/2021/04/Greenlining-Institute-Algorithmic-Bias-Explained-Report-Feb-2021.pdf) |
| What are the impacts of biases and flaws in algorithms and how do they impact people with the decisions they make? | [Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms (brookings.edu)](https://www.brookings.edu/articles/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/) |
| How do you reduce bias/flaws in an algorithm and/or their impacts on decisions made? | [18.1\_-\_REVIEWED-\_Serwin-\_Algorithmic\_Bias.pdf (iadclaw.org)](https://www.iadclaw.org/assets/1/7/18.1_-_REVIEWED-_Serwin-_Algorithmic_Bias.pdf) |