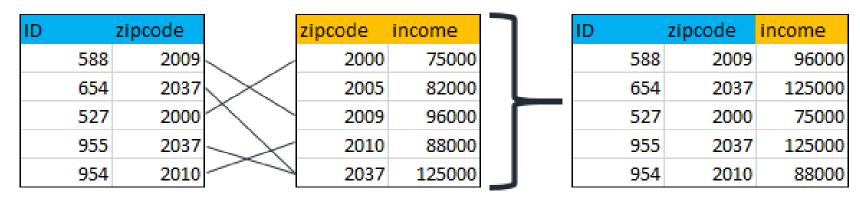
```
rel payons sints particle 45.62 families addiction 313 famous for goods and payolist, "13.65 fifting applies, standills and beautiful goods", home companies but
152 Septilit of Mildon's Presidency Office Communical, $552 Septility of Mildon's Presidency Office Communical Action Communical Communication and Action Communication Co
 rel survives state services, St. SI Security authorities, 113 Payment for goods and services, 113.06 Seeks and perindicals, have component the personnel recurrent, $6000, 44000, 15371.36
SEC SAMELIC OF SELECTION OF SEL
JAK Republic of Mildovo's Presidency DYFson Espectatus and other a presidency Office Especialist Section apparetus Coffice) of ministries and other a
oral purpose shots services $2.80 Semistive authorities, $13 Aspect for goods and services, $15.13 Telecommunication and motivates, have compared from personnel recorded.
the Sandile of Meldon's Frankley Office Caporatus), fitt Sandile of Meldon's Frankley Office Caporatus), ... the Cartrel approximation of Meldon's Frankley Office Caporatus)...
erall purpose state mervices, $1.82 Esecutive authorities, $12 Payment for goods and services, Special seams, S
 AND Resultic of Malabour's Presidency Office Companion), #AND Resulting Office Companion (Affice) of Milabour's Presidency Office Companion), #AND Resulting of Milabour's Presidency Office (Annual Companion)
and airpine state services. Mr. M. Electrics authorities, 113 Parent, for goods and services, 113, 18 Cornect resolv of agricular and inventors from component, for services), recornect,
NRT Republic of Wildows's Presidency Diffice Community. WALT Republic of Wildows's Presidency Diffice Community. Cethod of eccipints and other selections to the
eral purpose storm services, #1.82 Seniotive purposes for your and services, "12.28 the of generoses and local transplantacion distinction algor, have an
Working With Data
                                                                                                                                                                                                                            Stephen Downes
                                                                                                                                                                                                                                                                             December 4, 2021
 585 Court of Accounts, 8585 Clears of Maddags, ... 858 Central associate Coffice) of ministries and interior administrative authorities, 85 General parameter services, 765
tril services", 112 Margatury state social insurince proclum, 112.00 Margatury state social insurunce proclum, Special meior, Personnel, 6, 20106, 17415, 17
183 Court of Accesses, $165 Court of McCarry, agreement conflicts of abstacles and other additional or authorities at General partners about an access and access and
```

SET Republic of Mildres's Presidency Dffice Consentue). #SET Republic of Mildres's Presidency Dffice Consentus. Ceffice) of minimals and other selections such

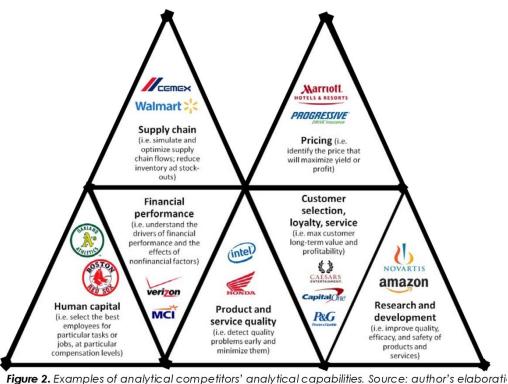
Combining Data

- Data seldom stands alone data is linked to other data, and can reveal more than was intended (e.g. criminal caught by sister's DNA) (Cohen, et.al., 2018)
- Data on travel may be used for discovering things far beyond mere travel patterns; data on purchases are not solely relevant to purchases; and so on. (Hand, 2019)



https://www.displayr.com/what-is-data-merging/

Unpacking Implications



CDS unpacks "the work these data assemblages do in the context of dataveillance, the attrition and loss of privacy, the impact of profiling and social sorting, and the downstream use of data, to mention but a few."

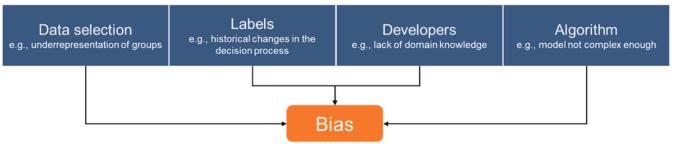
Prinsloo 2019 https://epress.lib.uts.edu.au/journals/index.php/JLA/article/view/6589/7307 https://www.semanticscholar.org/paper/When-big-data-meets-dataveillance%3A-the-hidden-side-Esposti/11aed548023909f4efd4c9b6d7d27d0cf70ae8af

Bias (Causes)

- An incomplete or skewed training dataset
- Labels used for training: training data is labeled in order to teach the model how to behave, and humans create these labels...

 Features and modeling techniques - the measurements used as inputs for machine-learning models, or the actual model training

itself



Josh Feast. (2019). 4 Ways to Address Gender Bias in Al. Harvard Business Review Online. November 20, 2019. https://hbr.org/2019/11/4-ways-to-address-gender-bias-in-ai

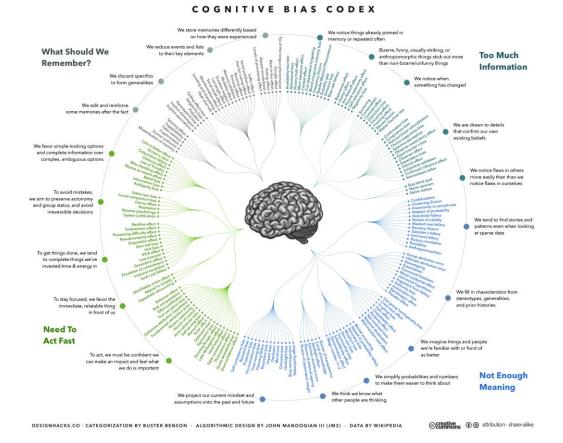
Bias (Types)



https://statanalytica.com/blog/bias-in-statistics/

Cognitive Bias Codex

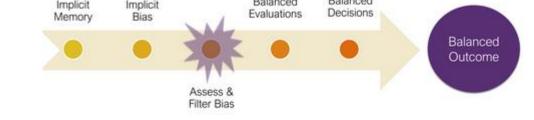
https://greatminds.co nsulting/insight/cogni tive-bias-a-mentalshortcut-that-causesa-mistake-inreasoning



https://greatminds.consulting/wp-content/uploads/2019/08/The Cognitive Bias Codex - 180 biases designed by John Manoogian III jm3.png

Bias (Remedies)

- Ensure diversity in the training samples
- Ensure that humans labeling come from diverse backgrounds.
- Measure accuracy levels separately for different demographic categories to identify when one category is being treated unfavorably.
- Solve for unfairness by collecting more training data associated with sensitive groups.



Josh Feast. (2019). 4 Ways to Address Gender Bias in Al. Harvard Business Review Online. November 20, 2019. https://hbr.org/2019/11/4-ways-to-address-gender-bias-in-ai

Data and Objectivity

- Data should never be accepted on faith
- There is no such thing as context-free data; data cannot manifest the kind of perfect objectivity that is sometimes imagined.
- Lisa Gitelman and Virginia Jackson point out in the introduction to the aptly titled "Raw Data Is an Oxymoron" (2013)

All from Radan, 2019: 16

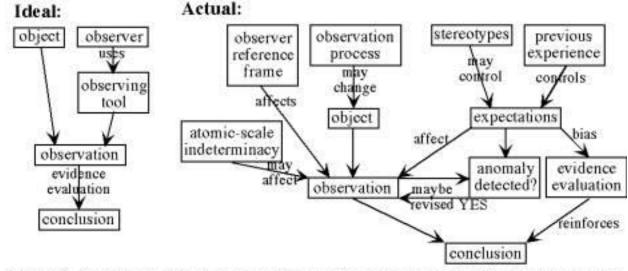


Figure 22. Flowcharts of ideal and actual interactions concerning experimental observations.

Data Risks

- The danger of stale and outdated data (Cohen, et.al., 2014)
- Sensitive files subject to regulations like GDPR, HIPAA and PCI (Varonis, 2019)
- Lack of space, time, and social context limitation on scope of data (Hand, 2018)
- Use for unexpected purposes and to reveal unexpected information
- Risk of exceptional intrusiveness since
- Potential for misuse, privacy breach, blackmail, and other crimes.
- Ghost users and accounts

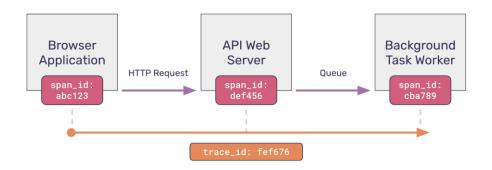
We need to be clear about who benefits and who incurs the risk



https://www.varonis.com/blog/data-risk-report-highlights-2019/

Data Tracing

- Using markers or signs to track the flow of data through a network
- "We call this new verification method 'radioactive' data because it is analogous to the use of radioactive markers in medicine."
- Examples: watermarks, embedded metadata, hash addressing



Alexandre Sablayrolles, Matthijs Douze, Hervé Jégou. (2020). Using 'radioactive data' to detect if a data set was used for training. Facebook Artificial Intelligence. February 05, 2020.

https://ai.facebook.com/blog/using-radioactive-data-to-detect-if-a-data-set-was-used-for-training/

Data Ownership

- Data ownership refers to both the possession of and responsibility for information.
 Ownership implies power as well as control.
 - Who owns the data?
 - Who owns the information the data is about?

https://ori.hhs.gov/education/products/n_illinois_u/datamanagement/dotopic.html

Loshin (2002) identifies a list of parties laying a potential claim to data:

- Creator The party that creates or generate data
- Consumer The party that uses the data owns the data
- Compiler This is the entity that selects and compiles information from different information sources
- Enterprise All data that enters the enterprise or is created within the enterprise is completely owned by the enterprise
- Funder the user that commissions the data creation claims ownership
- Decoder In environments where information is "locked" inside particular encoded formats, the party that can unlock the information becomes an owner of that information
- Packager the party that collects information for a particular use and adds value through formatting the information for a particular market or set of consumers
- Reader as owner the value of any data that can be read is subsumed by the reader and, therefore, the reader gains value through adding that information to an information repository
- Subject as owner the subject of the data claims ownership of that data, mostly in reaction to another party claiming ownership of the same data
- Purchaser/Licenser as Owner the individual or organization that buys or licenses data may stake a claim to ownership