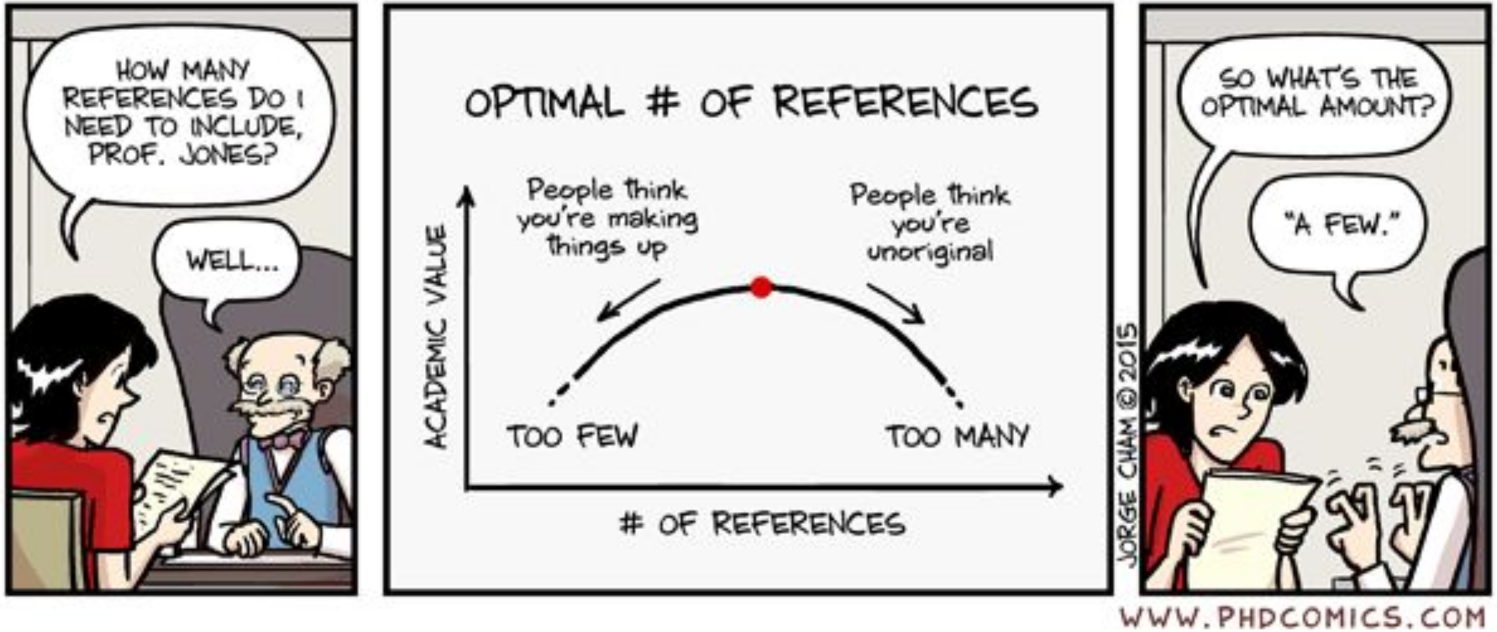
July 14th, 2022

Ethical Creation of Ethical Data

João Freitas
GM & Eng. Site Lead @ PagerDuty
ex-CTO @ Defined.ai

Contributors from Defined.ai: Rui Correia, André Monforte

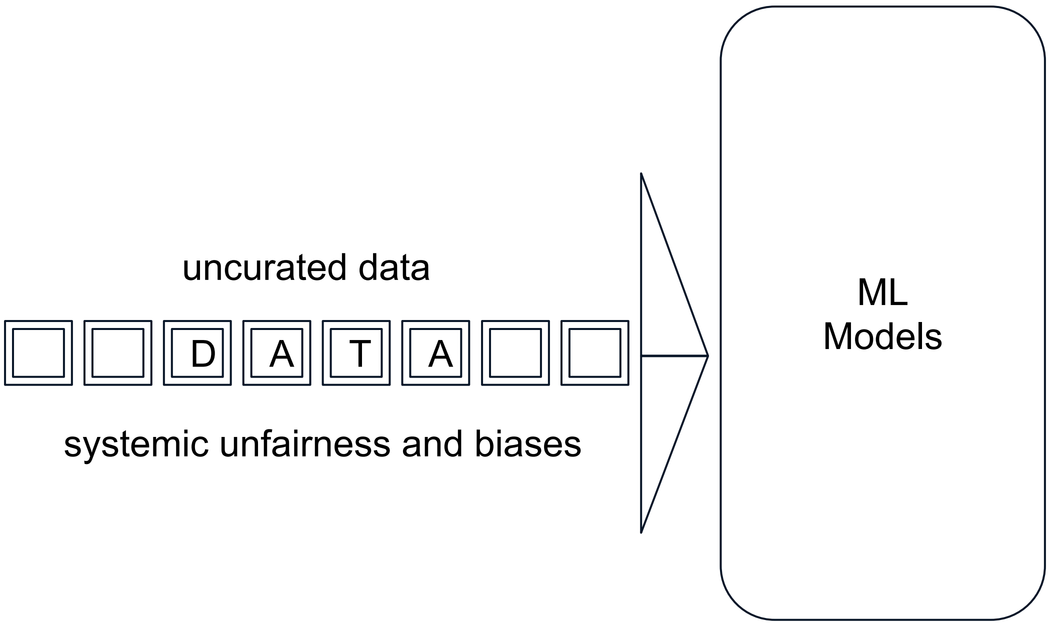
False Consensus Effect or … just an opinion



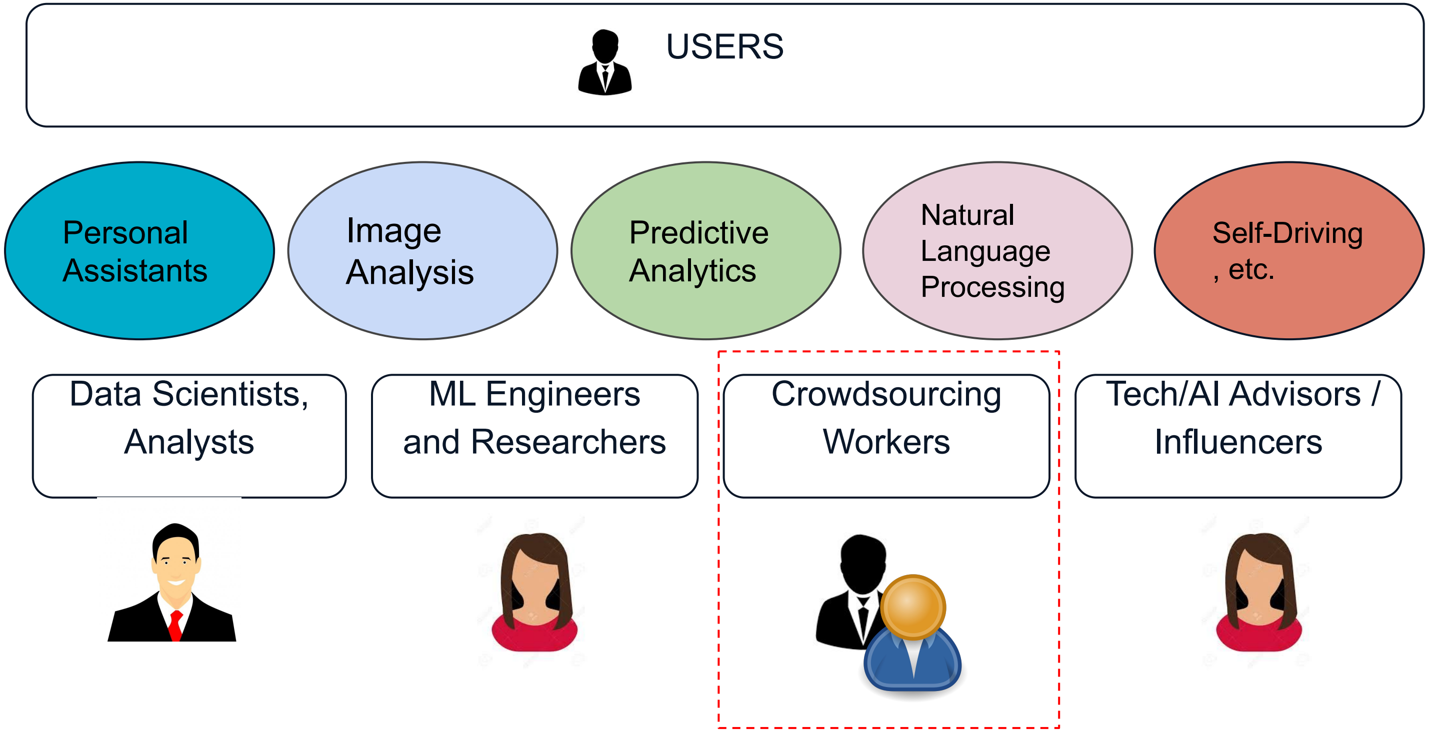
Increasing Interaction with technology

increasing data generation…

WIKIPEDIA
The Free Encyclopedia



Who is contributing to the data and the models?



Has AI become ubiquitous?

We are contributing to it and consuming it whether we are aware of it or not.

- mobile apps
- media
- online shopping
- autonomous vehicles
- etc.

…if so what are the ethical implications?

Fairness vs. Bias vs. Ethics

Fairness
Sociology

- Discrimination-free judgment
- Equal treatment
- Depends on individual perceptions, individual expectations, and context

Bias
Mathematics

- Measure of favoring/hindering something
- Bias can be a compensation mechanism, hence fair

Ethics
Philosophy

- Systematization of right and wrong
- Normative, hence superior political dimension for decision-making

e.g.: Intentional/Deliberate bias + unethical favoring = unfairness!

Fairness in AI

When the system acts **solely** based on domain-specific (and ethical) information

When the system’s performance is **comparable across distinct classes** of its user base.

Fairness in AI

When the system acts **solely** based on domain-specific (and ethical) information

V

Features & Algorithms

When the system's performance is **comparable across distinct classes** of its user base.

V

Data

Challenges to Fairness

Model definition goals
Usually does not take into account unfairness

Data collection & annotation
- Limited contributors
- Process errors

unfairness

Preparing data
- Attributes chosen (age, gender…)

unfairness

Model training
Algorithms can overgeneralize or overfit to data

unfairness

AI application

Testing
- Testing of the model might appear accurate
- Traceability is difficult

Unfairness in AI

Some examples

Image Recognition

- Inaccurate recognition or exclusion based on skin tone
- Unable to recognize certain faces in context

Predictions

- Inaccurate repeat criminal offender predictions
- Chooses most frequent group over another in job candidate screening

Customer Service

- Misunderstanding certain dialects in voice-controlled systems
- Generic, and potentially offensive, recommendation systems

Who is accountable and responsible for an unfair AI?

Challenges to Fairness

Unknowns

Not easy to initially detect, predict impact.

Traceability is difficult

Data source

- Models are tested for performance not always fairness
- Same data source used for training is often used for benchmarking

Modeling

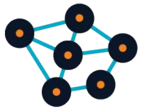
Models need to be versatile – not always equally fair and accurate for all parts of the world

Fairness

- Computational models are described by mathematics
- Fairness is hard to model

“Data-hungry” models

Training data is a vital component

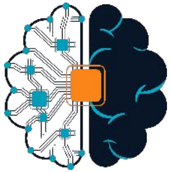


Machine Learning Algorithms

+

High Quality training data

=



AI

Data Bias

“Real-world” data perpetuate systemic inequalities in our society.

Fairness by ignorance
models don’t stop drawing inferences across features just because we haven’t explicitly labelled those features.

This can lead to

1. **Damaging brand image**
2. **Losing customers**
3. **Generating incorrect/ inaccurate results**
4. **Disregarding valued customer segments**

Data Bias
Challenges

Low Quality

Time Consuming

Hard to Scale

Compromises are often made to reduce time to market
Reducing Bias is expensive!

Ethical Creation of Data

We moved from a traditional hiring process of annotators to…crowdsourcing

"...a business practice that means literally to outsource an activity to the crowd"

Howe J., The rise of crowdsourcing, Wired 14(6) (2006)

- Decomposing generation and labelling efforts into everyday micro-tasks
- Opportunity for a wider set of people to potentially contribute to a dataset
- Connects data requesters to a pool of contributors
- Human-in-the-loop model evolved and created its own economy

Towards Unbiased Quality Data

Crowdsourcing

Crowd

The range of contributors will shape the data collected

+

Knowledge

Known contributors can be targeted to optimize data diversity:

- Languages (Native, Fluent, Basic)
- Location
- Age
- Gender
- Device

Less bias

Controlled fairness

(for the target user base)

Towards Unbiased Quality Data

Metadata

Data is no longer an obscure unit of information, but rather becomes contextualized with the process and the agents who contributed to it.

Towards Unbiased Quality Data

Quality Control

Contributor

Quality checks must verify people are who they say they are

- Language & job specific tests
- Real time audits
- Anti-spam features
- Reputation system

+

Data Quality

Use real time audits and gold sets to ensure accuracy

- Automatic spam detection
- Validation steps for data collections
- Gold sets
- Inter-annotator agreement

+

Task

Healthy obsession for a better UX

- Intuitive UI & contributor instructions
- Localization
- Clear payment information

=

Less bias

Quality in the dataset leads to quality in the model

Ethics in Data Collection

“Can we foresee a future crowd workplace in which we would want our children to participate?”

(A. Kittur et al. 2013)

Microtasks : An ethical economy?

- Flexible
- Affordable
- Source of human labour

- Subject to abuse -
- Unfair payments
- Explores vulnerabilities (power imbalance)
- Lacks proper regulation

Reason	% Choosing	Average Ranking
Work from home	94.6 [93.2, 95.7]	2.12
Flexible hours	91.4 [89.7, 92.8]	2.70
No commute	72.1 [69.5, 74.4]	5.05
Breaks between work	67.3 [64.6, 69.8]	5.30
No boss, supervisor, etc.	66.2 [63.6, 68.8]	4.91
No dress code	63.8 [61.1, 66.4]	6.30
Work from wherever	61.1 [58.4, 63.7]	5.67
Less stressful	57.7 [55.0, 60.4]	5.80
Learn interesting things	45.9 [43.2, 48.6]	6.45
Flexibility for family	43.4 [40.7, 46.1]	4.32
Immediate pay	35.6 [33.0, 38.2]	5.79
Make more money	25.4 [23.0, 27.9]	5.36
Physical and mental constraints make it hard to find alternative work	14.1 [12.3, 16.1]	4.25

Moss, A. J., Rosenzweig, C., Robinson, J., Jaffe, S. N., & Litman, L. (2020, April 28). Is it Ethical to Use Mechanical Turk for Behavioral Research? Relevant Data from a Representative Survey of MTurk Participants and Wages. https://doi.org/10.31234/osf.io/jbc9d

PAPA Framework (Mason, 1986)

Issues society would likely face in the information age

P

Privacy

ability of the individual to personally control information about oneself

A

Accuracy

extent to which data are correct, reliable and certified

P

Property

bundle of rights to exclusive use, to sell, or to generate income

A

Accessibility

parameters that influence human functioning in the environment

Ethics in Collection

Privacy

What type of data is being collected? how it is used and what is the end goal / purpose?

The right to be forgotten

Anonymization vs sharing PII

Ensures

1. **Compliance**
2. **Rights**
3. **Prevent breaches**
4. **Fraud**

Ethics in Collection
Property

Look at crowd as a collective good that needs to be appropriately compensated.

Watch out for flawed reputation systems which force need to accept large amounts of unfair work.

Ensures

1. **Crowd Growth**
2. **Crowd Satisfaction**
3. **Throughput**
4. **Brand**

Ethics in Collection
Accessibility

Crowd selection and data analysis based on factual information (in contrast to social constructs)

Voice
Gender > Pitch (& others)

Image
Ethnicity > Skin tone (& others)

Open Access
Explicit Requirements
Equal opportunity
Healthy reputation system

Access to ALL

ensures

1. **Uniform distribution of data in problem-relevant variables**
2. **Transparency**
3. **Metadata Quality**
4. **Brand**

Building a sustainable and ethical data economy

Compensation

Fair payment

- Time-based
- Skill-based
- Payment readjustment
- Bonus payment
- Timely manner
- Delivery detachment

Building a sustainable and ethical data economy

Compensation

Fair payment

- Time-based
- Skill-based
- Payment readjustment
- Bonus payment
- Timely manner
- Delivery detachment

Judgement

Ethical Judgment

- Source of poor performance
- Fast Judgement
- Clear Communication
- Avoid generalizations when using automation
- Feedback loop when assessment is automated

Building a sustainable and ethical data economy

Compensation
Fair payment

- Time-based
- Skill-based
- Payment readjustment
- Bonus payment
- Timely manner
- Delivery detachment

Judgement
Ethical Judgment

- Source of poor performance
- Fast Judgement
- Clear Communication
- Avoid generalizations when using automation
- Feedback loop when assessment is automated

Trust
A two-way street

- Authentication
- Authorization
- Factual dimensions & outliers
- Find verifiable proxies

Creation of Ethical Data

Using crowdsourcing

- Creating unbiased, fair, ethical data is a hard task.
- Crowdsourcing can provide a streamlined and intentional process to data creation
- Ultimately, add the capacity to Infer the culprit data points for a specific decision
  - Understanding gaps in the training data
- Details of the end system that will learn from the data are not always known,
  - e.g. user base
- Much harder with intermediaries (e.g. 3, 4 hops until it reaches the requester)

To build trust in data we need to record and share how datasets are created.

Creating the right mechanisms

Upon unfair/biased predictions we can go back to the data

- Diversity
- Inclusive and diverse community
- Access to a larger pool of people (with diverse opinions, experiences and perspectives)
- Features of the crowd -> Data -> Models
- Atomization of Cognition - how do we model the annotation/generation process
- Expert design - the contents are as important as the generation and annotation pipelines

A data unit enriched with metadata about the process and the contributing agents allows for data to become contextualized instead of an obscure unit of information, enabling the creation of unbiased data.

Knowing our dataset and its surroundings

- Data **Lineage** - keeping track of data units and knowing its origin (data traceability)
- Data **Governance** - who has access to the data, for how long, who owns it, guarantee integrity
- Data **Balancing** - dynamically adapt to contributor’s profiles, who is the user of the final system
- **Meta**data - technical specification about data units (creation and context).
    - Helps the requestor to make decisions and act upon unfairness and unethical practices
- Data **Standards** - helps to drive metadata consistency, meaning and application

Towards a Healthier Data Ecosystem

1. Goal-oriented ethical design requires thinking about outcomes: **Creation + maintenance**
2. How disparate the outcomes are for the people subject to the system? **Assessment**
3. Design considering the worst usage scenario: **What if production data is unethical?**
4. AI is everyone's business. Contribute to its sustainable and responsible advancement. **Leave no one out.**
5. Ethical **Data Stamp** (just like the sustainable products at the groceries store)
6. Data associated to its creators, with a context. **Data documented.**
7. Ethical, fair, unbiased data comes at a cost. A healthy data ecosystem needs to have these pillars at its core so that AI becomes truly ubiquitous **for all**.

How to define AI policies?

Policies for AI are still at its infancy

risks | economics | safety | lawful | trustworthy

| standards | innovation | investment | ethics

Some Examples:

- US National AI Research Resource (https://www.ai.gov/nairrtf/)
- EU AI Act (https://artificialintelligenceact.eu/) - proposed law
- Breck, Eric, et al. "The ML test score: A rubric for ML production readiness and technical debt reduction." 2017 IEEE International Conference on Big Data (Big Data). IEEE, 2017.
- IEEE Recommended Practice for Assessing the Impact of Autonomous and Intelligent Systems on Human Well-Being," in IEEE Std 7010-2020 , vol., no., pp.1-96, 1 May 2020, doi: 10.1109/IEEESTD.2020.9084219.

Example: Ethical Charter

https://bigscience.huggingface.co/blog/bigscience-ethical-charter

Intrinsic Values - “what is valuable for its own sake, in itself […], as an end”

- **Inclusivity -** non-discrimination + sense of belonging
- **Diversity -** expertise from various sources of knowledge, communities, scientific fields, and institutional contexts
- **Reproducibility -** reproduction of the research experiments and scientific conclusions
- **Openness -** process + results
- **Responsability -** social + environmental

Example: Ethical Charter

https://bigscience.huggingface.co/blog/bigscience-ethical-charter

Extrinsic Values - “what is valuable as a means, or for something else’s work”

- **Accessibility** - make outputs easily interpretable and explained

- **Transparency** - tools to interpret, monitor, explain, and make intelligible the artifacts

- **Interdisciplinary** - building bridges among linguistics, CS, law, sociology, philosophy, and other disciplines

Multilingualism - language coverage and parity

Best Practices

Fairness-centric approach to performance

Ensure models take real-world issues and user feedback into account

Lookout for bias in data

Understand input data as much as possible - representative and accurate?

Practice and promote an ethical data ecosystem

Rely on factual dimensions for collection and validation

Thank you

July 14th, 2022

Email: jfreitas@pagerduty.com