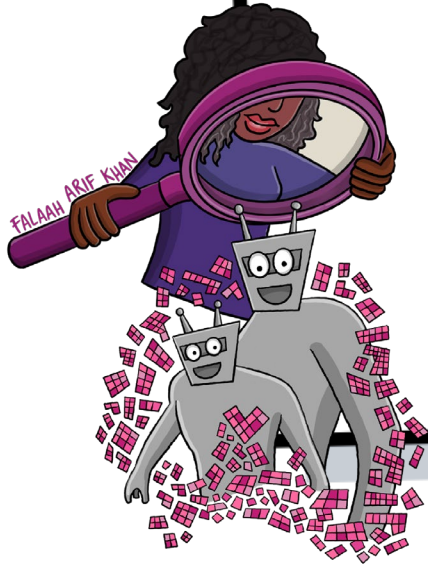


# Introduction to transparency

---

Julia Stoyanovich  
New York University





center  
for  
responsible  
ai

<http://r-ai.co>



# Responsible AI by example

## Welcome to the fastMRI Dataset



### About Us

Here at the [Center for Advanced Imaging Innovation and Research \(CA<sup>2</sup>RI](#)), in the Department of Radiology at NYU School of Medicine and NYU Langone Health, we bring people together to create new ways of seeing. We are committed to the translation of new imaging techniques and technologies into clinical practice, for the improvement of human health. In particular, we are pushing the boundaries of rapid image acquisition and advanced image reconstruction, with the aim of providing uniquely valuable biomedical information to advance the understanding of disease and improve the care of patients.

### fastMRI

We are partnering with Facebook AI Research (FAIR) on fastMRI – a collaborative research project to investigate the use of AI to make MRI scans up to 10X faster.

NYU Langone and FAIR are providing open-source AI models, baselines, and evaluation metrics.

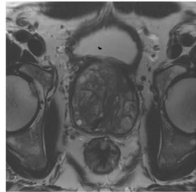
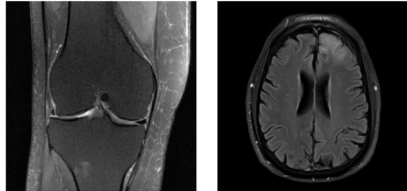
### The Dataset

The deidentified imaging dataset provided by NYU Langone comprises raw k-space data in several sub-dataset groups. Curation of these data are part of an IRB approved study. Raw and DICOM data have been deidentified via conversion to the vendor-neutral ISMRMD format and the RSNA clinical trial processor, respectively. We also performed manual inspection of each DICOM image for the presence of any unexpected protected health information (PHI), with spot checking of both metadata and image content.

**Knee MRI:** Data from more than 1,500 fully sampled knee MRIs obtained on 3 and 1.5 Tesla magnets and DICOM images from 10,000 clinical knee MRIs also obtained at 3 or 1.5 Tesla. The raw dataset includes coronal proton density-weighted images with and without fat suppression. The DICOM dataset contains coronal proton density-weighted with and without fat suppression, axial proton density-weighted with fat suppression, sagittal proton density, and sagittal T2-weighted with fat suppression. The exact distribution of contrasts is given in [table 1](#). Please note that this table does not include stats about the data that was originally held back for the fastMRI reconstruction challenge.

### Apply for Access

The application process includes acceptance of the Data Sharing Agreement (found below) and submission of an online application form. The application must include the investigator's institutional affiliation and the proposed uses of the data. NYU fastMRI data may be used for internal research or educational purposes only as described in the data use agreement and may not be redistributed in any way without prior permission. Read and agree to the data use agreement below to apply for access.

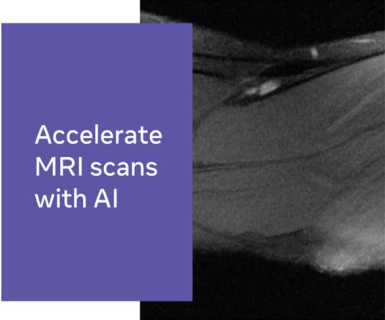


clear need for improvement  
can validate quality / usefulness  
technical readiness  
decision-maker readiness



### fastMRI

AI at Meta and NYU Langone Health demonstrate that AI can generate accurate MRIs using one-fourth of the raw data traditionally required for a full MRI.



<https://fastmri.med.nyu.edu/>

# Responsible AI by example:

# are we done?

## Welcome to the fastMRI Dataset



### About Us

Here at the [Center for Advanced Imaging Innovation and Research \(CA<sup>2</sup>RI](#)), in the Department of Radiology at NYU School of Medicine and NYU Langone Health, we bring people together to create new ways of seeing. We are committed to the translation of new imaging techniques and technologies into clinical practice, for the improvement of human health. In particular, we are pushing the boundaries of rapid image acquisition and advanced image reconstruction, with the aim of providing uniquely valuable biomedical information to advance the understanding of disease and improve the care of patients.

### fastMRI

We are partnering with Facebook AI Research (FAIR) on fastMRI – a collaborative research project to investigate the use of AI to make MRI scans up to 10X faster.

NYU Langone and FAIR are providing open-source AI models, baselines, and evaluation metrics.

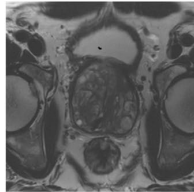
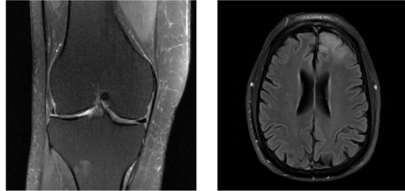
### The Dataset

The deidentified imaging dataset provided by NYU Langone comprises raw k-space data in several sub-dataset groups. Curation of these data are part of an IRB approved study. Raw and DICOM data have been deidentified via conversion to the vendor-neutral ISMRMD format and the RSNA clinical trial processor, respectively. We also performed manual inspection of each DICOM image for the presence of any unexpected protected health information (PHI), with spot checking of both metadata and image content.

**Knee MRI:** Data from more than 1,500 fully sampled knee MRIs obtained on 3 and 1.5 Tesla magnets and DICOM images from 10,000 clinical knee MRIs also obtained at 3 or 1.5 Tesla. The raw dataset includes coronal proton density-weighted images with and without fat suppression. The DICOM dataset contains coronal proton density-weighted with and without fat suppression, axial proton density-weighted with fat suppression, sagittal proton density, and sagittal T2-weighted with fat suppression. The exact distribution of contrasts is given in [table 1](#). Please note that this table does not include stats about the data that was originally held back for the fastMRI reconstruction challenge.

### Apply for Access

The application process includes acceptance of the Data Sharing Agreement (found below) and submission of an online application form. The application must include the investigator's institutional affiliation and the proposed uses of the data. NYU fastMRI data may be used for internal research or educational purposes only as described in the data use agreement and may not be redistributed in any way without prior permission. Read and agree to the data use agreement below to apply for access.

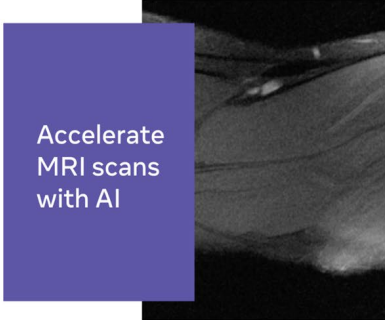


clear need for improvement  
can validate quality / usefulness  
technical readiness  
decision-maker readiness



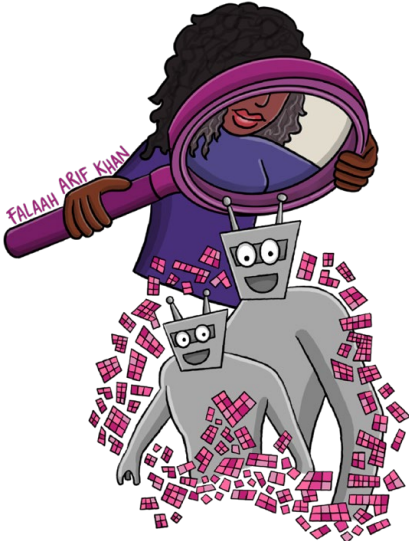
### fastMRI

AI at Meta and NYU Langone Health demonstrate that AI can generate accurate MRIs using one-fourth of the raw data traditionally required for a full MRI.

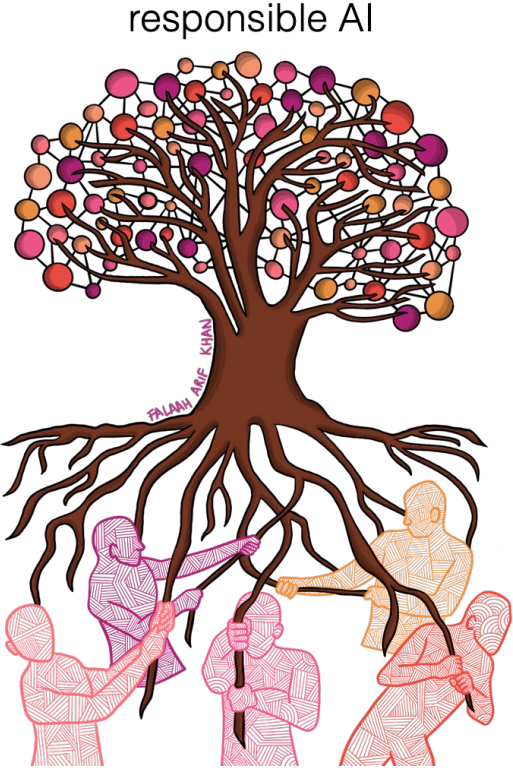


<https://fastmri.med.nyu.edu/>

# Terminology & vision



transparency, interpretability,  
explainability, intelligibility



responsible AI



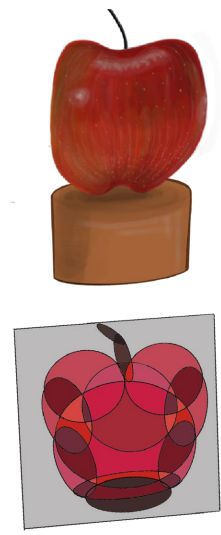
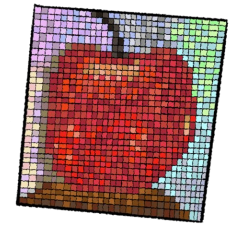
agency, responsibility



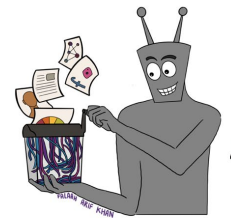
# Transparency: of what?



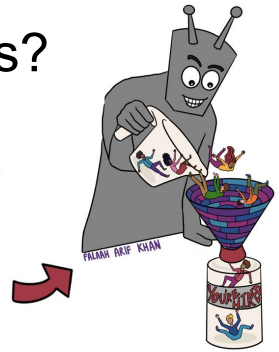
data?



model?



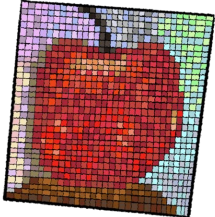
process?



# Challenge: transparency through the lifecycle



data?



model?



process?



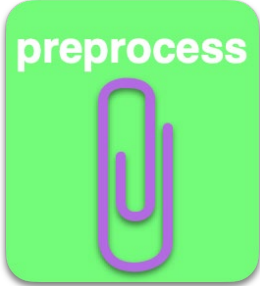
# Example: taming technical bias

**Goal:** design a model to predict appropriate level of compensation for job applicants

**Problem :** accuracy is lower for women who have more experience on the job - a **fairness concern**

demographics				

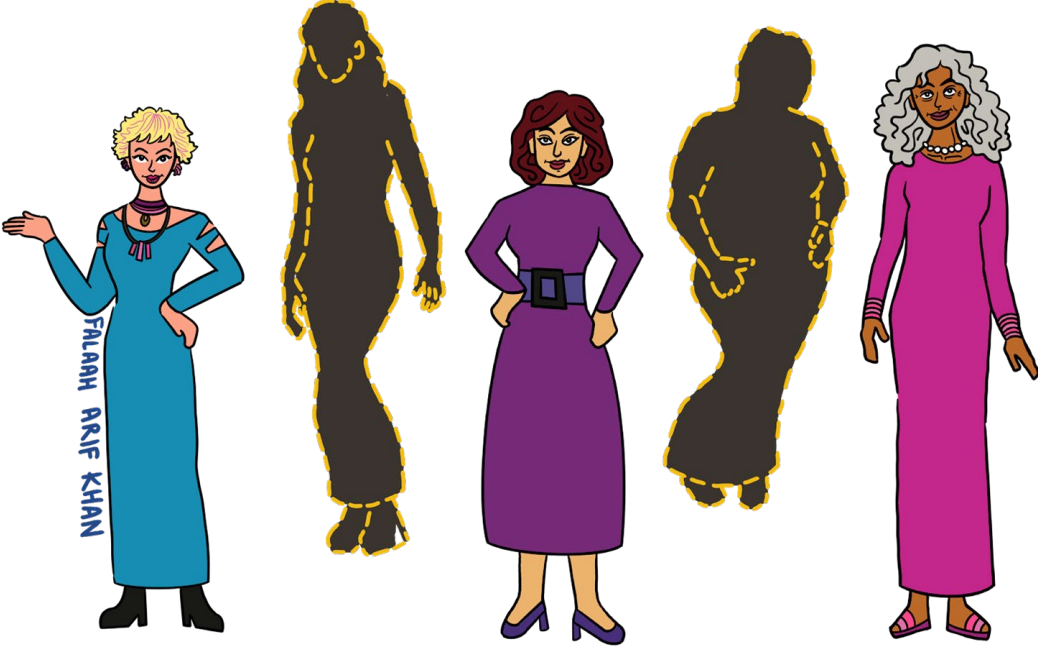
employment				



[Schelter,He,Khilmnani, Stoyanovich, 2020]



# Missing values: observed

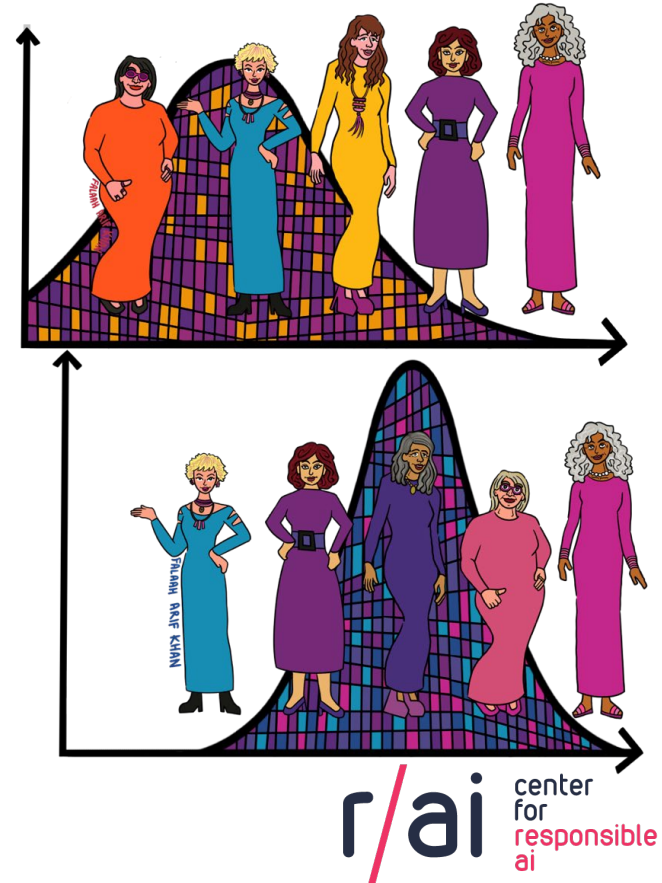


# Missing value imputation

are values **missing at random** (e.g., *gender, age, years of experience, disability status* on job applications)?

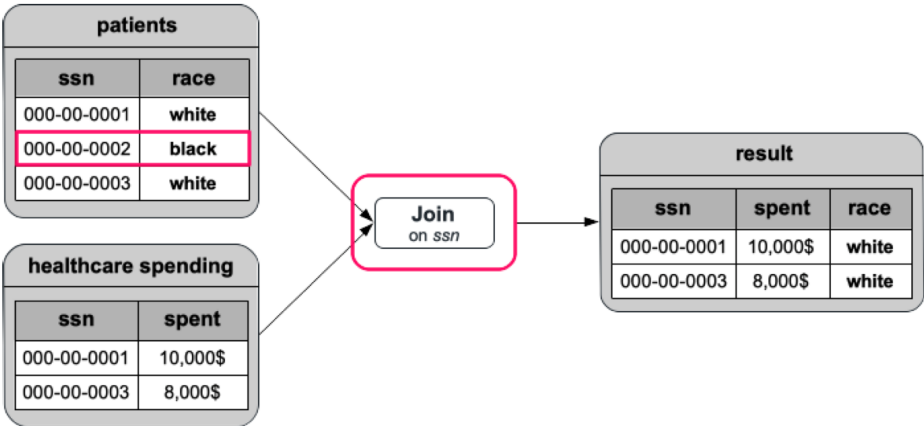
are we ever interpolating **rare categories** (e.g., *Native American*)

are **all categories** represented (e.g., *non-binary gender*)?



# Data filtering

operations like **selection** and **join**, can arbitrarily change demographic group proportions



age_group	county
60	CountyA
60	CountyA
20	CountyA
60	CountyB
20	CountyB
20	CountyB

50% vs 50%



age_group	county
60	CountyA
60	CountyA
20	CountyA

66% vs 33%

# Data distribution labels

**Potential issues in preprocessing pipeline:**

- 1 Join might change proportions of groups in data
- 2 Column 'age\_group' projected out, but required for fairness
- 3 Selection might change proportions of groups in data
- 4 Imputation might change proportions of groups in data
- 5 'race' as a feature might be illegal!
- 6 Embedding vectors may not be available for rare names!

**Python script for preprocessing, written exclusively with native pandas and sklearn constructs**

```
# load input data sources, join to single table
patients = pandas.read_csv(...)
histories = pandas.read_csv(...)
data = pandas.merge([patients, histories], on=['ssn'])

# compute mean complications per age group, append as column
complications = data.groupby('age_group')
    .agg(mean_complications=('complications', 'mean'))
data = data.merge(complications, on=['age_group'])

# Target variable: people with frequent complications
data['label'] = data['complications'] >
    1.2 * data['mean_complications']

# Project data to subset of attributes, filter by counties
data = data[['smoker', 'last_name', 'county',
            'num_children', 'race', 'income', 'label']]
data = data[data['county'].isin(counties_of_interest)]

# Define a nested feature encoding pipeline for the data
impute_and_encode = sklearn.Pipeline([
    (sklearn.SimpleImputer(strategy='most_frequent')),
    (sklearn.OneHotEncoder()))]
featurisation = sklearn.ColumnTransformer(transformers=[
    (impute_and_encode, ['smoker', 'county', 'race']),
    (Word2VecTransformer(), 'last_name')
    (sklearn.StandardScaler(), ['num_children', 'income'])]

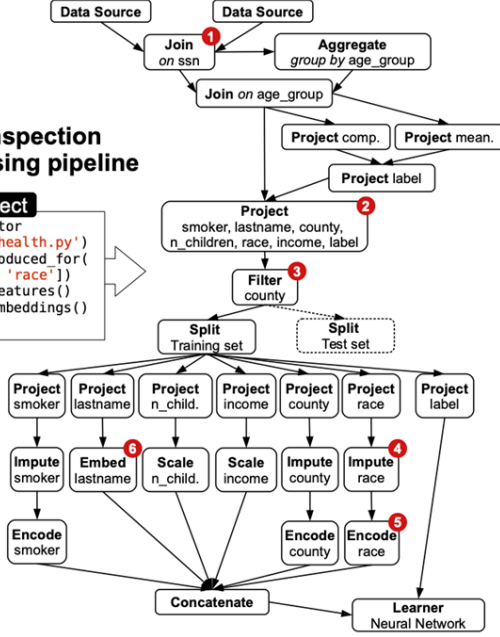
# Define the training pipeline for the model
neural_net = sklearn.KerasClassifier(build_fn=create_model())
pipeline = sklearn.Pipeline([
    ('features', featurisation),
    ('learning_algorithm', neural_net)])

# Train-test split, model training and evaluation
train_data, test_data = train_test_split(data)
model = pipeline.fit(train_data, train_data.label)
print(model.score(test_data, test_data.label))
```

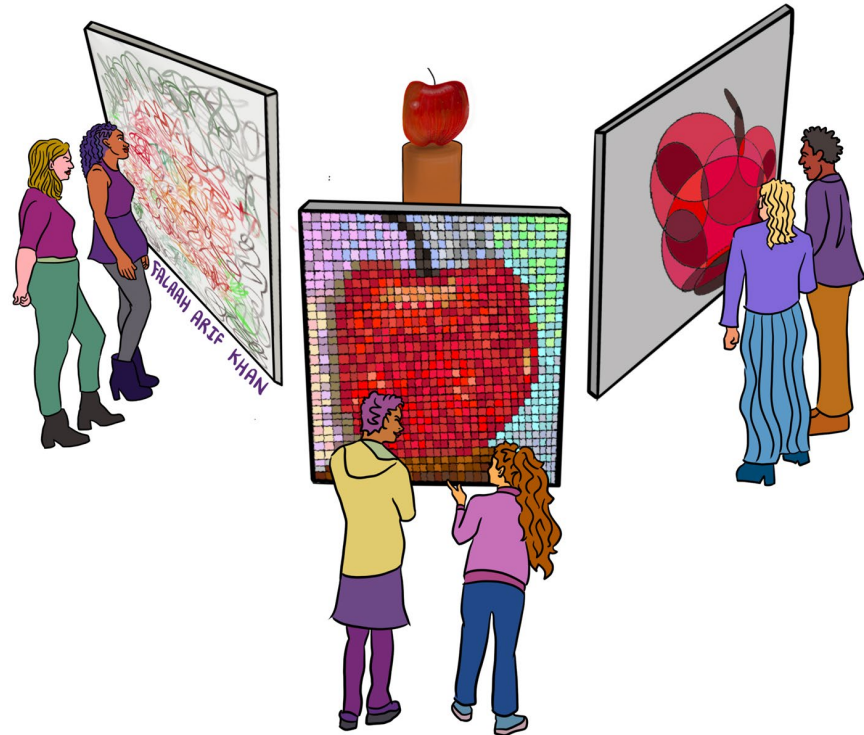
**Corresponding dataflow DAG for instrumentation, extracted by *minspect***

**Declarative inspection of preprocessing pipeline**

```
minspect
PipelineInspector
.on_pipeline('health.py')
.no_bias_introduced_for(
    ['age_group', 'race'])
.no_illegal_features()
.no_missing_embeddings()
.verify()
```



# Transparency: for whom and why?



To whom are we explaining?  
Why / for what purpose?  
Are explanations effective?



# The “nutritional labels” metaphor

**SIDE-BY-SIDE COMPARISON**

Original Label



New Label



**Note:** The images above are meant for illustrative purposes to show how the new Nutrition Facts label might look compared to the old label. Both labels represent fictional products. When the original hypothetical label was developed in 2014 (the image on the left-hand side), added sugars was not yet proposed so the “original” label shows 1g of sugar as an example. The image created for the “new” label (shown on the right-hand side) lists 12g total sugar and 10g added sugar to give an example of how added sugars would be broken out with a % Daily Value.

An example of the old nutrition labels, left, and the new one. The new nutrition labels will display calories and serving size more prominently, and include added sugars for the first time.  
PHOTO: FOOD AND DRUG ADMINISTRATION/ASSOCIATED PRESS

<https://www.wsj.com/articles/why-the-nutritional-labels-on-your-food-are-changing-or>

## ACCOUNTANT

### Acme Partners

---

**Qualifications:** BS in accounting, GPA >3.0, Knowledge of financial and accounting systems and applications

---

**Personal data to be analyzed:** An AI program could be used to review and analyze the applicant’s personal data online, including LinkedIn profile, social media accounts and credit score.

---

**Additional assessment:** AI-assisted personality scoring

**ALERT:** Applicants for this position DO NOT have the option to selectively decline use of AI analysis for any of their personal data or to review and challenge the results of such analysis.

<https://www.wsj.com/articles/hiring-job-candidates-ai-11632244313>

## Security & Privacy Overview

# Smart Device Co.

Smart Video Doorbell NS200  
Firmware version: 2.5.1 - updated on: 11/12/2020  
The device was manufactured in: China

**Security Mechanisms**

- Security updates:** Automatic - Available until at least 1/1/2022
- Access control:** Password - Factory default - User changeable, Multi-factor authentication, Multiple user accounts are allowed

**1**

---

**Data Practices**

Sensor data collection	Visual	Audio	Physiological	Location
<b>Sensor type</b>	Camera	Microphone		
<b>Purpose</b>	Providing device functions	Providing device functions, Research		
<b>Data stored on device</b>	Identified	No device storage		
<b>Data stored on cloud</b>	Identified	Identified - Option to delete		
<b>Shared with</b>	Manufacturer, Government	Manufacturer		
<b>Sold to</b>	Not disclosed	Not sold		

**Other collected data:** Motion, Account info, Payment info, Contact info, Device setup info, Device tech info, Device usage info

**Privacy policy:** [www.NS200.smartdeviceco.com/policy](http://www.NS200.smartdeviceco.com/policy)

**3**


---

**More Information**

**Detailed Security & Privacy Label:**  
[www.iotsecurityprivacy.org/labels](http://www.iotsecurityprivacy.org/labels)

CMU IoT Security and Privacy Label **CISPL 1.0** [iotsecurityprivacy.org](http://iotsecurityprivacy.org)

**4**

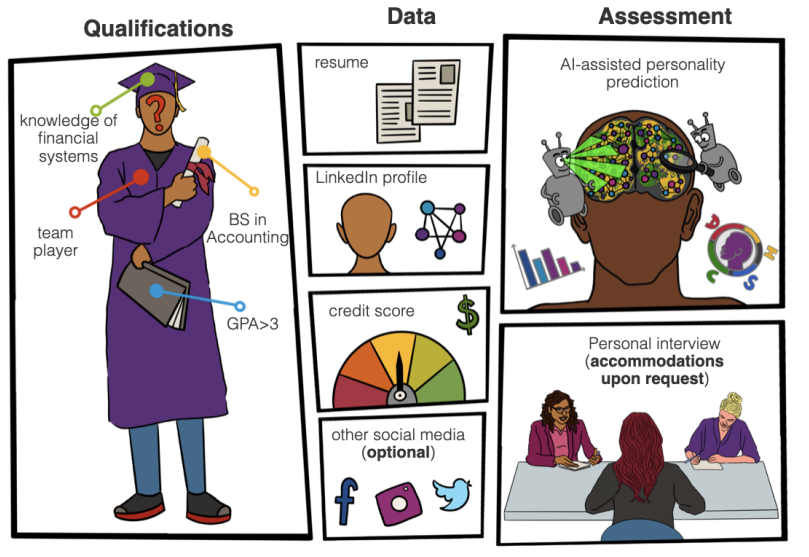


**PUBLIC DOMAIN**

<https://www.wsj.com/articles/imagine-a-nutrition-label-for->



# Nutritional labels for data & models



**comprehensible** : short, simple, clear  
**consultative** : providing actionable info  
**comparable** : implying a standard

also....

incrementally **computable** , as a by-product of multi-step computation

[Stoyanovich & Howe 2019]

# Challenge: transparency that “works”

## AI & JOURNALISM

### Workshop: Algorithmic Transparency

JULIA STOYANOVICH, Director, NYU Center for Responsible AI  
ANDREW BELL, Fellow, NYU Center for Responsible AI

SEPTEMBER 19TH, 11AM ET, ZOOM

**We are AI**  
taking control of technology  
powered by NYU Center for Responsible AI

r/ai center for responsible ai

QUEENS PUBLIC LIBRARY P2PU

#### NYU CENTER FOR RESPONSIBLE AI PRESENTS

### THE ALGORITHMIC TRANSPARENCY WORKSHOP

Join us for a workshop on algorithmic transparency from the NYU Center for Responsible AI  
**Date:** December 12, 2023 (Tuesday) from 12:00 pm - 2:00 pm  
**Venue:** NYU Tandon Future Labs (7th floor, 370 Jay Street, Brooklyn, NY 11201)  
*Free lunch for participants will be served beginning at 11:30 am.*



#### Course Instructors

The workshop will be co-taught by **Andrew Bell**, fellow at the NYU Center for Responsible AI (RAI) and **Julia Stoyanovich**, Director of NYU RAI.



#### Andrew Bell

Andrew Bell is a Computer Science Ph.D. Candidate being co-advised by Prof. Julia Stoyanovich and Dr. Oded Nov. He is a recipient of the National Science Foundation/Graduate Research Fellowship (NSF GRFP). His research interests lie at the intersection of machine learning and public policy and are more narrowly focused on the fairness and explainability of algorithmic decision systems. In Spring 2023, Andrew was a visiting research fellow at the Center for AI (CENTA) in Turin, Italy.



#### Julia Stoyanovich

Dr. Julia Stoyanovich is Associate Professor of Computer Science & Engineering and of Data Science, and Director of the Center for Responsible AI at NYU. Her goal is to make “responsibility” synonymous with “AI”. Julia has co-authored over 100 academic publications, and has written for the New York Times, the Wall Street Journal and Le Monde. She engages in technology policy, has been teaching responsible AI to students, practitioners and the public, and has co-authored comic books on the topic. She received her Ph.D. in Computer Science from Columbia University.

#### About the NYU Center for Responsible AI

The NYU Center for Responsible AI aims to make responsible AI synonymous with AI. The center, which is made up of over 15 researchers across a broad range of fields, conducts interdisciplinary research, engages in AI policy and regulation, and aims to teach different audiences about AI and its social impact. More information can be found at <https://airesponsibly.com/>



2023  
**ALL ABOARD!**  
**MAKING AI EDUCATION**  
**ACCESSIBLE**

Authors:  
 Falaah Arif Khan, Lucius Bynum, Amy Hurst, Lucas Rosenblatt,  
 Meghana Shanbhogue, Mona Sloane, Julia Stoyanovich

NYU CENTER FOR RESPONSIBLE AI PRESENTS

### WHAT IS RESPONSIBLE AI & HOW DOES IT APPLY TO YOUR WORK AT META?

**Responsible AI** is the science and the practice of making the design, development, and use of AI socially sustainable. In this workshop, the **NYU Center for Responsible AI** will give an overview of the discipline, and will present and extensively discuss concrete Meta-relevant case studies on ad targeting, content ranking, machine translation, algorithmic hiring, and more. We will take deep dives into algorithmic fairness, transparency and interpretability, and privacy and data protection, while keeping the conversation relevant to your work at Meta. No prior knowledge of responsible AI concepts or techniques is required, or even expected, for participation. Whether you are a responsible AI aficionado or a skeptic, and whether or not your role at Meta is technical — this workshop is for you!

This workshop will consist of two 120-minute sessions. When signing up for the workshop, you commit to attending both sessions and participating in the discussion. Note that the NYU team is conducting an educational research study in conjunction with the workshop, to learn more about effective ways to teach responsible AI concepts and techniques to industry practitioners. They welcome your participation in the workshop even if you do not wish to participate in the research study. You will have an opportunity to find out more about the research and opt into the study at the start of the workshop.

The workshop will be presented by **Dr. Julia Stoyanovich**, Institute Associate Professor of Computer Science and Engineering, Associate Professor of Data Science, and Director of the Center for Responsible AI at NYU. Please reach out to her at [stoyanovich@nyu.edu](mailto:stoyanovich@nyu.edu) if you have any questions about the workshop.

NYU Center for Responsible AI aims to make responsible AI synonymous with AI. We engage in basic and applied research, participate in technology policy & regulation efforts, and teach responsible AI to data science students, practitioners, and members of the public. For more information, visit <https://airesponsibly.com>.



# References & acknowledgements

Responsible Data Management , Communications of the ACM, 2022 <https://doi.org/10.1145/3488717>

Data Distribution Debugging in Machine Learning Pipelines , VLDB Journal, 2022 <https://doi.org/10.1007/s00778-021-00726-w>

The Algorithmic Transparency Playbook , NYU Center for Responsible AI, 2023, <https://r-ai.co/transparency-playbook>

Nutritional Labels for Data and Models , IEEE Data Engineering Bulletin, 2019 <http://sites.computer.org/debull/A19sept/p13.pdf>



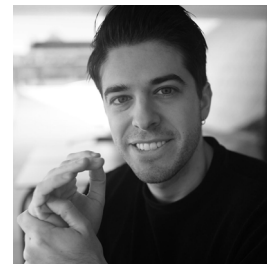
HV Jagadish



Bill Howe



Oded Nov



Andrew Bell



Supported in part by NSF Grants No. 1934464, 1926250, 1916647, 1916505, 2312930



# ALGORITHMIC TRANSPARENCY PLAYBOOK

A stakeholder-first approach to creating  
transparency for your organization's algorithms.

Andrew Bell  
Julia Stoyanovich, Ph.D.  
Oded Nov, Ph.D.

# Thank you! Questions?

---

Julia Stoyanovich  
New York University, USA