Human-Centered Data Science

Establishing critical reflective practice in the development of data-driven software

Claudia Müller-Birn

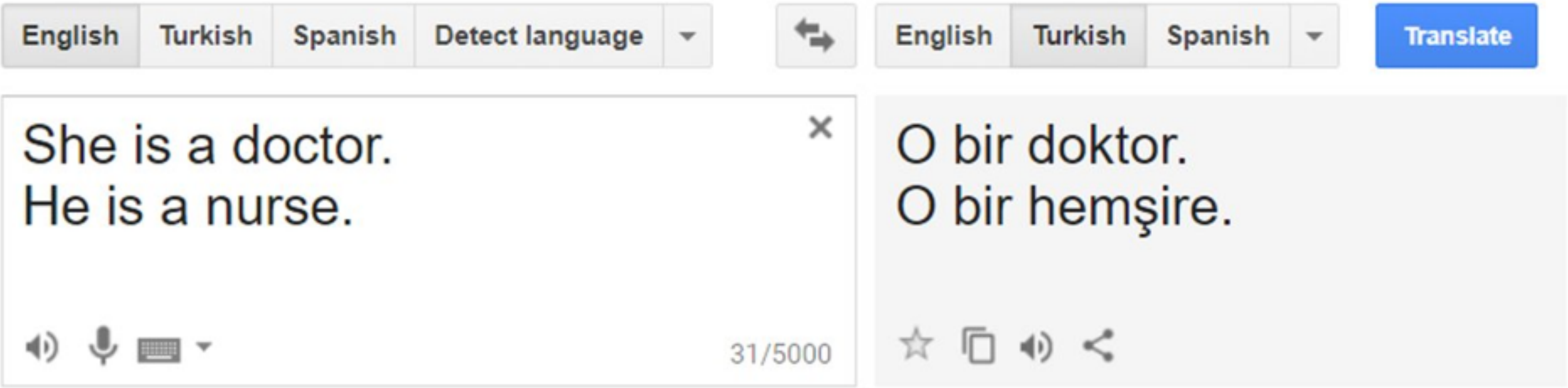
Human-Centered Computing, Institute of Computer Science, Freie Universität Berlin

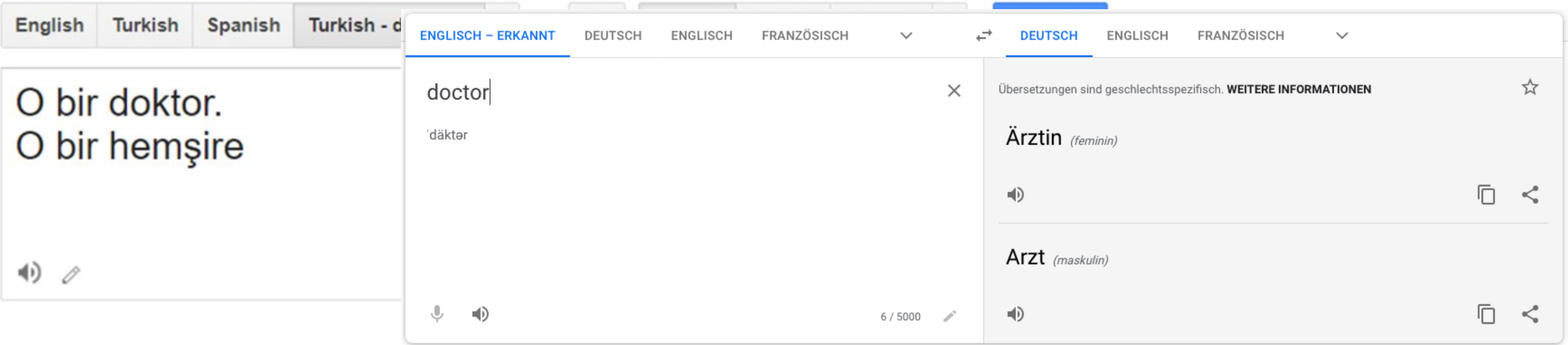
November 23, 2022

Overview

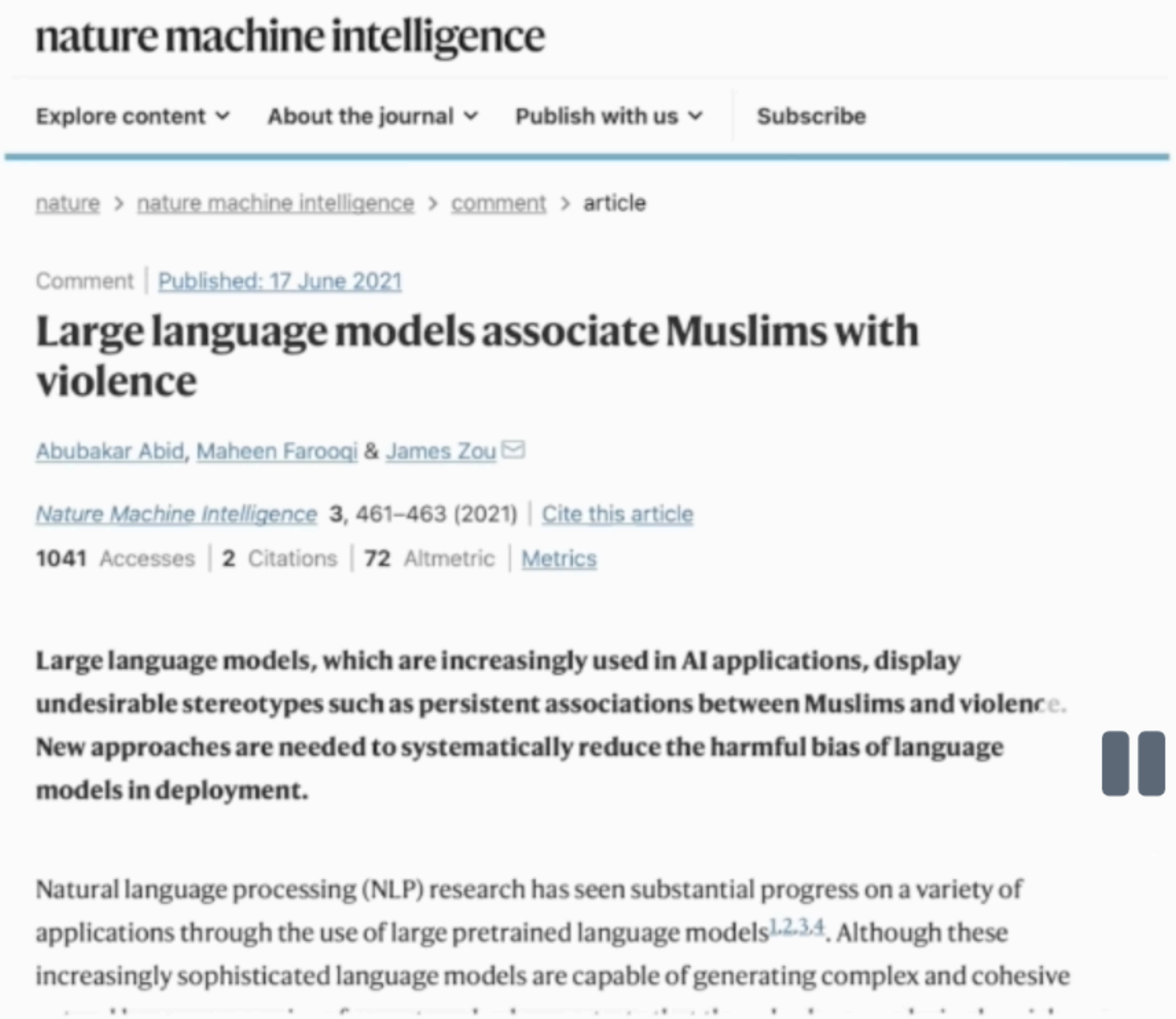
- **Increasing Societal Impact of Data-Driven Software**
  (examples with a focus on the AMS system)
- **The Need for Human-Centered Data Science**
  (definition, critical reflective practice, bias)
- **Mitigating Bias by Understanding Documentation As Reflexive Practice**
  (origins of data, datasheets of data sets)

Text Data


English
Turkish
Spanish
Detect language
English
Turkish
Spanish
Translate
She is a doctor.
He is a nurse.
31/5000
O bir doktor.
O bir hemşire.


English
Turkish
Spanish
O bir doktor.
O bir hemşire
ENGLISCH – ERKANNT
DEUTSCH
ENGLISCH
FRANZÖSISCH
DEUTSCH
ENGLISCH
FRANZÖSISCH
doctor
ˈdäktər
6 / 5000
Übersetzungen sind geschlechtsspezifisch. WEITERE INFORMATIONEN
Ärztin (feminin)
Arzt (maskulin)

Language Model


nature machine intelligence

Explore content | About the journal | Publish with us | Subscribe

nature > nature machine intelligence > comment > article

Comment | Published: 17 June 2021

Large language models associate Muslims with violence

Abubakar Abid, Maheen Farooqi & James Zou

Nature Machine Intelligence 3, 461–463 (2021) | Cite this article

1041 Accesses | 2 Citations | 72 Altmetric | Metrics

Large language models, which are increasingly used in AI applications, display undesirable stereotypes such as persistent associations between Muslims and violence. New approaches are needed to systematically reduce the harmful bias of language models in deployment.

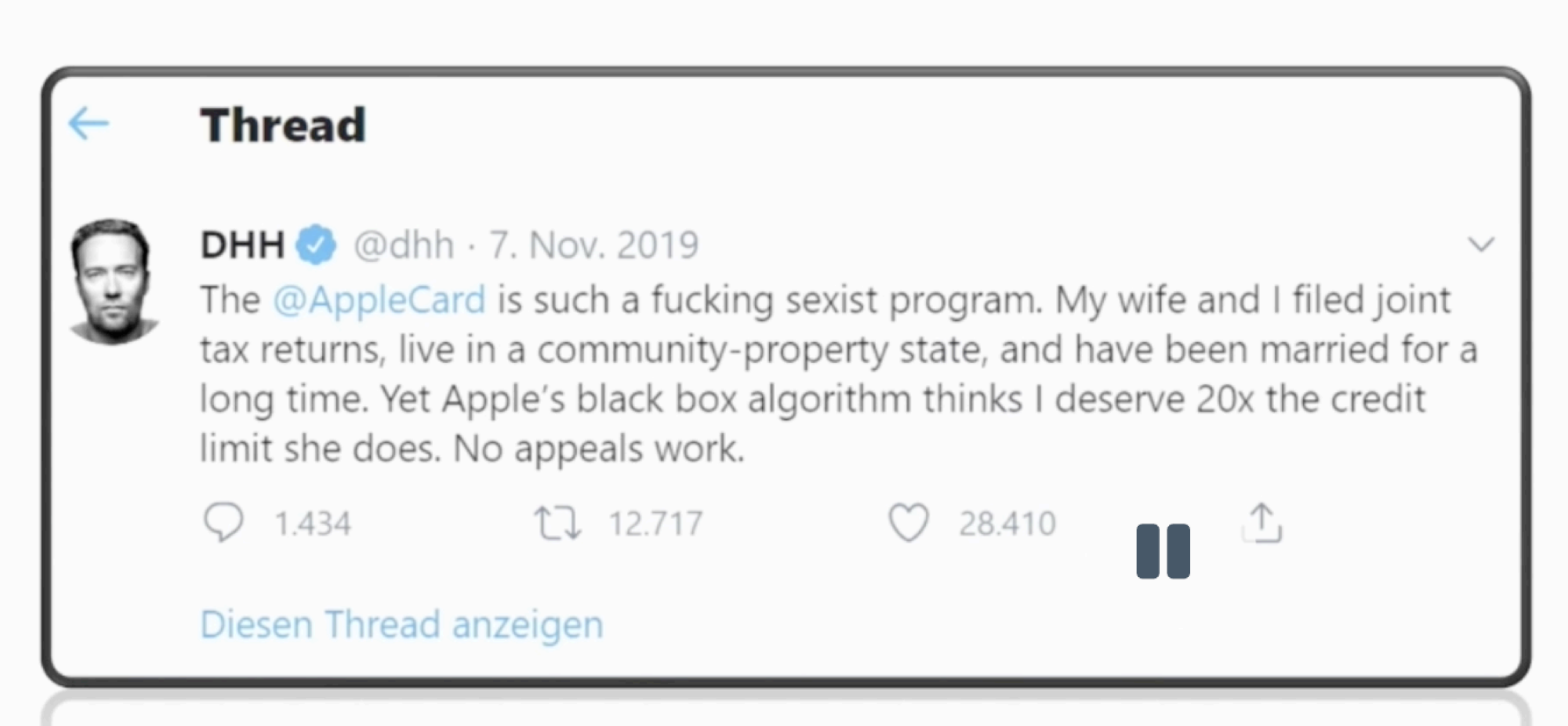
Natural language processing (NLP) research has seen substantial progress on a variety of applications through the use of large pretrained language models[1,2,3,4]. Although these increasingly sophisticated language models are capable of generating complex and cohesive

Quelle: https://www.nature.com/articles/s42256-021-00359-2

Image Data


diri noir avec banan
@jackyalcine
Following
Google Photos, y'all [redacted] up. My friend's not a gorilla.
Skyscrapers
Airplanes
Cars
Bikes
Gorillas
Graduation
© Twitter – @jackyalcine

Finance Data



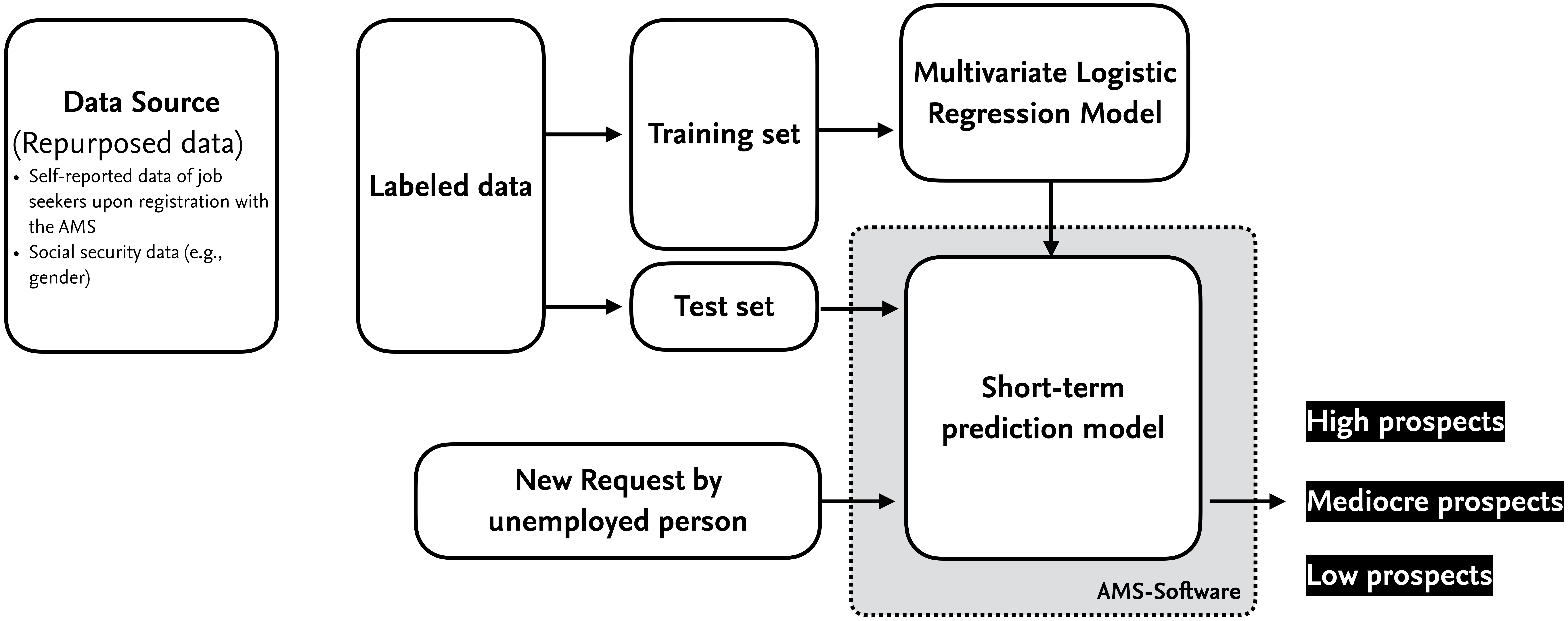
Public Unemployment Data

AMS
Arbeitsmarktservice
Österreich

Photo by Amy Hirschi on Unsplash

"Entwicklung eines Modells zur Prognose der regionalspezifischen Arbeitsmarkt-Integrationschancen von vorgemerkten Arbeitslosen."

Model Building Process



Insights taken from
Allhutter, D., Cech, F., 0007, F. F., Grill, G., & Mager, A. (2020). Algorithmic Profiling of Job Seekers in Austria - How Austerity Politics Are Made Effective. Frontiers Big Data, 3, 326.

Used Features

BE_INT

= f (0,10

– 0,14 x GENDER_FEMALE

– 0,13 x AGE-GROUP_30_49

– 0,70 x AGE-GROUP_50_PLUS

+ 0,16 x STATE_GROUP_EU

– 0,05 x STATE_GROUP_THIRD

+ 0,28 x EDUCATION_APPRENTICESHIP

+ 0,01 x EDUCATION_MATURA_PLUS

– 0,15 x CARE_TAKING

– 0,34 x LIVING_TYP_2

– 0,18 x LIVING_TYP_3

– 0,83 x LIVING_TYP_4

– 0,82 x LIVING_TYP_5

…

…

– 0,67 x IMPAIRED

+ 0,17 x OCCUPATION_PRODUCTION

– 0,74 x OCCUPATION_DAYS_LITTLE

+ 0,65 x FREQUENCY_CASE_1

+ 1,19 x FREQUENCY_CASE_2

+ 1,98 x FREQUENCY_CASE_3_PLUS

– 0,80 x CASE_LONG

– 0,57 x MN_PARTICIPATION_1

– 0,21 x MN_PARTICIPATION_2

– 0,43 x MN_PARTICIPATION_3)

Translated features from http://www.forschungsnetzwerk.at/downloadpub/arbeitsmarktchancen_methode_%20dokumentation.pdf

Der AMS-Algorithmus ist ein „Paradebeispiel für Diskriminierung“

The AMS algorithm is a prime example of discrimination

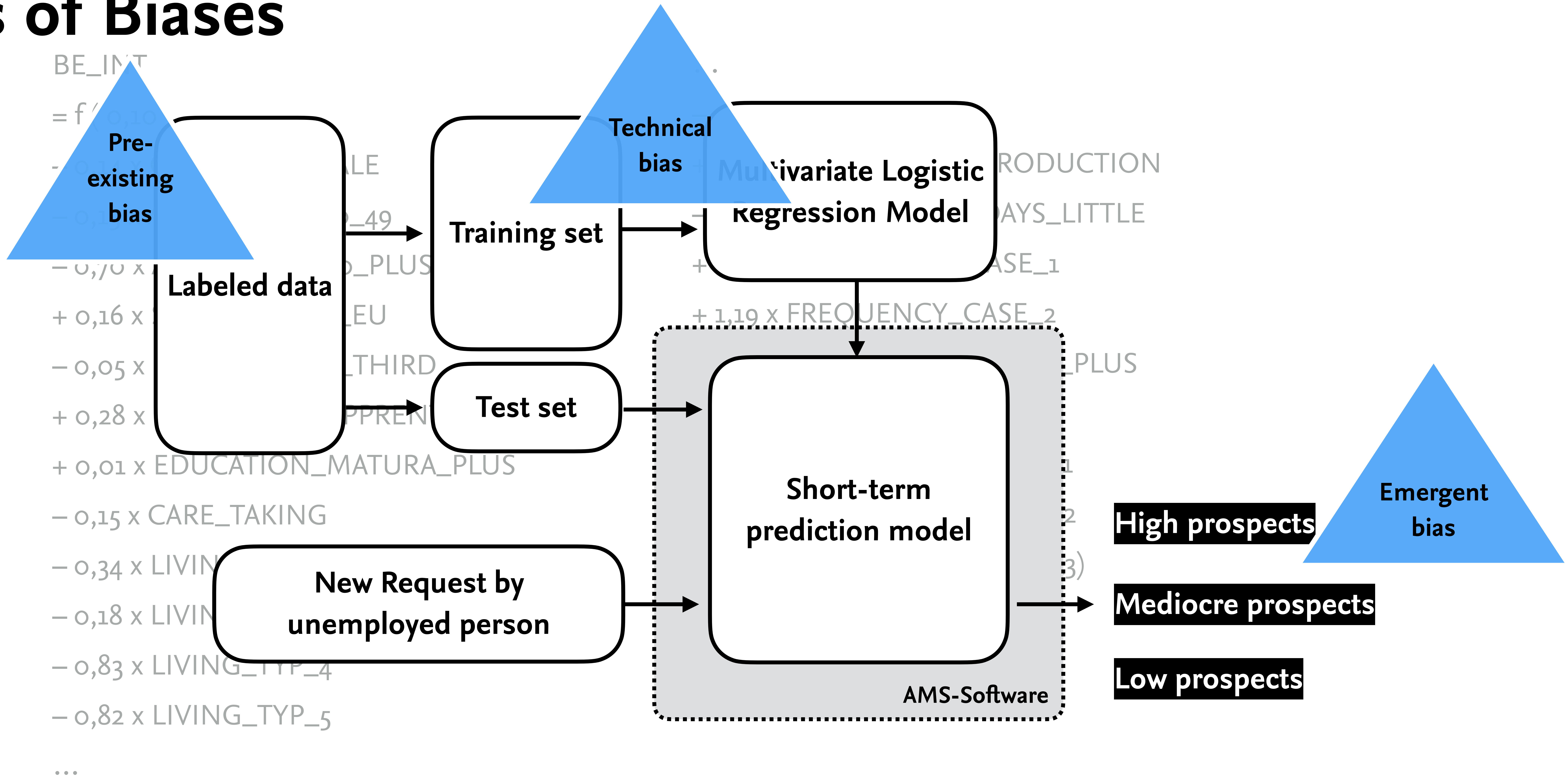
Der AMS-Algorithmus ist ein „Paradebeispiel für Diskriminierung“
Ein Computerprogramm soll ab 2019 die Arbeitsmarktchancen von Arbeitslosen berechnen. Experten von der TU und WU Wien schlagen Alarm.
futurezone.at

Austria's employment agency is rolling out a sorting algorithm that gives lower points to women and the disabled, in the name of efficiency.

A textbook example of automated – and possibly illegal – discrimination.

Austria's employment agency rolls out discriminatory algorithm, sees no pro...
AMS, Austria's employment agency, is about to roll out a sorting algorithm that gives lower scores to women and to the disabled. It is very likely illegal under...

Sources of Biases



Sources of Bias: Friedman, B., & Nissenbaum, H. (1996). Bias in computer systems. ACM Trans. Inf. Syst., 14(3), 330–347.

"technology's interaction with the social ecology is such that technical developments frequently have environmental, social, and human consequences that go far beyond the immediate purposes of the technical devices and practices themselves, and the same technology can have quite different results when introduced into different contexts or under different circumstances."

Kranzberg's First Law: "Technology is neither good nor bad; nor is it neutral."

Kranzberg, M. (1986). Technology and History: "Kranzberg's Laws." Technology and Culture, 27(3), 544–560.

Overview

- **Increasing Societal Impact of Data-Driven Software**
  (examples with a focus on the AMS system)

- **The Need for Human-Centered Data Science**
  (definition, critical reflective practice, bias)

- **Mitigating Bias by Understanding Documentation As Reflexive Practice**
  (origins of data, datasheets of data sets)

„

Human-Centered Data Science (HCDS) draws on the well-established traditions of human-centered design to inform a [responsible] data science practice.

HCDS pushes computational approaches to large-scale data to include the kind of rich detail, contextual knowledge, and deep understanding that qualitative research and mixed methods can bring to the understanding of data and society.

Aragon, Cecilia, Shion Guha, Marina Kogan, Michael Muller, und Gina Neff. Human-Centered Data Science: An Introduction. Cambridge, MA, USA: MIT Press, 2022.

Critical Reflective Practice of Human-Centered Data Science

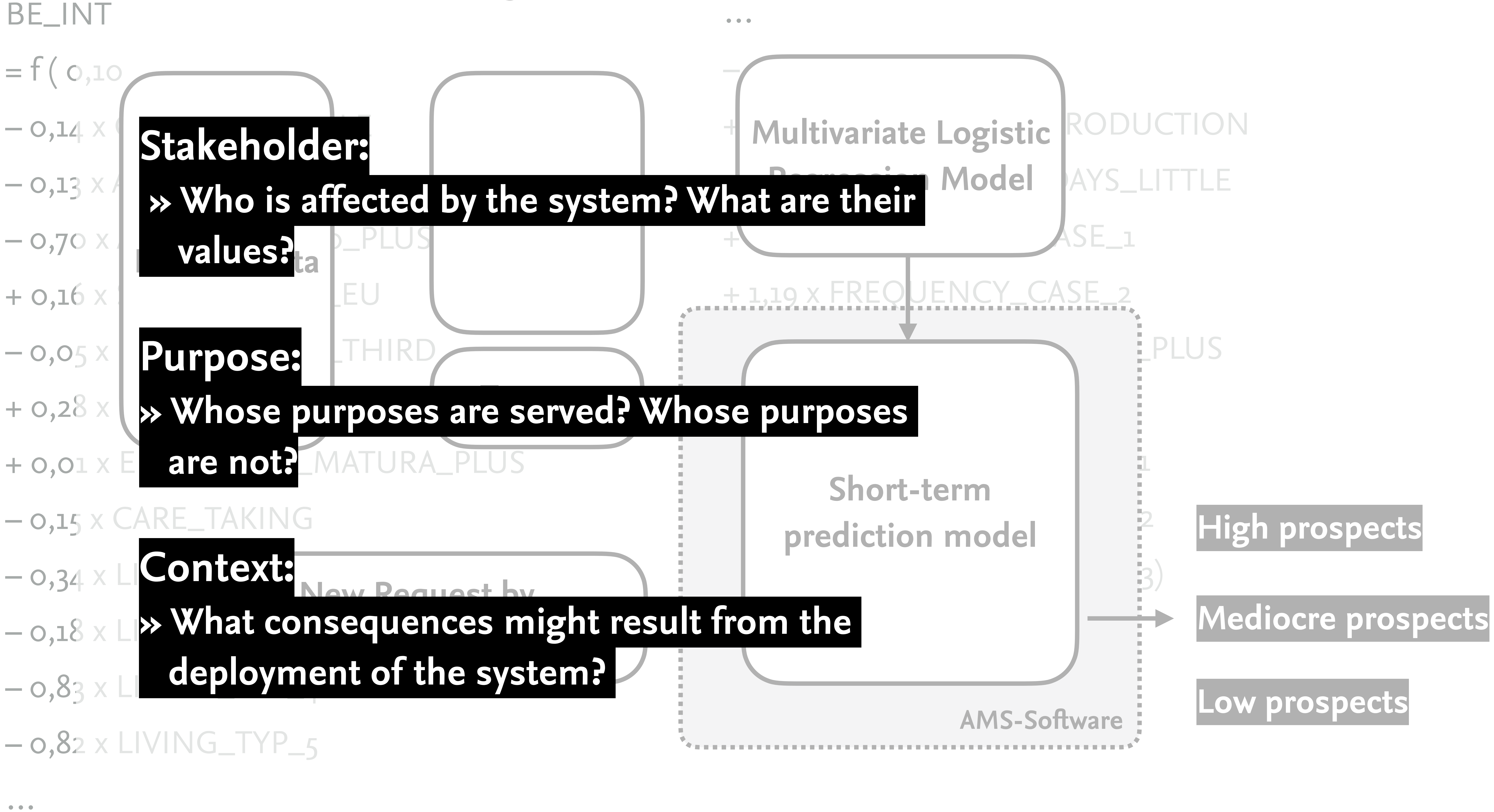
Barocas and Boyd emphasize that an “*ethical deliberation should be embedded in the everyday work of scientists.*”

Frauenberger et al. proposes, thus, an *in-action ethics* approach that emphasizes *ethos.*

Shilton suggests *value levers* that enable that values discussions occur without explicit intervention within technology design settings.

Solon Barocas and Danah Boyd. 2017. Engaging the ethics of data science in practice. Commun. ACM 60, 11 (Oct 2017), 23–25.
Christopher Frauenberger, Marjo Rauhala, and Geraldine Fitzpatrick. 2017. In-Action Ethics. Interacting with Computers 29, 2 (Mar 2017), 220–236.
Shilton, Katie. 2018. „Values and Ethics in Human-Computer Interaction“. Foundations and Trends® Human–Computer Interaction 12(2):107–71.

A Human-Centered Design Perspective on the AMS System



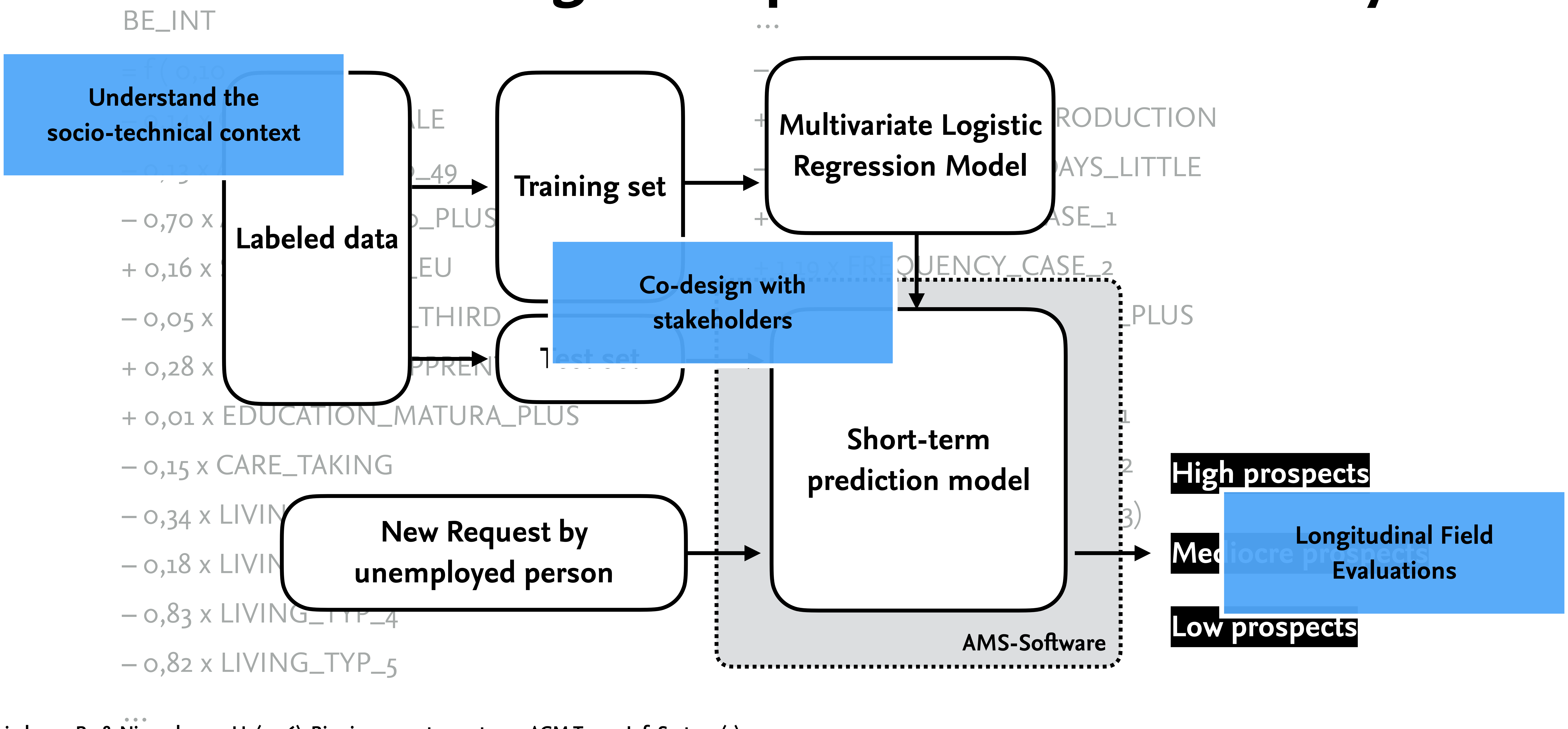
Stakeholder:
» Who is affected by the system? What are their values?

Purpose:
» Whose purposes are served? Whose purposes are not?

Context:
» What consequences might result from the deployment of the system?

Sources of Bias: Friedman, B., & Nissenbaum, H. (1996). Bias in computer systems. ACM Trans. Inf. Syst., 14(3), 330–347.

A Human-Centered Design Perspective on the AMS System



Sources of Bias: Friedman, B., & Nissenbaum, H. (1996). Bias in computer systems. ACM Trans. Inf. Syst., 14(3), 330–347.

Overview

- **Increasing Societal Impact of Data-Driven Software**
  (examples with a focus on the AMS system)

- **The Need for Human-Centered Data Science**
  (definition, critical reflective practice, bias)

- **Mitigating Bias by Understanding Documentation As Reflexive Practice**
  (origins of data, datasheets of data sets)

When Data Become Data

Rarely can a magic moment be established when things become data

Borgman, C. L. (2016). Big data, little data, no data: Scholarship in the networked world. MIT press. (p. 62)

What is the Origin of Data?

Situated knowledges emphasis on **disclosing the mechanisms for the production of data.** These mechanisms for data production include social, cultural, historical and material conditions.

Additionally, a reflection on your **own perspective** is necessary but also **on existing values** of all stakeholders.

Data need Context

Reflexivity is a precondition for restoring context in data creation.

D'Ignazio, C., & Klein, L. F. (2020). Data feminism. MIT Press.
Haraway, D. (1988). Situated knowledges: The science question in feminism and the privilege of partial perspective. Feminist studies, 14(3), 575-599.


BY TIMNIT GEBRU, JAMIE MORGENSTERN,
BRIANA VECCHIONE, JENNIFER WORTMAN VAUGHAN,
HANNA WALLACH, HAL DAUMÉ III, AND KATE CRAWFORD
Datasheets
for Datasets
acm

Inspiration: Electronic Components

Sensirion Pressure sensor 1 pc(s) SDP610-025Pa -25 Pa up to 25 Pa (L x W x H) 29 x 18 x 27.05 mm

Primary Like Image

Item no.: 1313587

Manufacturer no.: 1-100759-02

EAN: 2050002956520

SENSIRION
THE SENSOR COMPANY

With the sensor out of the SDP600 series, launched one of the first digital dynamic Sensirion differential pressure sensor. The sensor has a digital I²C interface and measures even the smallest pressure differences (10 Pa) with highest sensitivity an...

Full description

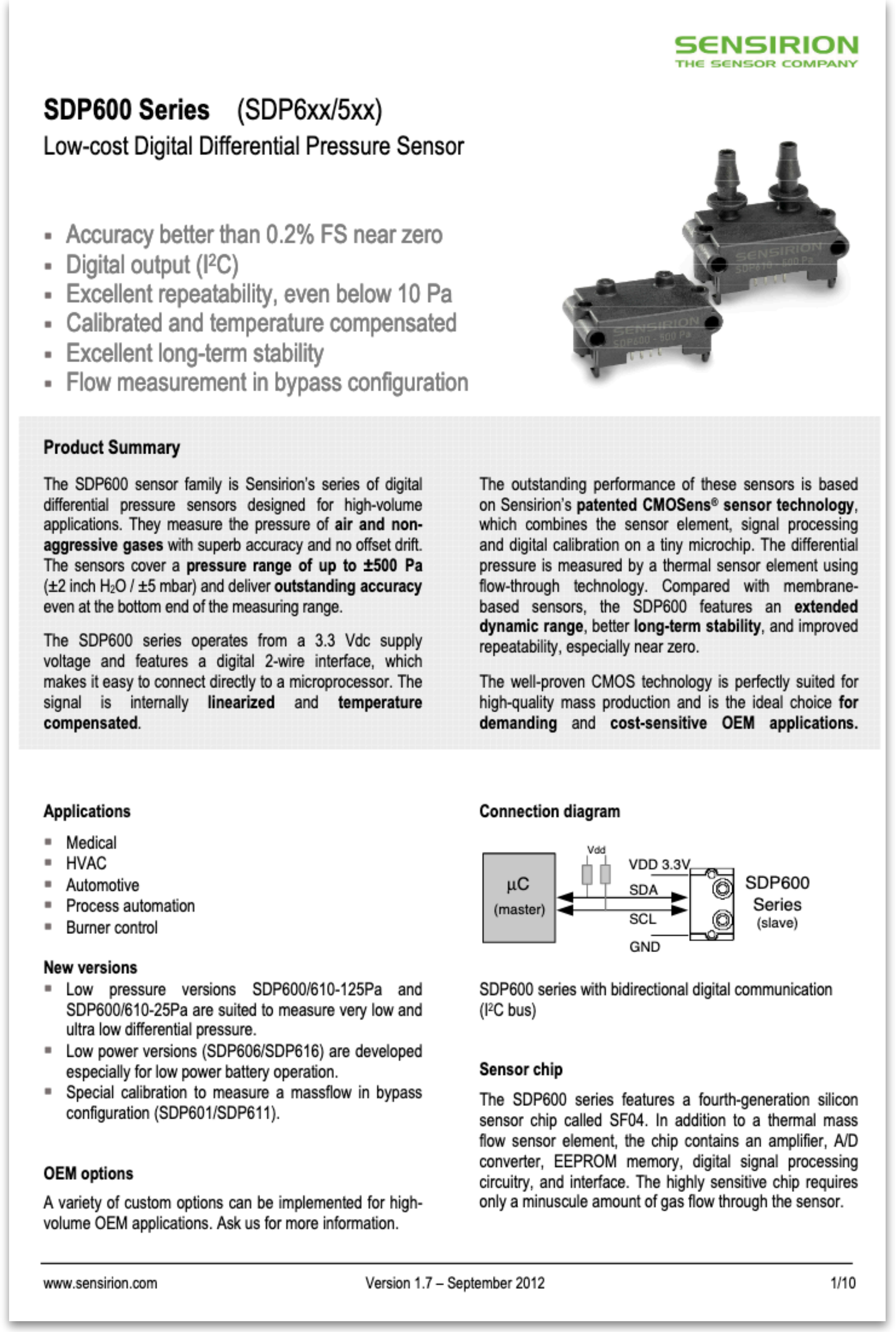
Guaranteed product originality

Documents & Downloads

Data sheets (1)

Datasheet 1313587 Sensirion Pressure sensor 1 pc(s) SDP610-025Pa -25 Pa up to 25 Pa (L x W x H) 29 x 18 x 27.05 mm

PDF



SENSIRION
THE SENSOR COMPANY

SDP600 Series (SDP6xx/5xx)
Low-cost Digital Differential Pressure Sensor

- Accuracy better than 0.2% FS near zero
- Digital output (I²C)
- Excellent repeatability, even below 10 Pa
- Calibrated and temperature compensated
- Excellent long-term stability
- Flow measurement in bypass configuration

Product Summary

The SDP600 sensor family is Sensirion's series of digital differential pressure sensors designed for high-volume applications. They measure the pressure of **air and non-aggressive gases** with superb accuracy and no offset drift. The sensors cover a **pressure range of up to ±500 Pa** (±2 inch H_2O / ±5 mbar) and deliver **outstanding accuracy** even at the bottom end of the measuring range.

The SDP600 series operates from a 3.3 Vdc supply voltage and features a digital 2-wire interface, which makes it easy to connect directly to a microprocessor. The signal is internally **linearized** and **temperature compensated**.

The outstanding performance of these sensors is based on Sensirion's **patented CMOSens® sensor technology**, which combines the sensor element, signal processing and digital calibration on a tiny microchip. The differential pressure is measured by a thermal sensor element using flow-through technology. Compared with membrane-based sensors, the SDP600 features an **extended dynamic range**, better **long-term stability**, and improved repeatability, especially near zero.

The well-proven CMOS technology is perfectly suited for high-quality mass production and is the ideal choice for **demanding** and **cost-sensitive OEM applications.**

Applications
- Medical
- HVAC
- Automotive
- Process automation
- Burner control

New versions
- Low pressure versions SDP600/610-125Pa and SDP600/610-25Pa are suited to measure very low and ultra low differential pressure.
- Low power versions (SDP606/SDP616) are developed especially for low power battery operation.
- Special calibration to measure a massflow in bypass configuration (SDP601/SDP611).

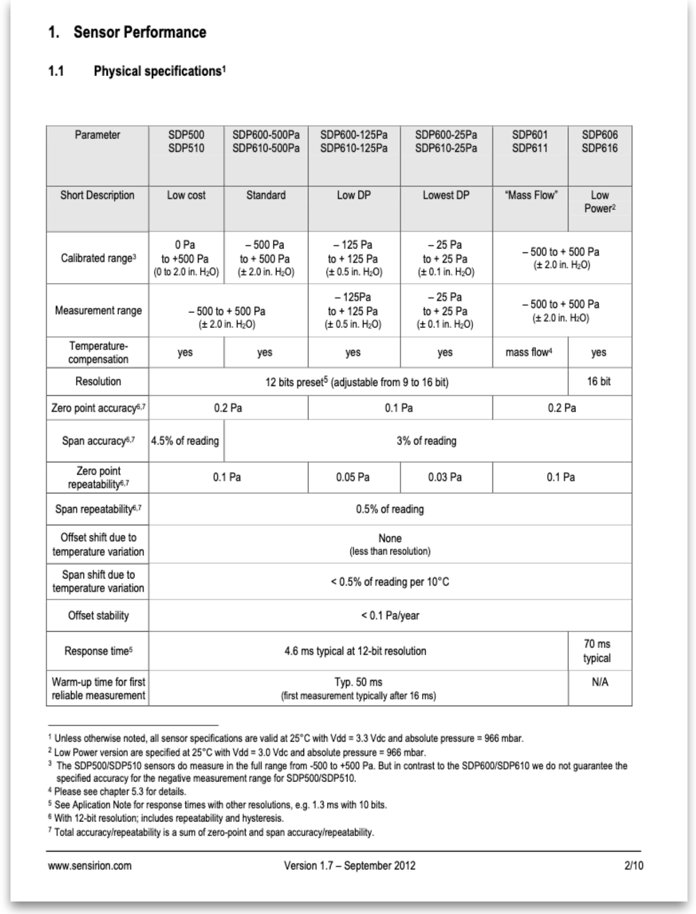
OEM options
A variety of custom options can be implemented for high-volume OEM applications. Ask us for more information.

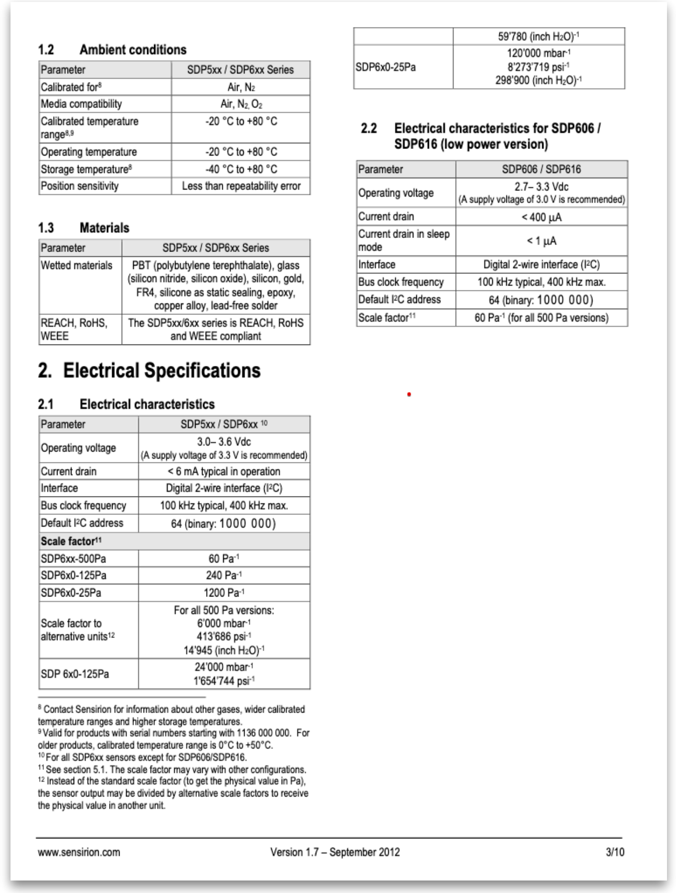
Connection diagram

SDP600 series with bidirectional digital communication (I²C bus)

Sensor chip
The SDP600 series features a fourth-generation silicon sensor chip called SF04. In addition to a thermal mass flow sensor element, the chip contains an amplifier, A/D converter, EEPROM memory, digital signal processing circuitry, and interface. The highly sensitive chip requires only a minuscule amount of gas flow through the sensor.

www.sensirion.com Version 1.7 – September 2012 1/10





Screenshots taken from https://www.conrad.com/p/sensirion-pressure-sensor-1-pcs-sdp610-025pa-25-pa-up-to-25-pa-l-x-w-x-h-29-x-18-x-2705-mm-1313587

Objective of «Datasheets for Datasets»

For Dataset Creators Encouraging careful reflection on the process of creating, distributing, and maintaining a dataset, including any underlying assumptions, potential risks or harms, and implications of use.

For Dataset Consumers Ensuring that they (policy makers, consumer advocates, investigative journalists, individuals*) have the information they need to make informed decisions about using a dataset for their chosen tasks and avoid unintentional misuse.

(*) individuals whose data is included in datasets, and individuals who may be impacted by models

Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford. 2021. Datasheets for datasets. Commun. ACM 64, 12 (December 2021), 86–92.

Sections of «Datasheets for Datasets»

- motivation
- composition
- collection process
- preprocessing/ cleaning/labeling
- uses
- distribution
- maintenance

Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford. 2021. Datasheets for datasets. Commun. ACM 64, 12 (December 2021), 86–92.

Sections of «Datasheets for Datasets»

motivation	Describe the motivations for creating the dataset, including funding, any specific tasks the authors had in mind, and who the authors are.
composition	Describe the composition of the dataset, like what kinds of data are in it, how it was collected, whether labels are associated with the data, and whether the dataset contains sensitive information.
collection process	Describe the data collection process, like how the data was collected, where or who is was collected from, who was involved in the collection process, and, if people are involved, if consent was given for the data to be collected.
preprocessing/ cleaning/labeling	Whether the data was process or labelled and how it was done.
uses	The tasks the dataset is intended to be used for, how it has already been used, and limitations of use. Distribution: How the dataset will be distributed and to who, and any restrictions on distribution.
maintenance	Who and how the dataset will be maintained, and if and how others will be able to build on it.
distribution	Whether the dataset is distributed to third parties outside of the owner with what license by employing any restrictions.

Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford. 2021. Datasheets for datasets. Commun. ACM 64, 12 (December 2021), 86–92.

Example

Motivation

For what purpose was the dataset created? Was there a specific task in mind? Was there a specific gap that needed to be filled? Please provide a description.

The dataset was created to enable research on predicting sentiment polarity—i.e., given a piece of English text, predict whether it has a positive or negative affect—or stance—toward its topic. The dataset was created intentionally with that task in mind, focusing on movie reviews as a place where affect/sentiment is frequently expressed.[1]

Who created the dataset (e.g., which team, research group) and on behalf of which entity (e.g., company, institution, organization)?

The dataset was created by Bo Pang and Lillian Lee at Cornell University.

Who funded the creation of the dataset? If there is an associated grant, please provide the name of the grantor and the grant name and number.

Funding was provided from five distinct sources: the National Science Foundation, the Department of the Interior, the National Business Center, Cornell University, and the Sloan Foundation.

Any other comments?

None.

Movie Review Polarity — **Thumbs Up? Sentiment Classification using Machine Learning Techniques**

these are words that could be used to describe the emotions of john sayles' characters in his latest , limbo . but no , i use them to describe myself after sitting through his latest little exercise in indie egomania . i can forgive many things . but using some hackneyed , whacked-out , screwed-up * non * - ending on a movie is unforgivable . i walked a half-mile in the rain and sat through two hours of typical , plodding sayles melodrama to get cheated by a complete and total copout finale . does sayles think he's roger corman ?

Figure 1. An example "negative polarity" instance, taken from the file neg/cv452_tok-18656.txt.

exception that no more than 40 posts by a single author were included (see "Collection Process" below). No tests were run to determine representativeness.

What data does each instance consist of? "Raw" data (e.g., unprocessed text or images)or features? In either case, please provide a description.

Each instance consists of the text associated with the review, with obvious ratings information removed from that text (some errors were found and later fixed). The text was down-cased and HTML tags were removed. Boilerplate newsgroup header/footer text was removed. Some additional unspecified automatic filtering was done. Each instance also has an associated target value: a positive (+1) or negative (-1) sentiment polarity rating based on the number of stars that that review gave (details on the mapping from number of stars to polarity is given below in "Data Preprocessing").

Is there a label or target associated with each instance? If so, please provide a description.

The label is the positive/negative sentiment polarity rating derived from the star rating, as described above.

Is any information missing from individual instances? If so, please provide a description, explaining why this information is missing (e.g., because it was unavailable). This does not include intentionally removed information, but might include, e.g., redacted text.

Everything is included. No data is missing.

Are relationships between individual instances made explicit (e.g., users' movie ratings, social network links)? If so, please describe how these relationships are made explicit.

None explicitly, though the original newsgroup postings include poster name and email address, so some information (such as threads, replies, or posts by the same author) could be extracted if needed.

Are there recommended data splits (e.g., training, development/validation, testing)? If so, please provide a description of these splits, explaining the rationale behind them.

The instances come with a "cross-validation tag" to enable replication of cross-validation experiments; results are measured in classification accuracy.

Are there any errors, sources of noise, or redundancies in the dataset? If so, please provide a description.

See preprocessing below.

Is the dataset self-contained, or does it link to or otherwise rely on external resources (e.g., websites, tweets, other datasets)? If it links

movie-review-data/rt-polaritydata.README.1.0.txt; http://www.cs.cornell.edu/people/pabo/movie-review-data/poldata.README.2.0.txt.

Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortma… Commun. ACM 64, 12 (December 2021), 86–92.

Key Takeaways

- **Data Science needs a Human-Centered perspective that is accompanied with a critical-reflexive practice.**
  - Besides considering the origins of your data, engage with concepts such us fairness, transparency, accountability, interpretability, privacy as well

- **General Remarks on educating data scientists**
  - Critical reflexive practice should be integrated in data science education - not as an add-on
  - Computing education should be joined by a humanistic education