



Ethik in Informatik und Digital Humanities

Vorlesung *Einführung in die Digital Humanities*
MSc Digital Humanities | Wintersemester 2020/21

Prof. Dr. Christof Schöch

(1) Einstieg: Wissenschaft und Verantwortung

Anstoß: Timnit Gebru

The image shows a Twitter mobile interface. On the left is a sidebar with navigation icons: Home, Explore, Notifications, Messages, Bookmarks, Lists, Profile, and More. At the bottom of the sidebar is a blue 'Tweet' button. The main content area is titled 'Tweet' and shows a pinned tweet from 'Timnit Gebru' (@timnitGebru). The tweet text reads: 'You know these last 2 weeks were supposed to be my vacations and we're currently on a road trip. Trying to find my email with my conditions that were so unacceptable to them, but one of them is basically point #1 here:' Below the text is a portrait photo of Timnit Gebru. A callout box below the photo contains the text: 'Standing with Dr. Timnit Gebru—#ISupportTimnit #BelieveBlackWomen We, the undersigned, stand in solidarity with Dr. Timnit Gebru, who was terminated from her position as Staff Research Scientist and...'. It also includes a link: googlewalkout.medium.com. At the bottom of the tweet card, it says '4:23 PM · Dec 4, 2020 · Twitter Web App' and shows engagement statistics: '1.2K Retweets 312 Quote Tweets 4.8K Likes'. Below the tweet are four interaction icons: a speech bubble, a retweet arrow, a heart, and an upward arrow.

<https://twitter.com/timnitGebru/status/1334881120920498180>

Freiheit und Verantwortung

- Wissenschaftliche Freiheit ist ein hohes Gut
- Aber: aus Freiheit ergibt sich Verantwortung
- Daher: Große Bedeutung der Wissenschaftsethik
- Spezieller Fall: Ethik der Digitalisierung

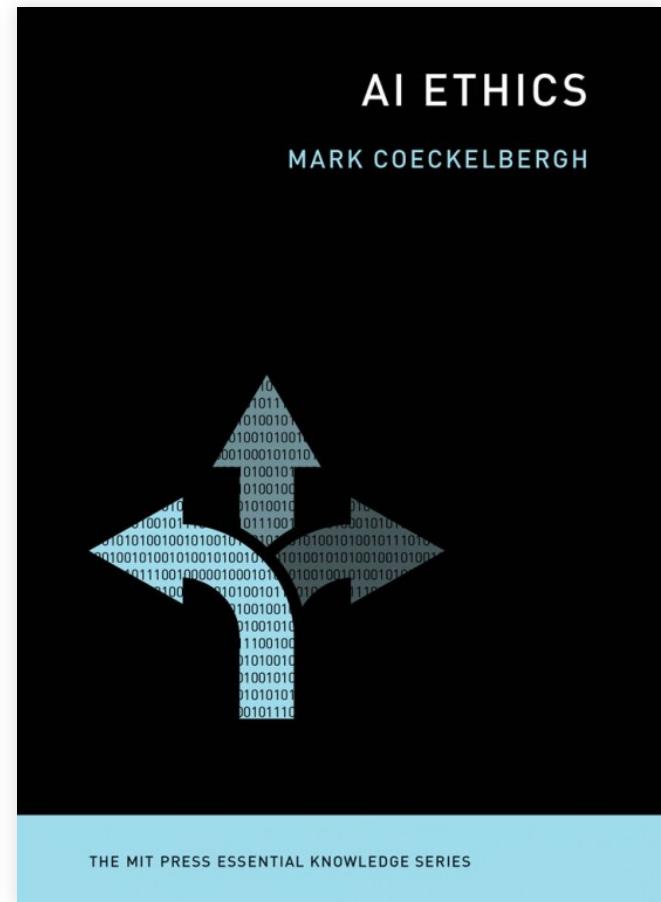
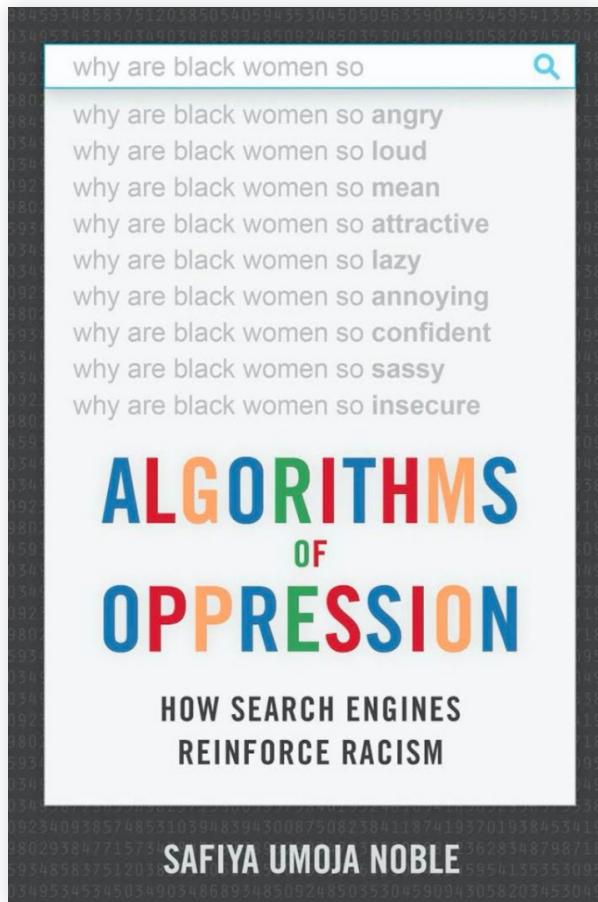
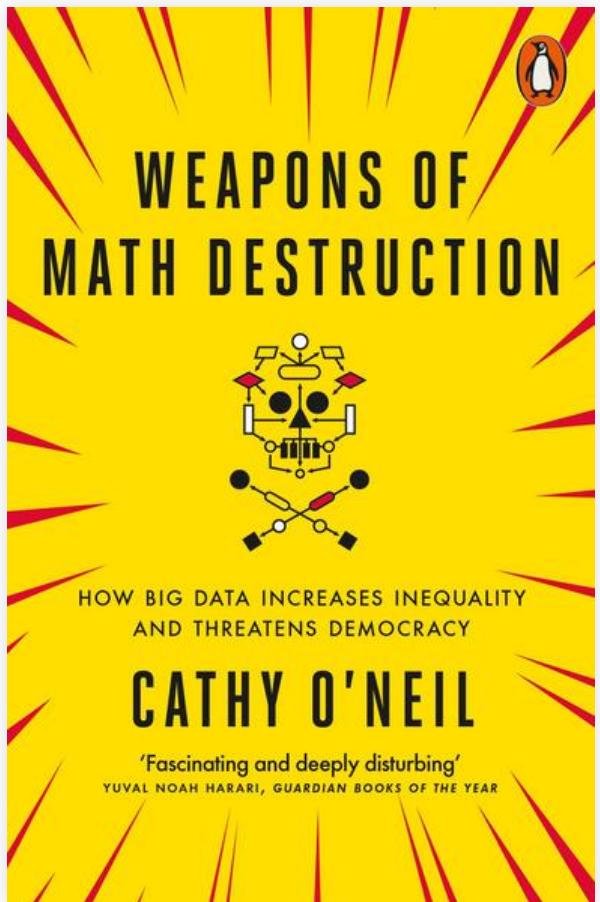
Grundlagen ethischer Entscheidung

- Folgen des persönlichen Rechtsgefühls
- Maximierung des gesellschaftlichen Nutzens
- Aus ethischen Prinzipien abgeleitete Entscheidungen
 - Menschenwürde
 - Freiheit, Gleichheit, Gerechtigkeit
 - Solidarität
 - Toleranz

Warum gerade Digitalisierung?

- Digitalisierung und Informatik
 - ... entwickeln sich schnell (Regulierungslücke)
 - ... betreffen viele Menschen in vielen Bereichen (Impact)
 - ... betreffen Menschen in wichtigen Dingen ("high stakes scenarios")
 - ... sind für Betroffene nicht verständlich (Intransparenz)

Ethik und Digitalisierung



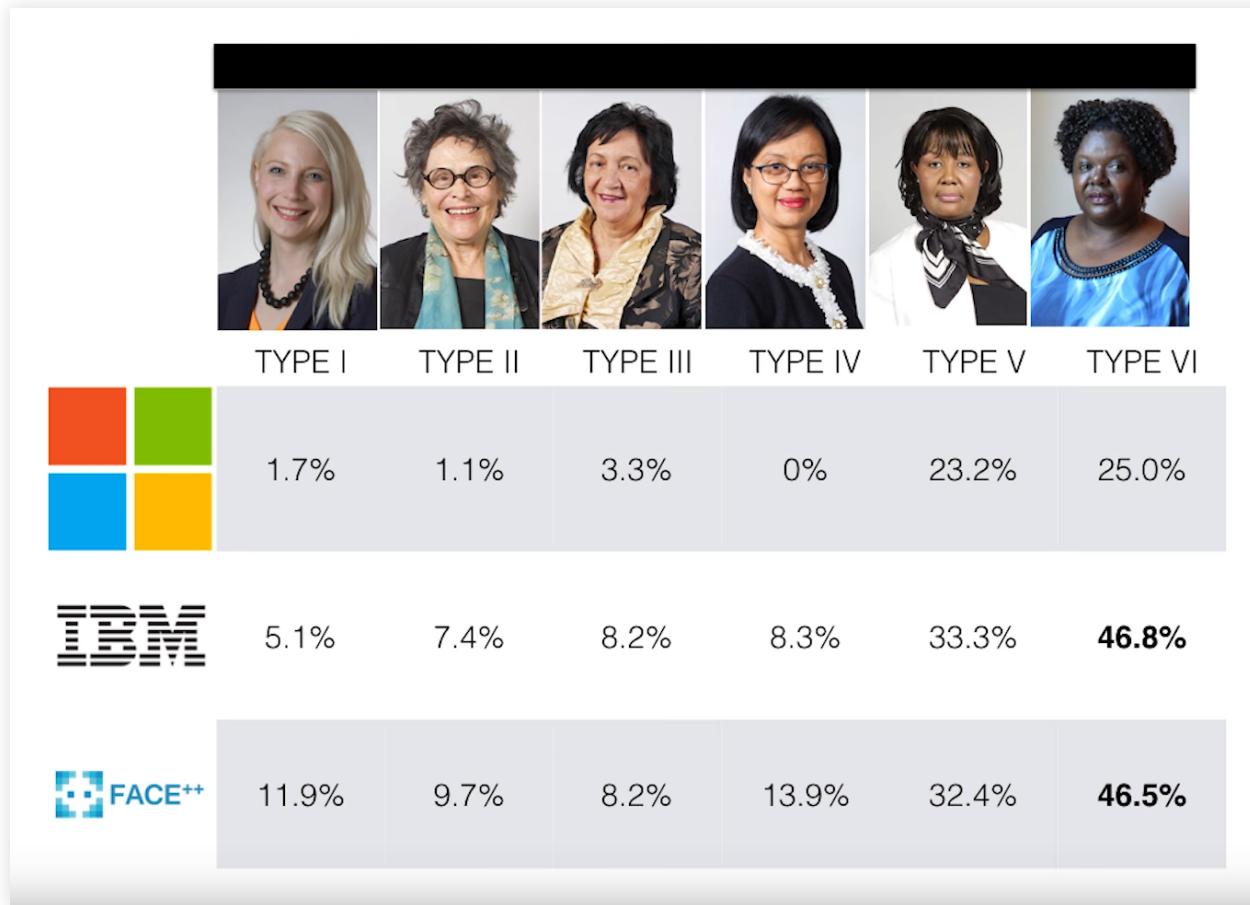
Aspekte von Ethik in Informatik und Digital Humanities

1. Biases in Datensätzen und Algorithmen
2. Digitalisierung und Privatsphäre
3. Verantwortung und Künstliche Intelligenz
4. Erklärbarkeit von Deep Learning
5. Ökologische Nachhaltigkeit bei Big Data
6. Diversität der Akteure

(2) Biases in Datensätzen und Algorithmen

Beispiel: Face Recognition

3.1



<https://news.mit.edu/2018/study-finds-gender-skin-type-bias-artificial-intelligence-systems-0212> ; Buolamwini and Gebru 2018

3.2

Beispiel DH: Reproduktion von Stereotypen

Vive la différence: Tracing the (Authorial) Gender Signal by Multivariate Analysis of Word Frequencies

Jan Rybicki

Institute of English Studies, Jagiellonian University, Kraków,
Poland

Abstract

Multivariate analysis of word frequencies is used to identify the gender of authors in a corpus of 18th- and early 19th-century English sentimental and Gothic fiction. Results obtained with most frequent words are compared to those produced with medium-frequency Burrows's Zeta words characteristic for both genders. Gender-sensitive words from two periods (18th/19th c. and 19th/20th c.) are compared in terms of their usefulness for gender identification in literary texts.

Correspondence:

Jan Rybicki, Institute of
English Studies, Jagiellonian
University, Al. Mickiewicza
9, 33-332, Kraków, Poland.

E-mail:

jkrybicki@gmail.com

Rybicki 2015, <https://academic.oup.com/dsh/article-abstract/31/4/746/2748261>

Beispiel DH: Diversity Paradox in ELTeC

Distant [?] Reading

ELTeC Summary Page

As well as the following summary statistics, this page provides links to human-readable versions of each text currently included in the European Literary Text Collection (ELTeC). Click on a language code in the table below to see now available in that language. Then click on the identifier of a text to see a simple rendering of the text as produced by CETEicean. The original source files are stored in a GitHub repository at COST-ELTeC, and may be downloaded freely from there.

The E5C column gives the [conformance score](#) calculated for each repository and is displayed in green if conformance is high. The other columns give counts for each of the four balance criteria, with numbers in red indicating criterion is unsatisfied. Hovering over the last figure in each column displays the E5C score calculated for that criterion.

This remains a work in progress! Comments and reports of any problems are much appreciated: send them to the [WG1 Issue Tracker](#).

Language	Last update	Texts	Words	AUTHORSHIP			LENGTH			TIME SLOT				REPRINT COUNT				
				Male	Female	1-title	3-title	Short	Medium	Long	1840-59	1860-79	1880-99	1900-20	range	Frequent	Rare	E5C
cze	2020-11-16	16	366626	14	2	12	0	16	0	0	5	6	5	0	6	0	15	33.85
deu	2020-11-15	98	12086096	65	33	36	8	20	37	41	24	24	25	25	1	46	46	93.85
eng	2020-11-21	99	12198190	49	50	69	10	26	27	46	21	22	31	25	10	32	67	97.69
fra	2020-11-15	100	8712219	66	34	58	10	32	38	30	25	25	25	25	0	44	56	101.54
gre	2019-09-22	11	42524	10	1	11	0	11	0	0	0	1	6	4	6	3	4	37.83
hun	2020-11-15	100	6948590	79	21	71	9	47	31	22	22	21	27	30	9	32	67	100.00
ita	2019-11-21	34	3328244	32	2	19	3	13	10	11	5	12	10	7	7	12	0	55.97
lav	2020-07-11	2	106045	2	0	2	0	0	2	0	0	0	1	1	1	0	1	21.54
lit	2020-08-20	25	636132	18	7	16	1	19	3	2	5	3	3	14	11	6	18	55.38
nor	2020-11-17	50	3195845	36	14	20	8	25	17	8	5	2	28	15	26	30	20	71.54
por	2020-11-21	100	6688254	83	17	73	9	41	41	18	13	37	19	31	24	26	60	94.62
rom	2020-11-15	80	4905678	65	11	43	7	35	29	16	4	14	23	39	35	24	56	83.08
sly	2020-11-15	100	5682120	89	11	26	5	53	39	8	2	13	36	49	47	48	52	78.46
spa	2020-11-15	81	6874582	65	16	42	5	30	27	24	16	15	25	25	10	42	39	90.77
srp	2020-11-19	70	3151549	63	7	23	9	43	26	1	2	9	29	30	28	26	35	70.77
swe	2020-11-15	58	4960085	29	28	18	8	16	24	18	15	3	20	20	17	17	41	76.92
ukr	2020-11-21	37	1451622	26	11	20	5	23	12	2	3	9	7	18	15	25	12	63.08

Summary produced: 2020-11-24

<https://distantreading.github.io/ELTeC/>

(3) Digitalisierung und Privatsphäre

Beispiel: De-Anonymisierung

I Know What You Did Last Summer: risks of location data leakage in mobile and social computing.

Lukasz Jedrzejczyk*, Blaine A. Price*, Arosha K. Bandara*, Bashar Nuseibeh*

Department of Computing, The Open University, UK*

L.Jedrzejczyk; B.A.Price; A.K.Bandara; B.Nuseibeh@open.ac.uk

ABSTRACT

Advances in mobile and web technologies have brought unwanted access to often sensitive data, ranging from our personal details, where we work, where we live and even behavioral patterns. The increasing use of social networks and location-aware mobile applications raise a number of concerns, including the issue of ensuring users' privacy. In order to explore those concerns we conducted an exploratory study in re-identifying people based on their movements and publicly available information. We observed anonymous users of a location-based social networking application in their natural environment and demonstrated how to re-identify them based on that data. In addition to discovering location-based private data, we were also able to find quite a number of facts from their private lives. This article reports on the methodology we used, ethical issues related to informed consent and user's reaction to the fact of being re-identified.

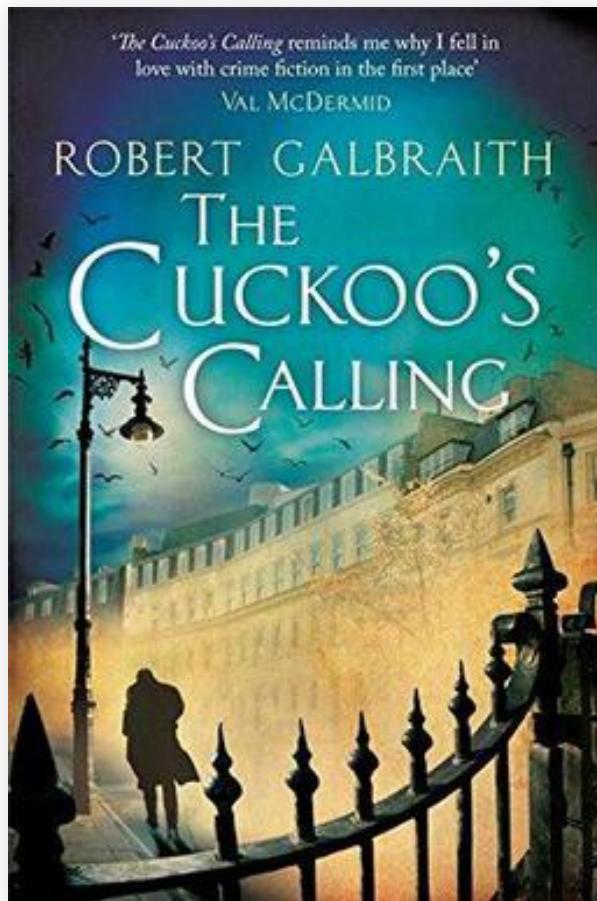
location records over a period of time. To avoid observation bias, we wanted to look at publicly available data provided by people who didn't know we were studying them. Some of the key questions driving our research were:

- *Is it possible to re-identify anonymous users based on the publicly accessible movement data combined with other, freely available, data ?*
- *How well do users understand the risks of location data sharing?*
- *What choices make users more vulnerable to location data leakage?*

The research presented in this article demonstrates the main vulnerabilities in commercial applications that can lead to inadvertent data leakage and retention. We show how we used publicly accessible data and data-mining tools to re-identify anonymous users of location based service and how

Siehe Jedrzejczyk et al. 2009

Beispiel DH: Galbraith / Rowling



The Rowling Case: A Proposed Standard Analytic Protocol for Authorship Questions FREE

Patrick Juola

Digital Scholarship in the Humanities, Volume 30, Issue suppl_1, December 2015, Pages i100–i113, <https://doi.org/10.1093/dsh/fqv040>

Published: 19 October 2015

PDF Split View Cite Permissions Share ▾

Abstract

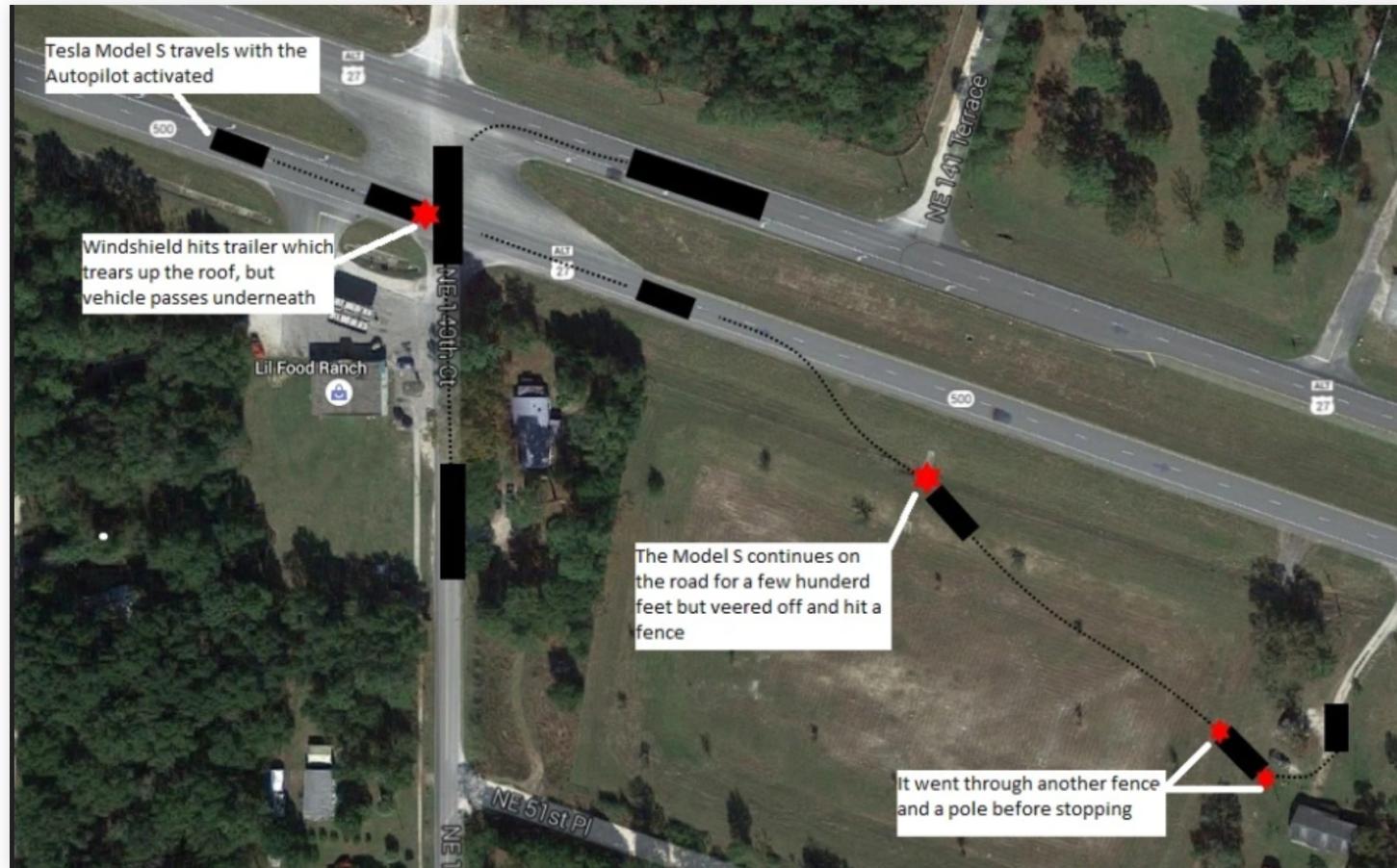
We propose a possible solution to one of the major weaknesses in the application of authorship attribution—the absence of clear-cut standards for accurate analytic practice. To address this, we propose a specific practice as a possible standard and present four recent cases applying this standard. The key elements of this protocol are the use of an ad hoc distractor set in conjunction with multiple analyses structured as a set of elimination tests. This protocol (or close variants of it) has been used in at least four separate cases across a wide variety of documents and consumers. It is mathematically supported while still being easy to understand. We are confident that the proposed protocol will provide a relatively straightforward and understandable way to reduce controversy regarding stylometric authorship attribution, and thereby increase its uptake and credibility.

Issue Section: Original Articles

Juola 2015, https://academic.oup.com/dsh/article/30/suppl_1/i100/363234

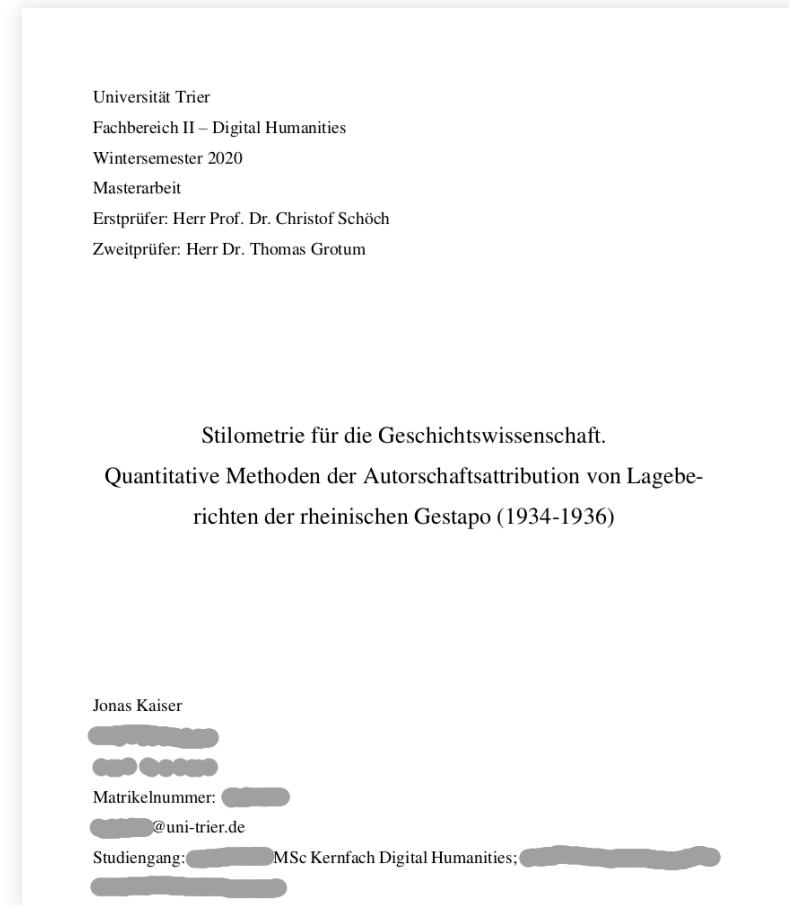
(4) Moraleische Verantwortung und Künstliche Intelligenz

Beispiel: Verantwortung für Unfälle selbstfahrender Autos



<https://electrek.co/2016/07/01/understanding-fatal-tesla-accident-autopilot-nhtsa-probe/>

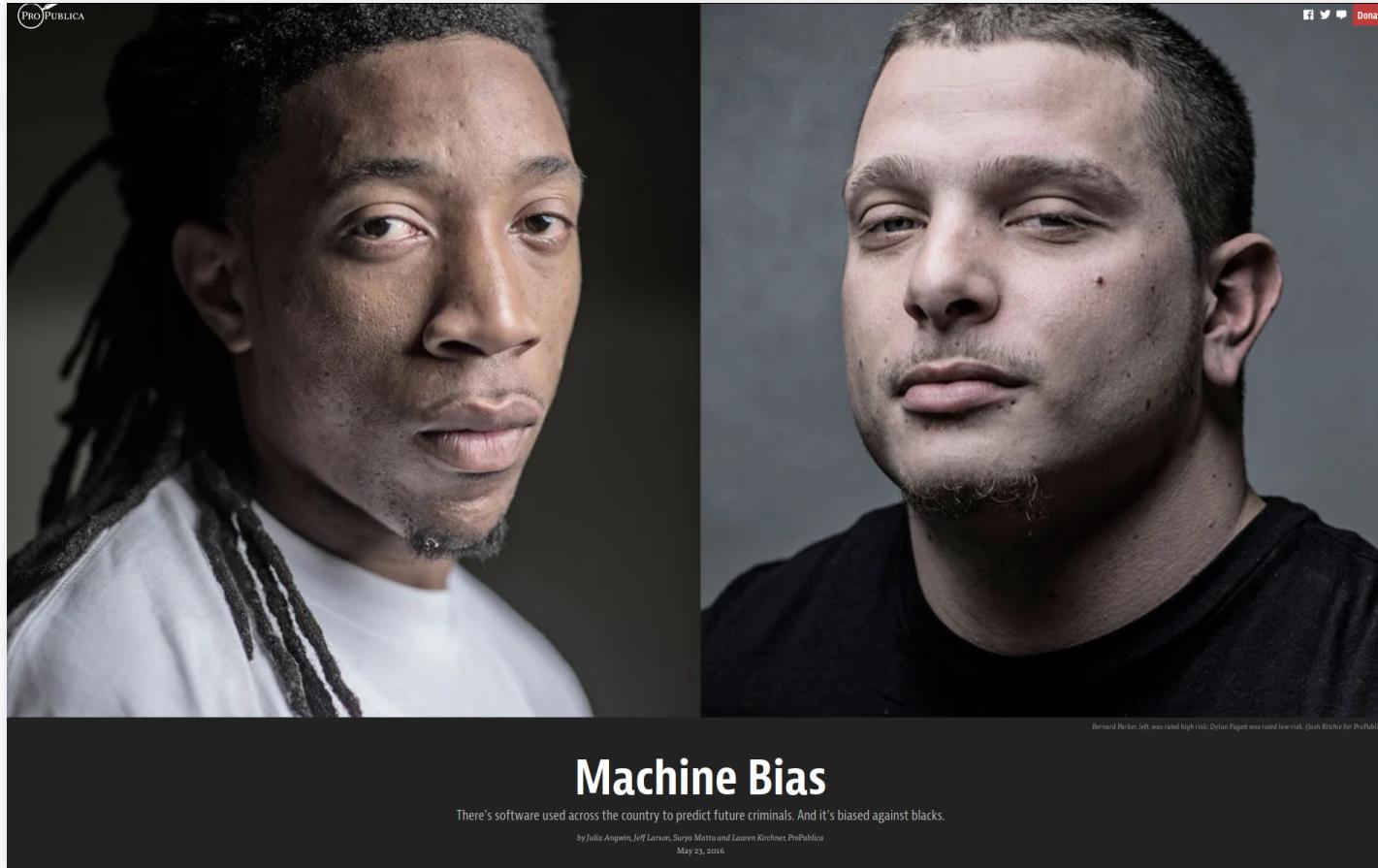
Beispiel DH: Autorschaft von Gestapo-Berichten



(5) Erklärbarkeit von Deep Learning

Beispiel: Rückfallrisiko

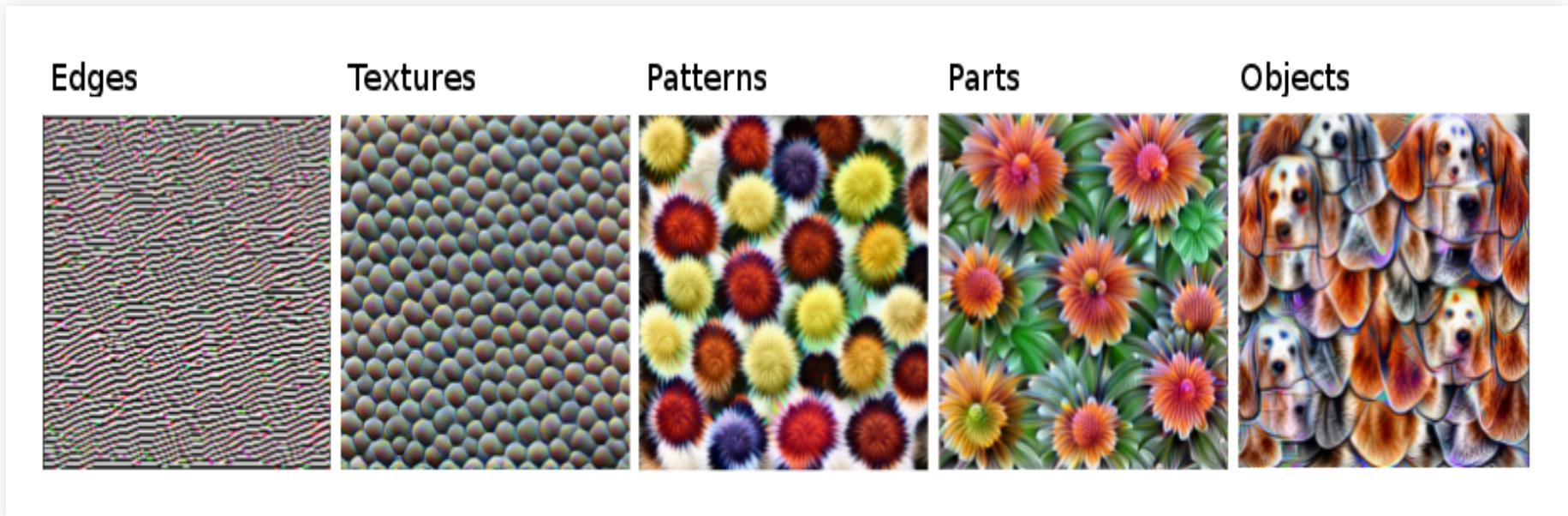
6.1



Angwin et al. 2016, <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

6.2

Beispiel DH: Interpretierbarkeit statt Performanz

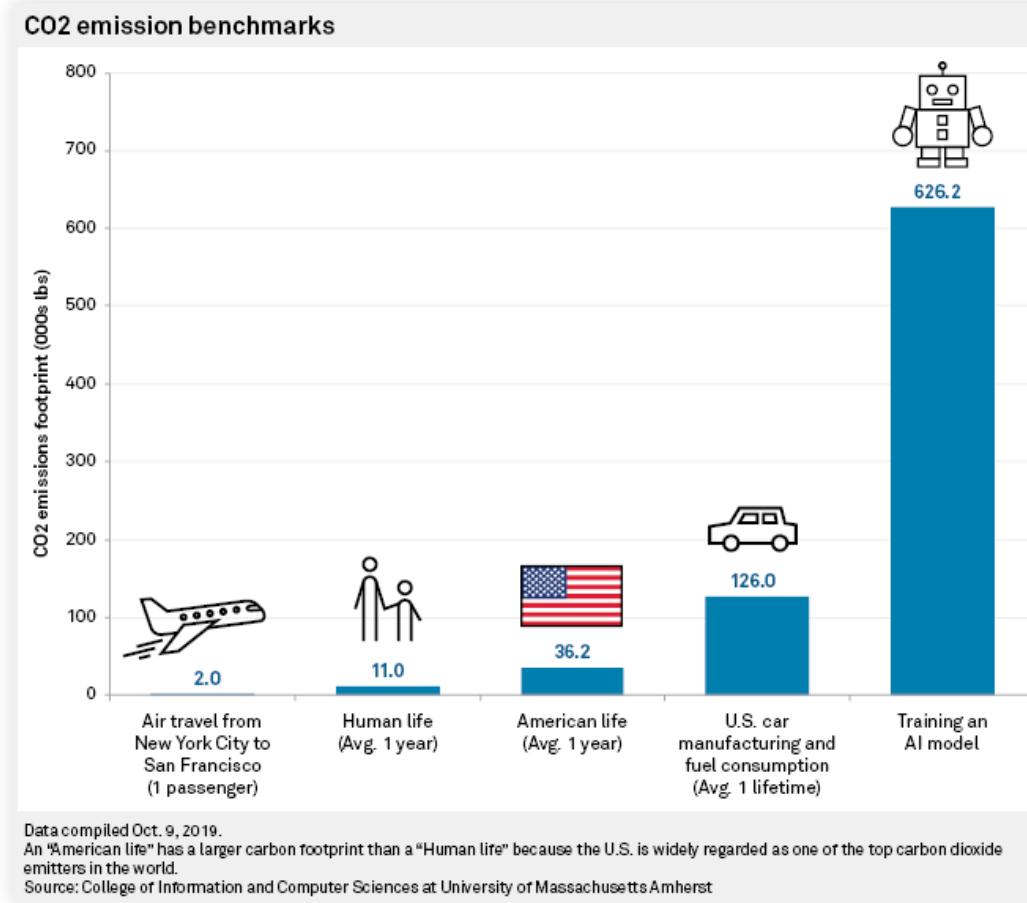


<https://christophm.github.io/interpretable-ml-book/cnn-features.html>

(6) Ökologische Nachhaltigkeit bei Big Data-Anwendungen

Beispiel: Trainieren von großen Sprachmodellen

7.1



<https://www.spglobal.com/marketintelligence/en/news-insights/trending/HyvwuXMO9YgqHfj7J6tGIA2>

7.2

Beispiel DH: Die DH Conference

The screenshot shows the 'Conference' section of the ADHO website. The page header includes the ADHO logo and navigation links for About, Statements, Committees, SIGs, Publications, Conference (which is highlighted in blue), and Awards. The main content area is titled 'Conference' and contains a paragraph about the history of the joint conference, mentioning its first meeting in 1989 at the University of Toronto. Below this is a bulleted list of 43 past conference locations, each with a city and date.

Conference

The ADHO organizes and sponsors an annual conference. The first joint conference was held in 1989 at the [University of Toronto](#). That event was the 16th annual meeting of ALLC and the ninth annual meeting of the ACH-sponsored International Conference on Computers and the Humanities (ICCH). Since then, the joint conference has grown to include additional organizations, and it has been held in cities around the world:

- University of Graz, Graz, Austria (11-15 July 2022) - *postponed until summer 2023*
- University of Tokyo, Tokyo, Japan (26-30 July 2021) - *postponed until summer 2022*
- Carleton University and the University of Ottawa, Ottawa, Canada (22-24 July 2020)
 - Virtual Conference Website.
- Universiteit Utrecht, Utrecht, Netherlands (9-12 July 2019)
- El Colegio de México, the Universidad Nacional Autónoma de México (UNAM), and Red de Humanidades Digitales (RedHD), Mexico City, Mexico (24-30 June 2018)
- McGill University and l'Université de Montréal, Montreal, Canada (8-11 August 2017)
- Jagiellonian University and the Pedagogical University of Kraków, Poland (11-16 July 2016)
- University of Western Sydney, Australia (29 June - 3 July 2015)
- University of Lausanne - EPFL, Switzerland (7-12 July 2014)
- University of Nebraska-Lincoln, USA (16-19 July 2013)
- University of Hamburg, Germany (16-22 July 2012)
- Stanford University, USA (19-22 June 2011)
- King's College London, UK (7-10 July 2010)
- University of Maryland, College Park, Maryland (22-25 June 2009)
- University of Oulu, Finland (25-29 June 2008)
- University of Illinois Urbana-Champaign (2-8 June 2007)
- The Sorbonne, Paris, France (5-9 July 2006)
- University of Victoria, British Columbia, Canada (15-18 June 2005)
- Goteborg University, Sweden (11-16 June 2004)
- University of Georgia, USA (29 May - 2 June 2003)
- University of Tübingen, Germany (23-28 July 2002)
- New York University, USA (13-16 July 2001)
- University of Glasgow, Scotland, UK (21-25 July 2000)
- University of Virginia, Charlottesville, Virginia, USA (9-13 June 1999)
- Lajos Kossuth University, Debrecen, Hungary (5-10 July 1998)
- Queen's University, Kingston, Ontario, Canada (3-7 June 1997)
- University of Bergen, Norway (25-29 June 1996)
- University of California, Santa Barbara, California (11-15 July 1995)
- The Sorbonne, Paris, France (19-23 July 1994)
- Georgetown University, Washington, DC (16-19 June 1993)
- Oxford University, Oxford, England (5-9 April 1992)
- Arizona State University, Tempe, Arizona (17-21 March 1991)
- University of Siegen, Germany (4-9 June 1990)

<https://adho.org/conference>

(7) Diversität der Akteure

Beispiel: Timnit Gebru bei Google

8 . 1

Süddeutsche Zeitung  Coronavirus Politik Wirtschaft Meinung Panorama Sport München Bayern Kultur Gesellschaft Wissen Reise Auto mehr...

15. Dezember 2020, 11:23 Uhr Künstliche Intelligenz

Warum musste Google-Expertin Timnit Gebru gehen?



Timnit Gebru liegt im Clinch mit ihrem alten Arbeitgeber Google. (Archivbild) (Foto: AFP)

Die Computeringenieurin war bei Google für die ethischen Fragen künstlicher Intelligenz zuständig. Bis sie rausflog – was aus ihrer Sicht ebenjene ethischen Fragen aufwirft.

Von Michael Moorstedt

<https://www.sueddeutsche.de/digital/gebru-google-alphabet-kuenstliche-intelligenz-ki-ethik-silicon-valley-1.5146265>

8 . 2

Beispiel DH: Vortrag in Shanghai



CDH 2020
Benevolence and Excellence: Digital Humanities and Chinese Culture Call for Papers

DH 2020: Benevolence and Excellence: Digital Humanities and Chinese Culture Call for Papers
Submitted by admin on Tue, 12/10/2019 - 16:56
DH (Digital Humanities) has flourished in China for over ten years. Over the past years, we strive to keep up with the world, re-examine the past and configure the present. When Prof. Dr. Melissa Terras from the University of Edinburgh visited Nanjing University in November 2019, she asked us about the origin and features of DH in China. She raised an insightful question that left us in deep thought: what unique vision can the origin of 'Ren'-oriented Chinese humanism, and humanities and spiritual legacy represented by the Confucian-based disciplines of 'Six Arts' bring to the DH research?
[Read more](#)
[Log in](#) to post comments

DH 2020: Benevolence and Excellence: Digital Humanities and Chinese Culture Call for Participants
Submitted by admin on Tue, 12/10/2019 - 17:02
DH (Digital Humanities) has flourished in China for over ten years. Over the past years, we strive to keep up with the world, re-examine the past and configure the present. When Prof. Dr. Melissa Terras from the University of Edinburgh visited Nanjing University in November 2019, she asked us about the origin and features of DH in China. She raised an insightful question that left us in deep thought: what unique vision can the origin of 'Ren'-oriented Chinese humanism, and humanities and spiritual legacy represented by the Confucian-based disciplines of 'Six Arts' bring to the DH research?

<http://dh2020.library.sh.cn/en>

Abschluss

Lektürehinweise

Referenzlektüre

- Timnit Gebru: "How To Stop Artificial Intelligence From Marginalizing Communities?" *TEDx College Park*, 2018. URL: <https://www.youtube.com/watch?v=PWCToVt1CJM>

Vertiefende Darstellungen

- Mark Coeckelbergh: *AI Ethics*. Boston: MIT Press, 2020.
- Safiya Umoja Noble. *Algorithms of Oppression. How Search Engines Reinforce Racism*. New York: New York University Press, 2018.
- Katharina Zweig. *Ein Algorithmus hat kein Taktgefühl*. München: Heyne, 2019.
- Cathy O'Neil. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown, 2016.

Bibliographie

- Buolamwini, Joy, and Timnit Gebru. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." In Conference on Fairness, Accountability and Transparency, 77–91, 2018.
- Denton, Emily, Ben Hutchinson, Margaret Mitchell, Timnit Gebru, and Andrew Zaldivar. "Image Counterfactual Sensitivity Analysis for Detecting Unintended Bias." ArXiv:1906.06439 [Cs, Stat], October 3, 2020. <http://arxiv.org/abs/1906.06439>.
- Raji, Inioluwa Deborah, Timnit Gebru, Margaret Mitchell, Joy Buolamwini, Joonseok Lee, and Emily Denton. "Saving Face: Investigating the Ethical Concerns of Facial Recognition Auditing." In Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, 145–151. AIES '20. New York, NY, USA: Association for Computing Machinery, 2020. <https://doi.org/10.1145/3375627.3375820>.
- Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford. "Datasheets for Datasets." ArXiv Preprint ArXiv:1803.09010, 2018.
- Jedrzejczyk, Lukasz, Blaine A. Price, Arosha K. Bandara, Bashar Nuseibeh, W. Hall, and M. Keynes. "I Know What You Did Last Summer: Risks of Location Data Leakage in Mobile and Social Computing." Department of Computing Faculty of Mathematics, Computing and Technology The Open University, 2009, 1744–1986.
- Juola, Patrick. "The Rowling Case: A Proposed Standard Protocol for Authorship Attribution." Digital Scholarship in the Humanities 30, no. suppl. 1 (2015): 100–113. <https://doi.org/10.1093/llc/fqv040>.
- Angwin, Julia, Jeff Larson, Surya Mattu, and Lauren Kirchner. "Machine Bias , Jeff Larson, Surya Mattu and Lauren Kirchner." ProPublica, 2016. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing?token=xbvF5KLcDIV6vr6B2AF9D0LIUK_IwLni.
- Malte Rehbein und Christian Thies: "Ethik". *Digital Humanities: Eine Einführung*, hg. von Fotis Jannidis, Ulrich G. Klemke und Michael Riegger, Stuttgart, Metzler, 2017, S. 252–257.

Christof Schöch, 2020
<http://www.christof-schoech.de>

Lizenz: Creative Commons Attribution 4.0