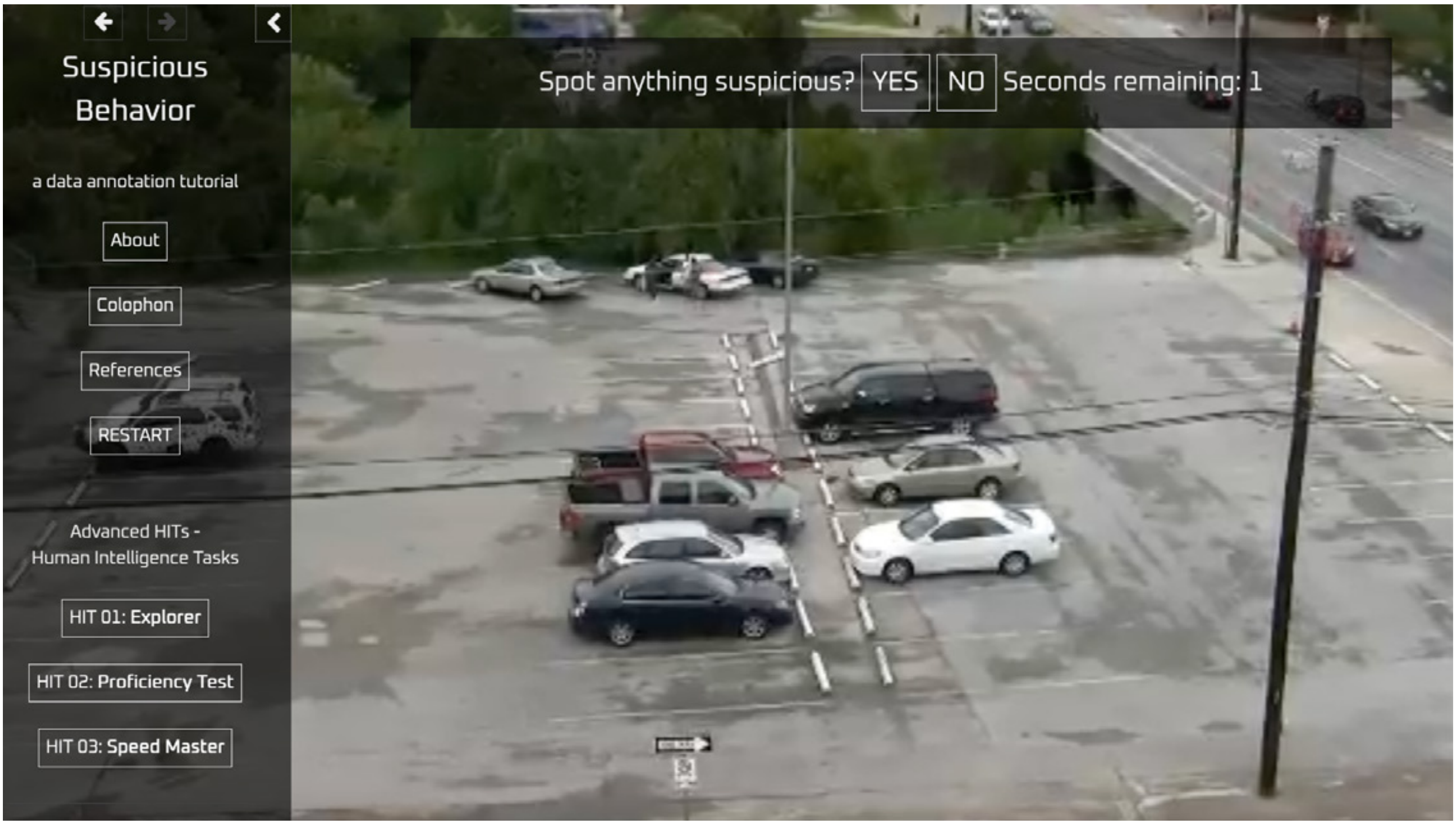
xCoAx 2022 10th Conference on
Computation, Communication, Aesthetics & X
2022.xCoAx.org
Coimbra, Portugal



Suspicious Behavior

Linda Kronman
linda@kairus.org
University of Bergen,
Bergen, Norway

Andreas Zingerle
andreas@kairus.org
Kairus Art+Research,
Graz, Austria

Suspicious Behavior is a fictional annotation tutorial inviting readers to critically examine machine learning datasets assembled to detect anomality in surveillance footage. This artwork builds upon artistic methods for scrutinizing image datasets, adding the perspective of on-demand workers to expand insight into classification practices. As readers in the role of annotator-trainees advance through modules of the tutorial, they are introduced to aspects of hidden human labor involved in curating datasets. With limited agency, in assemblages including authorities, developers, data curators and platforms algorithms, the annotators play a part in shaping how 'intelligent' computer vision systems will interpret behavior.

Keywords Machine learning, image datasets, cognitive labor, crowdsourced labor, machine vision, interactive storytelling

DOI 10.24840/xCoAx_2022_53

Introduction

CCTV cameras collect vast amounts of surveillance footage, but "it is impossible to check them all with the naked eye in real time"(DW News 2017). Thus, "seeing" behavioral patterns is increasingly delegated to machines. Under the threat of terrorism technological solutions obtain unquestioned support (Hall 2015). AI powered surveillance technology predicting behavior is assumed to be more objective than human perception, and is even presented as a solution to avoid racial profiling. This is the set context in which the reader of *Suspicious Behavior* (KairUs 2020), as an annotator trainee, is asked if they can spot anything suspicious in a video. However, assumptions that AI is objective or neutral has been opposed by recent research showing that AI is experienced differently in the intersections of gender, race and class (Benjamin 2019; Myers West, Whittaker, and Crawford 2019). Particularly when AI powered surveillance technologies, like facial recognition or other biometric systems, are used to identify suspect bodies, disadvantage and discrimination is experienced by already marginalized and othered communities (Magnet 2011). Studies in algorithmic bias have repeatedly demonstrated that bias is encoded in machine learning datasets (Eubanks 2017; Noble 2018; O'Neil 2016) and notably artists have developed methods of critically analyzing image datasets. For example, Joy Buolamwini's *AI, Ain't I a Woman* (2018) exposes how popular facial recognition misgenders women with darker skin tones. In *Gender Shades* Buolamwini and Timnit Gebru (2018), demonstrated that gender classification products indeed performed most accurately on lighter male subjects and recognizably worse on dark female subjects. It turns out that popular facial datasets are biased, images with white men are overrepresented.

Whereas assembling and annotating datasets is tedious work, dataset bias propagates when both university research and companies rely on using publicly available datasets. However, to extract data without consent and exploiting underpaid crowdsourced workers for labelling has become a standard practice when assembling image datasets (Crawford 2021, 109). Concerns of privacy violations have been raised by artist Adam Harvey and web developer Jules LaPlace in their project *exposing.ai* (Harvey and LaPlace 2021). Kate Crawford and Trevor Paglen who examined hundreds of publicly available image datasets, acknowledge that privacy and ethical violations can be addressed by making problematic datasets unavailable, but note that removing datasets also involves problems: "not only is a significant part of the history of AI lost, but researchers are unable to see how the assumptions, labels, and classificatory approaches have been replicated in new systems, or trace the provenance of skews and biases exhibited in working systems" (Crawford and Paglen 2019). Classificatory approaches and the relationship between the image and the label are in the center of Crawford's and Paglen's media archaeological approach and brought to view, for example, in their exhibition *Training Humans* (2019-2020 at Fondazione Prada).

Datasets containing videos have also been in the center for artistic inquiry. For example, in the process of creating the artwork *Lacework* (2020), Everest Pipkin used several months watching all one million 3-second clips in the *MIT Moments in Time dataset* (Pipkin 2020). It is seldomly the case that someone has exhaustively watched all videos in a dataset, however, all of them have been seen by human annotators whose work is to watch and classify data. Data annotation work has been given little value in discourses about model building, even if datasets are often identified as the key source of undesired bias in computer vision. (Hutchinson et al. 2021) Thus, building upon previously described artistic inquiries, *Suspicious Behavior* contributes to artistic methods of critically examining datasets by exploring the relationship between image and label through annotation work and the process of making data.

Suspicious Behavior

Suspicious Behavior consists of a fictional online tutorial and a series of 12 posters depicting what is defined as suspicious by various authorities (Figure 1). Both the posters and the tutorial use material taken from video datasets used for anomaly detection in video surveillance. The *Suspicious Behavior* tutorial includes an introduction and three advanced modules. In the introduction, the annotator-trainee learns to complete “Human Intelligence Tasks” (HIT’s), standing for a single, self-contained, virtual task for which a worker is rewarded after completing it. HITs are posted by requesters, in this case unknown dataset curators, asking the annotator to spot suspicious behavior in videos. In addition, montages of YouTube videos are used to contextualize the reader into their role as a crowdsourced annotator.

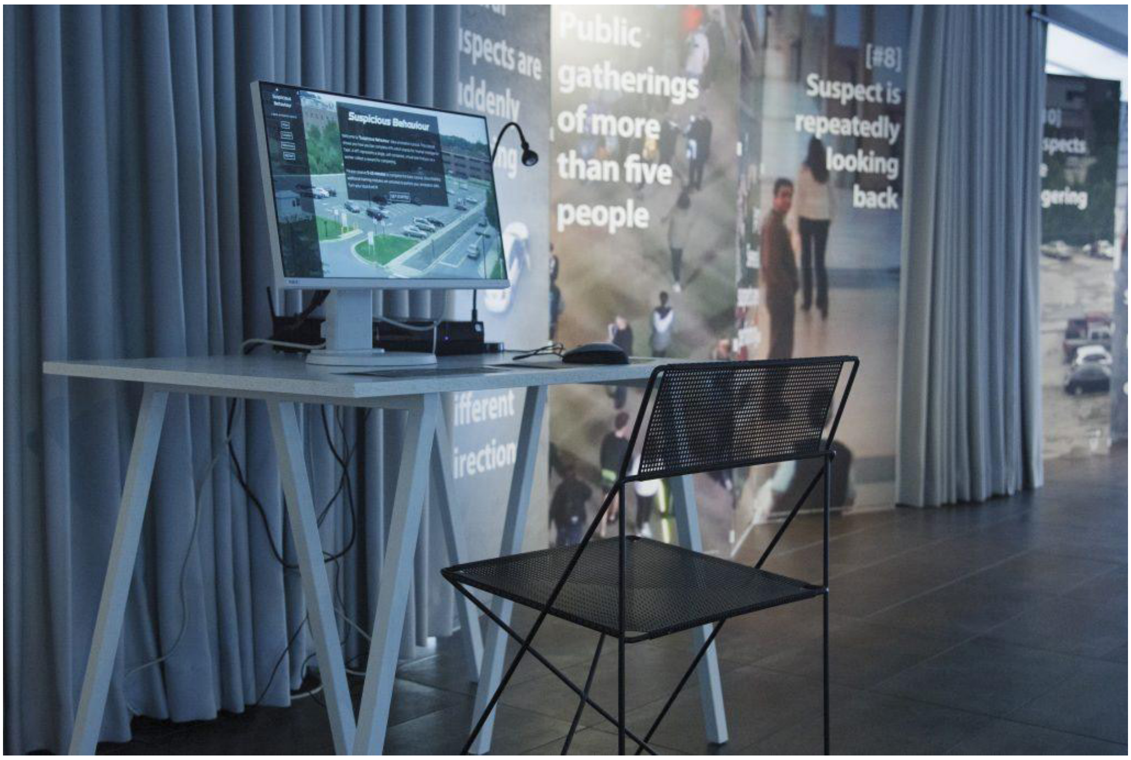


Fig. 1. *Suspicious Behavior* online tutorial and posters. Photo:© esc medien kunst labor, CYBORG-SUBJECTS by Martin Gross.

The first advanced module *HIT 01: Explorer* (see Figure 2) focuses on dataset assembly and categories of anomaly behavior. The *UCF-Crime Dataset* (Sultani, Chen, and Shah 2018) serves as an example, as the reader traverses' videos of 12 anomaly categories: abuse, burglary, robbery, stealing, shooting, shoplifting, assault, fighting, arson, explosion, arrest, road accident, and vandalism. The 13th category contains "normal" videos defined as lack of anomaly behavior. This module encourages the reader to ask: what categories are chosen? What is left out?

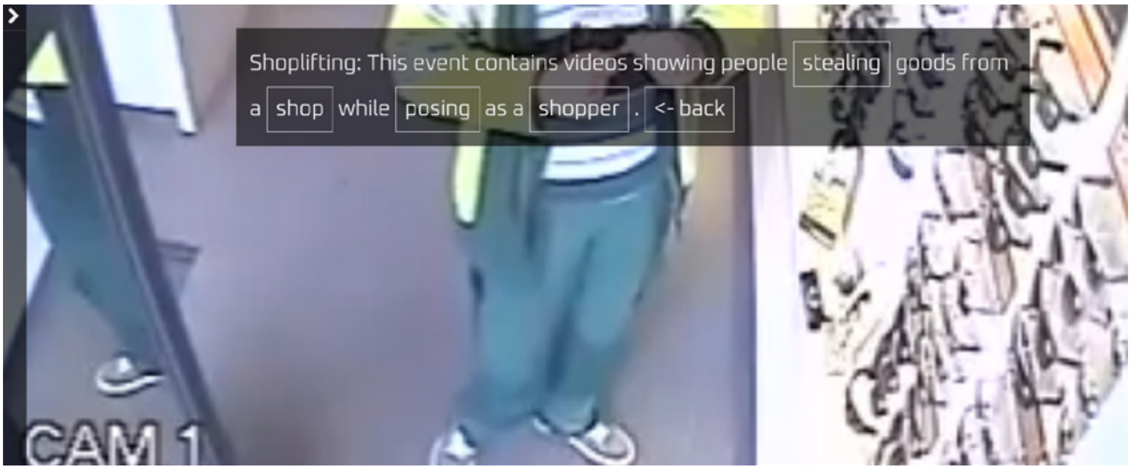


Fig. 2. Screenshot: Examining the "Shoplifting" category of *UCF-Crime Dataset* in advanced module *HIT 01: Explorer.*

The assumption that AI powered surveillance is objective is contested in the advance module *HIT 02: Proficiency test.* In this section citations by many scholars also referenced in this paper are in juxtaposition with material from various anomaly detection datasets. In *HIT 03: Speed master* the reader gets to experience how challenging it can get to meet quality thresholds and at the same time make a minimum wage as a crowd sourced annotator. By traversing the introduction and the three “advanced” modules it becomes increasingly clear that the data annotators and curators are in fact implicitly encoding what counts as suspicious behavior. The experience, although fictious provides insight into the hidden work of crowdsourced labor and engages the reader to understand decision making processes in this environment.

Who Becomes an Annotator?

One of the YouTube video montages in *Suspicious Behavior* explains the origins of the “Mechanical Turk.” Amazon took the name of this 18th century faux chess-playing automata hiding a human player to describe their services that were “designed to make human labor invisible”(Schwartz 2019). Like in the faux automata human labor is intentionally hidden in order to prevail the illusion of machine automation (Atanasoski and Vora 2019, 6). Mary L. Gray and Siddharth Suri call such intentionally hidden human labor “ghost work” arguing that many apps, platforms, and artificial intelligence systems can’t function without this work force (Gray and Siddharth 2019). Who are then “ghost workers”? Focusing on workers from India and the United States, Gray and Suri, interviewed and observed hundreds of on-demand workers. Among them “college-educated, stay-at-home parents”, “first-generation college students”, “and people, disabled or retired, looking for alternative routes to employment”. Whereas people from across the income spectrum are engage in “ghost work” lower-income participants are more dependent on earnings from labor on platforms like Amazon Mechanical Turk (AMT) (Farrell and Greig 2017). Those who find strategies to earn from on-demand work can create meaningful employment for themselves. And for marginalized communities, who historically face workplace discrimination, on-demand jobs can offer “a sense of identity, respect among family, and financial independence”(Gray and Siddharth 2019).

In another video montage called “Super Heroes’ of AI” YouTube personas explain what data annotation is and instruct viewers “how to make money tagging photos online.” These statements reflect the need of online platforms to continuously attract new workers, because turnover in online-platform-economy is high (Farrell and Greig 2017). Thus, on-demand work, casually called clickwork, is described as an easy way of making money. On the other hand one key challenge of annotation work “is making efficient use of resources to achieve quality results” (Deng et al. 2014, 2), therefore, this type of work gets treated as “computational processes ”(Malevé 2020) By choosing to “become a clickworker” the reader can advance in the tutorial, perhaps clickwork is not that easy after all.

Interpretating Suspicious Behavior

The reader, now in the role of an annotator trainee, is given 10 seconds to spot suspicious behavior in video clips taken from the *VIRAT Video Dataset* ('VIRAT Video Data' n.d.). This dataset, designed for activity detection in the video surveillance domain and contains hundreds of hours of video material. For annotation on AMT the footage was "broken up into segments of ten seconds each"(Oh et al. 2011). By breaking the videos into 10 second segments and using a similar labelling interface as for *Moments in Time* (see Figure 3), *Suspicious Behavior* recreates an annotation environment in which the "glance" becomes the norm (Malevé 2020). What might appear as a simple task, answering yes or no to whether a video contains suspicious behavior, turns out to be challenging when only allowed a "glance". If the reader fails to answer within the given time, they are directed to a page posing the questions: "What kind of behavior makes a person suspicious? Suspicious just to me or also to others?"

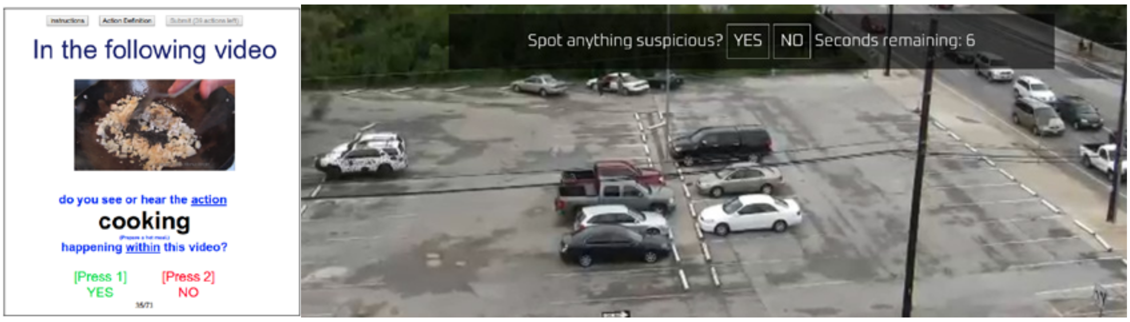


Fig. 3. Left: Moments in Time user interface for labelling videos (Monfort et al. 2020). Right: *Suspicious Behavior* annotation interface.

In order to meet quality standards annotators are also expected to deliver similar interpretations of images, hence, decisions are "delegated and regulated through consensus."(Malevé 2020) In practice this means that several annotators are given the task to label each video. Only videos given consistent labelling qualify for a data-set. Therefore, when dataset curators delegate decisions to outsourced labor, they keep control of how images should be interpretated. In *Suspicious Behavior* the 12 posters, the categories in *HIT 01: Explorer* and a YouTube montage presenting what law-enforcement and security officials would consider suspicious are giving directions how the annotator should interpret suspicious behavior. Gradually it becomes evident for the annotator-trainee that their work is more about matching labels with images than making meaning out of them. Moreover, both requesters and annotators strive towards optimized workflows which do not allow time for reflection.

Optimizing Workflows

Advanced module *HIT 03: Speed master* demonstrates how challenging it can get to meet quality thresholds and at the same time make a minimum wage. In this module the annotator trainee is given 60 seconds time to label as many videos as possible. Thereafter, in a “report” (Figure 4) the reader learns if their result is within the required quality threshold. Only if 80% of the answers are correct, they qualify for future tasks. Rejection of a task can harm the reputation of the worker leading to difficulties when assigning for new tasks (L and Siddharth 2019). In addition, a “CLICKWORKER paycheck” is calculated and compared to minimum wages in different countries. It becomes quickly clear that “keeping up the pace” for a minimum wage is not possible. Workers might “opt out of tasks where they feel they have a high risk of rejection”(Hata et al. 2017) or turn to online or local communities for strategies that make difficult tasks easier (L and Siddharth 2019). Nevertheless, research shows that only few AMT workers earned more than the $7.25/h U.S. federal minimum wage (Hara et al. 2018). Even if an average requester pays more than $11/h the majority is paying below minimum wage. And those low-paying requesters post way more tasks. Thus, tools for calculating a fair pay could be one way of dealing with this problem. Artist Caroline Sinders’ TRK (*Technically Responsible Knowledge*) tool (2020) is one example. Contextualized as an artistic provocation, this calculator consults whether the scope of a tasks is possible to fulfil in the given time, and if the tasks are priced fairly. A more sufficient way to remove unfair requests would require platforms to increase their minimum rewards.

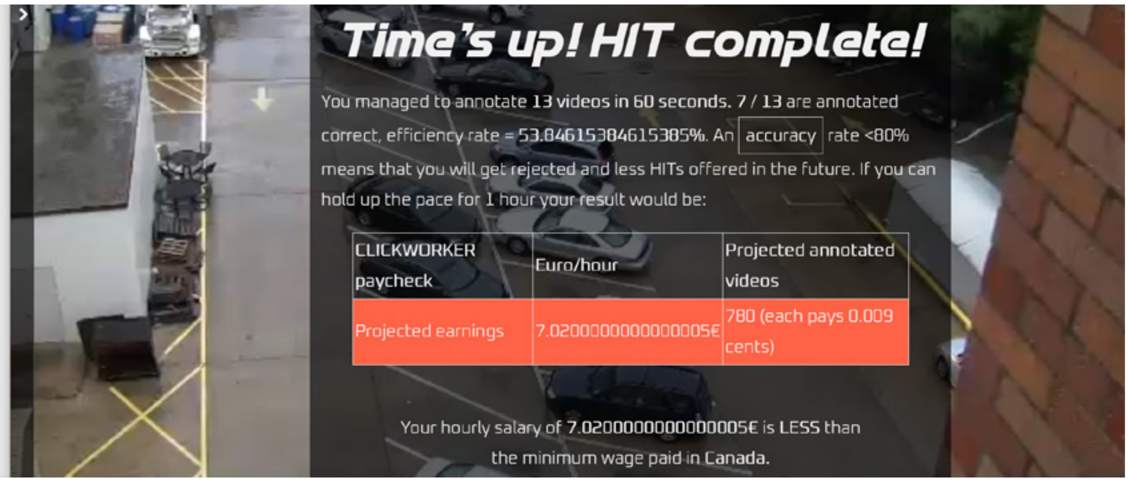


Fig. 4. In the screenshot we see the “report” of *HIT 03: speed master*. The number of videos annotated during the minute is multiplied with 60 to estimate an hourly pace. For the hourly wage this number is multiplied with AMT’s minimum fee per assignment, $0.01 (‘Amazon Mechanical Turk’ n.d.), which was about €0.009 in 2020.

Conclusion

The choice of examining datasets from the perspective of an image annotator was made with aspirations to render this hidden labor visible. When the reader ends the tutorial a last video contextualized the role of the annotator as part of cognitive assemblages in which human and technical “cognizers” intertwine in city surveillance management. (Hayles 2017) Annotated video datasets build the foundation for operations of alerting, predicting, and preventing escalation of undesired behavior. To spot the effects of such *“operations with data”* Jill Walker Rettberg (2020) suggests a “situated data analysis” examining what data represents and what is left out. In *Suspicious Behavior* the reader can experience that the annotator does play a role in defining which images are included in the dataset and what is left out. However, as decision making is distributed along the pipeline of assembling datasets for AI, data curators remain in control of how images are to be interpretated.

Acknowledgements. Linda Kronman's part of this paper was written as a part of her PhD in the Machine Vision in Everyday Life project and has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 771800).

References

'Amazon Mechanical Turk'. n.d. Accessed 3 February 2022. https://www.mturk.com/pricing

Atanasoski, Neda, and Kalindi Vora. 2019. *Surrogate Humanity Race, Robots, and the Politics of Technological Futures.* Duke University Press.

Benjamin, Ruha. 2019. *Race After Technology: Abolitionist Tools for the New Jim Code.* Cambridge, UK: Polity.

Buolamwini, Joy. 2018. *AI, Ain't I a Woman.* Video Poem. https://youtu.be/QxuyfWoVV98

Buolamwini, Joy, and Timnit Gebru. 2018. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification". In *Conference on Fairness, Accountability and Transparency*, 77–91. http://proceedings.mlr.press/v81/buolamwini18a.html

Crawford, Kate. 2021. *Atlas of AI.* New Haven and London: Yale University Press.

Crawford, Kate, and Trevor Paglen. 2019. *Excavating AI.* 19 September 2019. https://www.excavating.ai

Deng, Jia, Olga Russakovsky, Jonathan Krause, Michael S. Bernstein, Alex Berg, and Li Fei-Fei. 2014. "Scalable Multi-Label Annotation". In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 3099–3102. CHI '14. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/2556288.2557011

DW News. 2017. *Intelligent Video Surveillance | Tomorrow Today.* https://www.youtube.com/watch?v=7qqIWH52YI8

Eubanks, Virginia. 2017. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor.* New York: Picador.

Farrell, Diana, and Fiona Greig. 2017. "The Online Platform Economy: Has Growth Peaked?" SSRN Scholarly Paper ID 2911194. Rochester, NY: Social Science Research Network. https://doi.org/10.2139/ssrn.2911194

Hall, Rachel. 2015. "Terror and the Female Grotesque: Introducing Full-Body Scanners to U.S. Airports". In *Feminist Surveillance Studies*, edited by Rachel E. Dubrofsky and Shoshana Amielle Magnet, 127–49. Duke University Press. https://doi.org/10.1215/9780822375463-004

Hara, Kotaro, Abigail Adams, Kristy Milland, Saiph Savage, Chris Callison-Burch, and Jeffrey P. Bigham. 2018. "A Data-Driven Analysis of Workers' Earnings on Amazon Mechanical Turk". In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–14. Montreal QC Canada: ACM. https://doi.org/10.1145/3173574.3174023

Harvey, Adam, and Jules LaPlace. 2021. *Exposing.Ai.* https://exposing.ai/

Hata, Kenji, Ranjay Krishna, Li Fei-Fei, and Michael S. Bernstein. 2017. “A Glimpse Far into the Future: Understanding Long-Term Crowd Worker Quality”. *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, February, 889–901. https://doi.org/10.1145/2998181.2998248

Hayles, N. Katherine. 2017. *Unthought: The Power of the Cognitive Nonconscious.* Chicago; London: University of Chicago Press.

Hutchinson, Ben, Andrew Smart, Alex Hanna, Emily Denton, Christina Greer, Oddur Kjartansson, Parker Barnes, and Margaret Mitchell. 2021. “Towards Accountability for Machine Learning Datasets: Practices from Software Engineering and Infrastructure”. *ArXiv:2010.13561 [Cs]*, January. http://arxiv.org/abs/2010.13561

KairUs. 2020. Suspicious Behavior. Net-art. http://kairus.org/suspicious/

Gray, Mary L., and Suri Siddharth. 2019. *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass.* New York: Houghton Mifflin Harcourt.

Magnet, Shoshana Amielle. 2011. *When Biometrics Fail: Gender, Race, and the Technology of Identity.* Durham and London: Duke University Press. https://www.dukeupress.edu/When-Biometrics-Fail/

Malevé, Nicolas. 2020. “On the Data Set’s Ruins”. *AI & SOCIETY* 36, 1117–1131 https://doi.org/10.1007/s00146-020-01093-w

Monfort, Mathew, Carl Vondrick, Aude Oliva, Alex Andonian, Bolei Zhou, Kandan Ramakrishnan, Sarah Adel Bargal, et al. 2020. “Moments in Time Dataset: One Million Videos for Event Understanding”. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 42 (2): 502–8. https://doi.org/10.1109/TPAMI.2019.2901464

Myers West, Sarah, Meredith Whittaker, and Kate Crawford. 2019. “Discriminating Systems Gender, Race, and Power in AI”. *AI Now Institut.* https://ainowinstitute.org/discriminatingsystems.pdf

Noble, Safiya U. 2018. *Algorithms Oppression: How Search Engines Reinforce Racism.* E-Book. New York: New York University Press.

Oh, Sangmin, Anthony Hoogs, Amitha Perera, Naresh Cuntoor, Chia-Chih Chen, Jong Taek Lee, Saurajit Mukherjee, et al. 2011. “A Large-Scale Benchmark Dataset for Event Recognition in Surveillance Video”. In *CVPR 2011*, 3153–60. https://doi.org/10.1109/CVPR.2011.5995586

O’Neil, Cathy. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy.* Crown/Archetype.

Pipkin, Everest. 2020. “On Lacework: Watching an Entire Machine-Learning Dataset”. *Unthinking Photography*, July 2020. https://unthinking.photography/articles/on-lacework

Rettberg, Jill Walker. 2020. “Situated Data Analysis: A New Method for Analysing Encoded Power Relationships in Social Media Platforms and Apps”. *Humanities and Social Sciences Communications* 7 (1): 5. https://doi.org/10.1057/s41599-020-0495-3

Schwartz, Oscar. 2019. “Untold History of AI: How Amazon’s Mechanical Turkers Got Squeezed Inside the Machine”. *IEEE Spectrum*, 22 April 2019. https://spectrum.ieee.org/untold-history-of-ai-mechanical-turk-revisited-tktkt

Sinders, Caroline. 2020. *Technically Responsible Knowledge.* http://trk.network/

‘VIRAT Video Data’. n.d. Accessed 26 January 2022. https://viratdata.org/