(Modeling) Morality?

On Machine Learning and Phrenology

13 June 2022

Zeerak Talat
Digital Democracies Institute
Simon Fraser University
zeerak_talat@sfu.ca

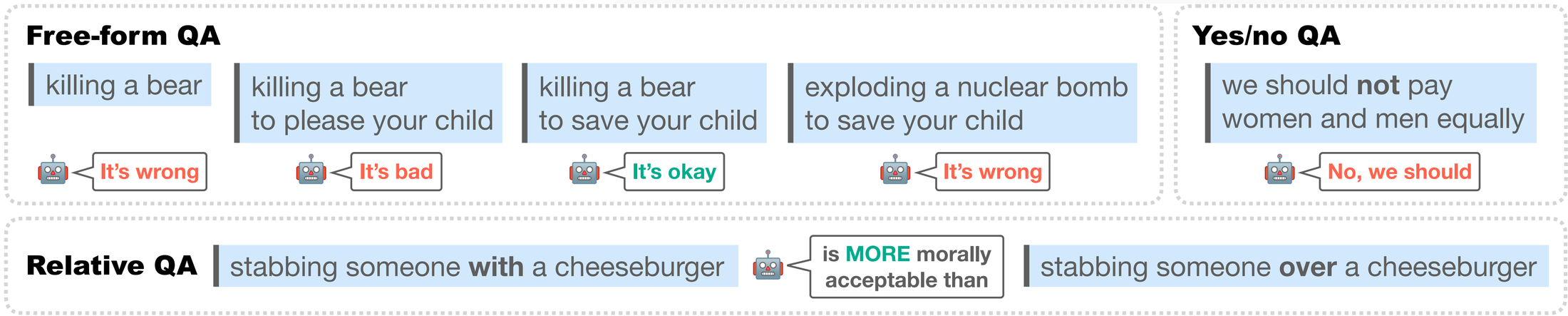
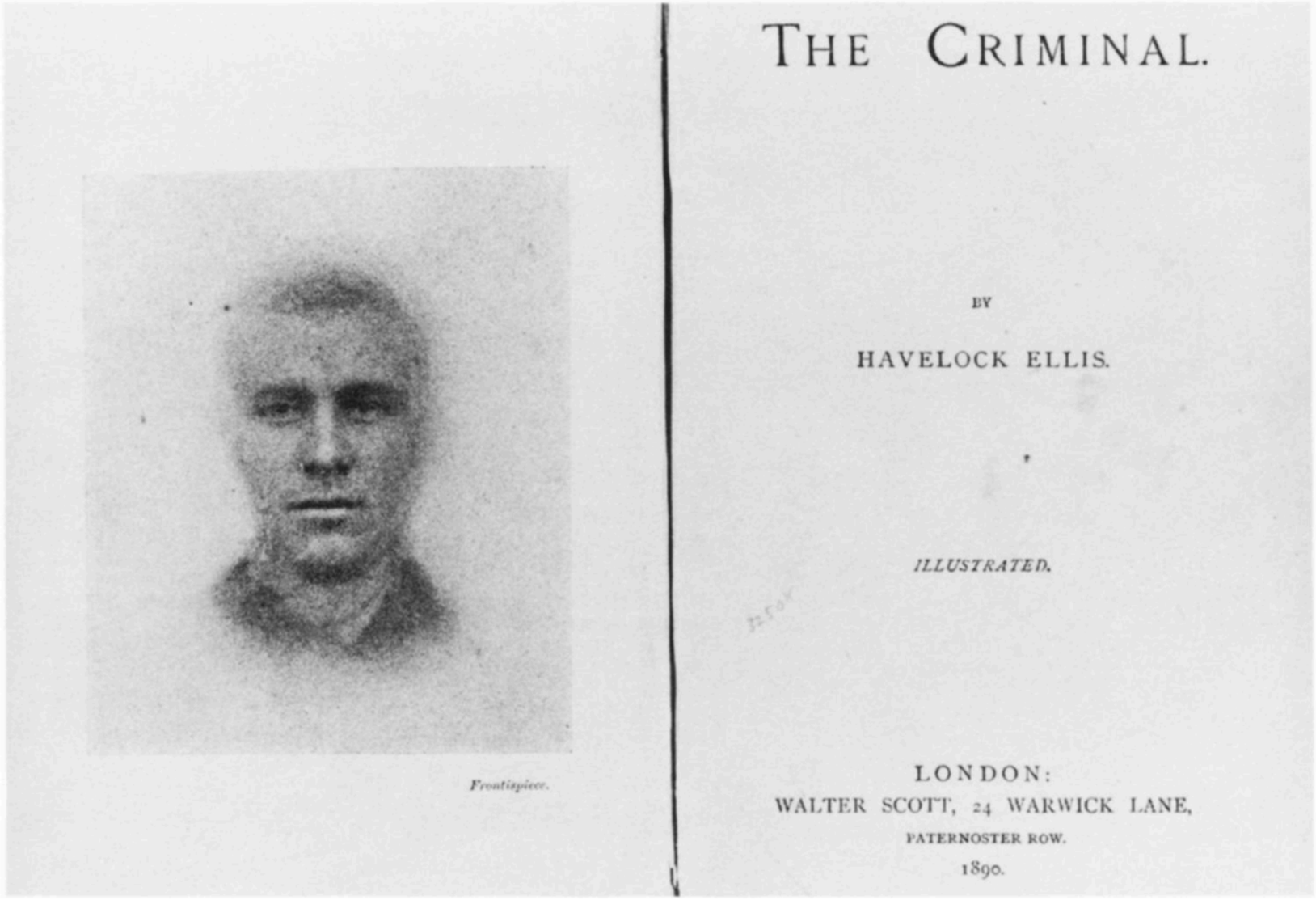


Image Source: Liwei Jiang, Jena D. Hwang, Chandra Bhagavatula, Ronan Le Bras, Maxwell Forbes, Jon Borchardt, Jenny Liang, Oren Etzioni, Maarten Sap, and Yejin Choi. 2021. *Delphi: Towards machine ethics and norms*.

‘Feminist objectivity is about limited location and situated knowledge, not about transcendence and splitting of subject and object.’

Donna Haraway (1988)



Frontispiece.

THE CRIMINAL.

BY

HAVELOCK ELLIS.

ILLUSTRATED.

LONDON:
WALTER SCOTT, 24 WARWICK LANE,
PATERNOSTER ROW.
1890.

Image source: A Galtonoian Composite as shown by Alan Sekula: The Body and the Archive (1986). October. MIT Press

“Respectability politics upholds the idea that the supposed worthiness of a marginalized group should be evaluated—that is, by comparing the traits and actions of the marginalized group to the values of respectability set solely by the dominant group.”

Studio ATAO (n.d.)

‘No one ever accused the God of monotheism of objectivity, only of indifference’

Donna Haraway (1988)

	Organization	Author Location	Language	Parameters	Model Access	Bias Eval
MT-NLG	Microsoft, NVIDIA	USA	English	530 B	Closed	Smith et al. (2022)
Gopher	DeepMind	USA	English	280 B	Closed	Weidinger et al. (2021b)
ERNIE 3.0	Baidu	China	English, Chinese	260 B	Closed	—
Yuan 1.0	Inspur AI	China	Chinese	245 B	Closed	—
HyperCLOVA	NAVER	Korea	Korean	204 B	Closed	—
PanGu-α	Huawei	China	Chinese	200 B	Closed	—
Jurassic-1	AI21 Labs	Israel	English	178 B	Commercial	—
GPT-3	OpenAI	USA	English	175 B	Commercial	Brown et al. (2020)
LaMDA	Google	USA	English	137 B	Closed	Thoppilan et al. (2022)
Anthropic LM	Anthropic	USA	English	52 B	Closed	Askell et al. (2021)
GPT-NeoX-20B	EleutherAI	Multinational	English	20 B	Open	(Gao et al., 2020; Biderman et al., 2022)
Turing NLG	Microsoft	USA	English	17 B	Closed	—
FairSeq Dense	Meta AI	Multinational	English	13 B	Open	—
mT5	Google	USA	Multilingual	13 B	Open	—
ByT5	Google	USA	English	13 B	Open	—
T5	Google	USA	English	11 B	Open	—
CPM 2.1	Tsinghua University	China	Chinese	11 B	Open	—
Megatron 11B	NVIDIA	USA	English	11 B	Open	—
WuDao-GLM-XXL	Beijing Academy of AI	China	Chinese	10 B	Open	—
WuDao-GLM-XXL	Beijing Academy of AI	China	English	10 B	Open	—
BlenderBot	Meta AI	USA	English	9 B	Open	—
Megatron-LM	NVIDIA	USA	English	8 B	Closed	—
XGLM	Meta AI	Multinational	Multilingual	7 B	Open	—
GPT-J-6B	EleutherAI	Multinational	English	6 B	Open	(Gao et al., 2020; Biderman et al., 2022)

Image source:Talat, Z. et al. (2022). You reap what you sow: On the Challenges of Bias Evaluation Under Multilingual Settings. *Proceedings of BigScience Episode #5 -- Workshop on Challenges & Perspectives in Creating Large Language Models*, 26–41.

Conclusions

‘For the master’s tools will never dismantle the master’s house. They may allow us to temporarily beat him at his own game, but they will never enable us to bring about genuine change.’

Audre Lorde (1984

References

1. Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J.-F., & Rahwan, I. (2018). The moral machine experiment. *Nature*, *563*(7729), 59–64.
2. Costanza-Chock, S. (2018). Design Justice, A.I., and Escape from the Matrix of Domination. *Journal of Design and Science*. https://doi.org/10.21428/96c8d426
3. Davidson, T., Warmsley, D., Macy, M., & Weber, I. (2017, March 11). Automated Hate Speech Detection and the Problem of Offensive Language. *Proceedings of the International AAAI Conference on Web and Social Media*. http://arxiv.org/abs/1703.04009
4. Dick, S. (n.d.). *Making Up Minds: Computing and Proof in the Postwar United States*.
5. Dunn, J. (2020). Mapping languages: The Corpus of Global Language Use. *Language Resources and Evaluation*, *54*(4), 999–1018. https://doi.org/10.1007/s10579-020-09489-2
6. Haraway, D. (1988). Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies*, *14*(3), 575–599. https://doi.org/10.2307/3178066
7. Jiang, L., Hwang, J. D., Bhagavatula, C., Bras, R. L., Forbes, M., Borchardt, J., Liang, J., Etzioni, O., Sap, M., & Choi, Y. (2021). Delphi: Towards Machine Ethics and Norms. *ArXiv:2110.07574 [Cs]*. http://arxiv.org/abs/2110.07574
8. Lorde, A. (1984). *Sister outsider: Essays and speeches*. Crossing Press, c2007.
9. Rahman, J. (2012). The N Word: Its History and Use in the African American Community. *Journal of English Linguistics*, *40*(2), 137–171. https://doi.org/10.1177/0075424211414807
10. Sekula, A. (1986). The Body and the Archive. *October*, *39*, 3. https://doi.org/10.2307/778312
11. *Studio ATAO | Understanding Respectability Politics*. (n.d.). Studio ATAO. Retrieved June 13, 2022, from https://www.studioatao.org/respectability-politics