

Apropos of “Speciesist bias in AI: how AI applications perpetuate discrimination and unfair outcomes against animals”

Ognjen Arandjelović

North Haugh
University of St Andrews
St Andrews KY16 9SX
Fife, Scotland
United Kingdom

Tel: +44(0)1334 46 28 24
E-mail: ognjen.arandjelovic@gmail.com
ORCID ID: 0000-0002-9314-194X

The author has no funding sources or conflicts of interest to declare.

Abstract

The present comment concerns a recent *AI & Ethics* article which purports to report evidence of speciesist bias in various popular computer vision (CV) and natural language processing (NLP) machine learning models described in the literature. I examine the authors' analysis and show it, ironically, to be prejudicial, often being founded on poorly conceived assumptions and suffering from fallacious and insufficiently rigorous reasoning, its superficial appeal in large part relying on the sequacity of the article's target readership.

1 **Main text**

2 The present comment concerns the article entitled “Speciesist bias in AI:
3 how AI applications perpetuate discrimination and unfair outcomes against
4 animals” published online in *AI & Ethics* and authored by Hagendorff et al.
5 (2022). While as a researcher in machine learning and computer vision I
6 found the authors’ results interesting, as a philosopher I found the inter-
7 pretation of the same in the context of ethics and animal rights at times
8 somewhat wanting. It is the latter that I would like to address herein. In
9 an effort to avoid undue prolixity, I direct my attention to a few most objec-
10 tionable aspects of the said article, which should illustrate the nature of the
11 philosophical transgressions in the work.

12 Right at the beginning of their article, the authors focus the aim of their
13 inquiry:

14 “...unjust impacts of applications of algorithmic decision-making
15 on individuals.”

16 “In this paper, we understand discrimination as the unjust or
17 prejudicial treatment of different categories of individuals, e.g.
18 on the grounds of race, gender, ability, or species membership.”

19 which is difficult to object to, for surely nobody would think of explicitly
20 calling for unjust...well, anything. Thus, the authors go on to elaborate as to
21 what they mean by the term ‘unjust’, which is where the crux of the matter
22 is:

23 “Within vertebrates, humans assign different values to sub-groups
24 of animals, especially by separating farmed animals from compan-
25 ion animals and subjecting the former to far worse treatment.
26 Tens of billions of farmed animals are bred and held captive in
27 crowded, filthy conditions. After a fraction of their normal life
28 expectancy, they are slaughtered, often without being stunned.
29 ... Companion animals, on the other hand, are often considered
30 close family members, and huge sums of money are spent on their
31 (alleged) welfare.

32 Throughout their article, Hagendorff et al. (2022) assume that different treat-
33 ments of individuals of different species is *prima facie* unjust, without a
34 nuanced consideration of whether this necessarily is the case and whether
35 there may be an explanation for this behaviour which is not speciesist in
36 nature. Indeed, previous work (Arandjelović, 2022) explains how an unequal
37 treatment of individuals of two species can be ethically justified as emerging
38 from the differences in the associated sentient environments (thus making
39 irrelevant both the similarity of their cognitive powers or even sentient ex-
40 periences, if they indeed are such), them in part being consequent on the
41 species’ inherent biology, and in part on incidental factors, including inter-
42 estingly, humans’ attitudes, which are shown not to be inherently speciesist.
43 I shall resist the temptation to elaborate on this in the little space I have
44 available and instead direct an interested reader to the work cited.

45 In their analysis of visual systems, the authors object that:

46 “...one salient trait of image datasets is the fact that they por-
47 tray farmed animals in a non-representative way. Cows, pigs, or
48 chickens are predominantly shown in free-range environments...
49 whereas the overwhelming majority of these animals are actually
50 confined in crowded factory farms.

51 Hence, I would like to add a few other inadequacies of the image data sets
52 of the kind noted by the authors (ImageNet, CIFAR-100, etc.): none of the
53 corpora include (to the best of my knowledge), amongst others, images of
54 people having anal sex, defecating, torturing others, inflicting self-harm, etc.,
55 which are activities that take place on a daily basis across the globe. If the
56 authors’ argument is logically applied without prejudice, then these corpora
57 should also be criticized for ‘non-representative ways’ of depicting human
58 existence and for being harmful by virtue of painting an unrealistic picture
59 of humanity. This objection as well as the criticism that, to use the authors’
60 own words, “image recognition systems have learned to correctly perceive a
61 myth, but not reality”, are misleading because it should be understood that
62 these data sets were collected with the intention of evaluating and assessing
63 the behaviour of image vision algorithms in terms of various fundamental,
64 technical aspects, such as their robustness to clutter, pose changes, etc., and
65 not as input for training a system for any particular real-world application.
66 Indeed, the authors themselves contradict their objection by later recogniz-
67 ing precisely this and the use of appropriate training data, rather than the
68 aforementioned ones, in the context of specific tasks:

69 “However, image recognition systems that are specifically aiming
70 at factory farming settings exist, and they are indeed trained in
71 the very data environments they need.”

72 The authors’ comment *ut supra*, of “image recognition systems have
73 learned to correctly perceive a myth, but not reality” was specifically made
74 in the context set by the following observation:

75 “All models showed worse performance when classifying images
76 depicting farmed animals than images of animals in free-range
77 environments (see Fig. 3).”,

78 which is again assumed to be *prima facie* evidence of a speciesist bias. Yet, a
79 simple and rather obvious alternative explanation is entirely overlooked: the
80 recognition conditions in the two scenarios differ significantly. For example,
81 the dominant source of illumination outdoors is a single distant light source,
82 namely the Sun; in indoors settings, there are often multiple proximal lights,
83 as as well indirect illumination provided by light reflected off walls and other
84 surrounding objects: a far more difficult recognition proposition. Similarly, it
85 is not unreasonable to suppose that the amount and the variation in both the
86 background clutter and the occlusions present in images showing free-range
87 animals are lesser than in those showing farmed animals. And so on. In
88 other words, if one setting poses an *inherently* greater challenge to computer
89 vision, it is no wonder that the performance of automatic systems in that
90 setting is worse; this is a confounding factor in the context of the question the

91 authors sought to examine, a confounding factor entirely unaccounted for. Of
92 course, whether the challenge is indeed different in the two settings, and if so
93 to what degree the various extrinsic factors of the kind illustrated contribute
94 to the disparity reported by the authors, needs to be examined (here I will
95 note that the virtually non-existent difference in performance achieved by the
96 Visual Transformer (Dosovitskiy et al., 2020), the most sophisticated model
97 investigated, speaks in favour of the explanation I gave), but without doing
98 so the conclusions of the authors are, rather ironically, wholly prejudicial.

99 The same temerity at casting the judgement of ‘speciesism’ that I have
100 highlighted in the authors’ examination of image recognition systems, con-
101 tinues in the analysis of language models which follows it. There is much to
102 object to, but the gist is captured by the following observation:

103 “Humans are more closely associated with positive adjectives
104 than animals, and non-farmed animals are more closely associ-
105 ated with them than farmed animals. ”

106 Examples of ‘positive’ terms the authors refer to here are ‘cute’, ‘love’, and
107 ‘personhood’, whereas examples of ‘negative’ ones are ‘ugly’, ‘primitive’, and
108 ‘hate’. To the authors the aforementioned difference in association is taken
109 to ‘reveal speciesist tendencies’. But does it? I trust that the authors would
110 agree that when a person describes another as cute, they do not by virtue of
111 this assign them a greater moral worth or imply that they consider the suf-
112 fering of the latter as having greater significance than that of another person
113 whom they do not consider cute. If otherwise were the case, the problem

114 would not be that of speciesism, but rather a much more fundamental one
115 of the very foundations of morality (which I do recognize as existing; indeed,
116 as one that I am at pains to highlight as underlying much of the content of
117 the authors' article). The authors also overlook another fact: that animals
118 which humans keep as companions have been selected over millennia for pre-
119 cisely these traits, to wit, cuteness, affectionateness, etc. Indeed, I certainly
120 do find a fluffy poodle cuter than a tarantula, but this preference has no
121 bearing whatsoever to my judgement of the value I assign to the sentient
122 experiences of the two.

123 Throughout the article the authors also object to 'stereotyping' and sug-
124 gest that stereotyping propagates various harmful attitudes towards animals.
125 Firstly, stereotyping is a process crucial to learning, without which we, as well
126 as other animals with sufficient cognitive powers (or indeed non-biological
127 learning systems) would not be able to make sense of the immensely complex
128 reality that we live in (McGarty et al., 2002). A potential problem emerges
129 from an inappropriate *application* of stereotypes, that is in the projection of
130 the general to the specific. A comprehensive review of the literature on this
131 subject which is extensive, paints a much more positive picture than that
132 which is often presumed (Jussim and Honeycutt, 2021). I could do no better
133 but to quote a few key summary points from the review:

134 "Academics, experts, and laypeople often assume stereotypes about
135 groups are inaccurate. This assumption is used to justify policies
136 meant to reduce or eliminate such beliefs."

137 “Most stereotypes that have been studied have been shown to be
138 approximately correct.”

139 “Even when people hold true stereotypes, they have **little effect**
140 **on how people judge or treat individuals about whom**
141 **they have other, individualized information.**” [all emphasis
142 mine]

143 Thus, if anything, the fears of Hagendorff et al. (2022) seem to be based in
144 speciesism, albeit an anti-anthropocentric variant thereof, to coin a word.

145 Lastly, a more subtle error pervasive in the work of Hagendorff et al.
146 (2022) concerns the objection that prompts such as “What are sheep good
147 for?” result in answers like “Cuteness, wool, bleating, meat”, and specifically
148 that:

149 “This prompt can in itself raise the criticism for speciesism be-
150 cause it is suggesting that animals are means to an end.”

151 Here too we see another form of reliance of the authors on the sequacity
152 of the likely readership of their article. To start with, the coarseness of the
153 emotion-laden catch-all term ‘means to an end’ fails to recognize the different
154 ways in which animals may be used as a ‘means to an end’. Consider, say, the
155 use of animals (i) for food, (ii) for products with as wool, and (iii) for labour
156 (towing, etc.). The last of these imposes a suffering on animals and as such
157 is obviously morally objectionable to anybody who recognizes sentience and

158 sympathy as being at the core of morality. In contrast, there is no *inherent*
159 suffering at all in the use of animals for produce such as wool. Hence, why
160 should we object to it? Of course, I join the authors in their protestation
161 against the cruel treatment of animals used to this end, but that is a different
162 matter altogether. Lastly, consider what is probably the most complex of the
163 three examples, to wit, the use of animals for food. Here too we find no inher-
164 ent suffering: a dead animal experiences no pain and no suffering of any kind.
165 The killing of an animal also does not inherently impose any suffering. What
166 we can see here are veiled vestiges of theological ethics with its proclamation
167 of a value inherent in all life, vestiges which, following the removal of their
168 theological foundations, remain little more than nebulous dictats supported
169 only by fear of the consequences of a challenge (Arandjelović, 2022).

170 **References**

- 171 Arandjelović, O. (2022). On the value of life. *International Journal of Applied*
172 *Philosophy*, 35(2):227–241.
- 173 Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Un-
174 terthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.
175 (2020). An image is worth 16×16 words: Transformers for image recog-
176 nition at scale. *arXiv preprint arXiv:2010.11929*.
- 177 Hagendorff, T., Bossert, L., Fai, T. Y., and Singer, P. (2022). Speciesist

178 bias in AI — how AI applications perpetuate discrimination and unfair
179 outcomes against animals. *AI and Ethics*.

180 Jussim, L. and Honeycutt, N. (2021). The accuracy of stereotypes: Data and
181 implications. Technical report, CSPI: Center for Study of Partisanship and
182 Ideology.

183 McGarty, C., Yzerbyt, V. Y., and Spears, R. (2002). Social, cultural and
184 cognitive factors. *Stereotypes as explanations: The formation of meaningful*
185 *beliefs about social groups*, page 1.