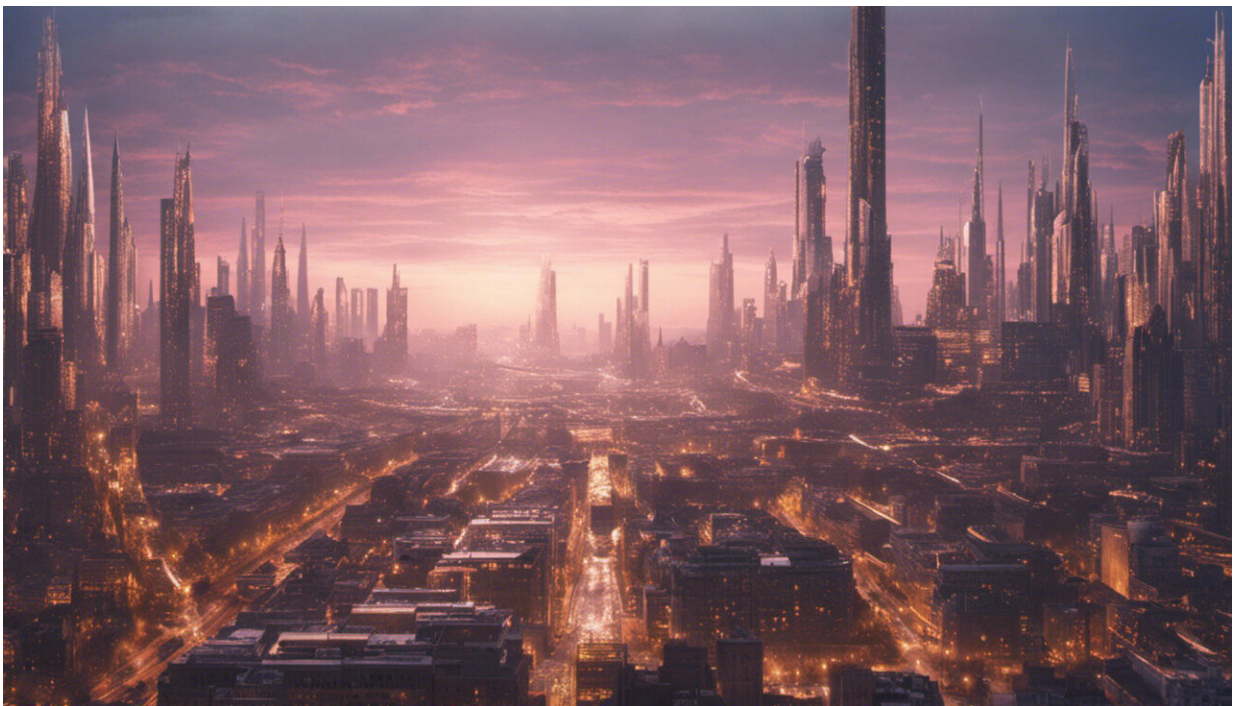


Machine gaydar—AI is reinforcing stereotypes that liberal societies are trying to get rid of

September 13 2017, by Ansgar Koene



Credit: AI-generated image ([disclaimer](#))

Following the old saying that "knowledge is power", companies are seeking to infer increasingly intimate properties about their customers as a way to gain an edge over their competitors. The growth of Artificial Intelligence (AI), algorithms that use machine learning to analyse large

multifaceted data sets, provides an especially attractive way to do this. In particular, the rapid advancement in AI capabilities for pattern discrimination and categorisation are leading researchers to explore its capabilities for increasingly complex data mining tasks.

This technology is no longer restricted to simple categorisation of directly traceable online behaviours ("likes" of particular brands, for example) or image objects (cat vs dog). AI is also being deployed to try to infer such intimate characteristics as [personality, gender and age](#) from [language usage on social media](#), and to use face image analysis to predict the likelihood of someone committing a crime or being a [terrorist or paedophile](#). Most recently, a group of Stanford researchers have used AI to [predict sexual orientation from facial images](#). Clearly the development of such methods for inferring intimate details about people carries strong implications for personal privacy.

A potentially even more problematic aspect of this push towards algorithmic categorisation of people is the accompanying tendency towards simplistic reduction. To train AI to categorise humans, one needs to provide discretely defined target categories and large sets of labelled data. This forces one to reduce complex humans into discrete socio-psychological classes.

Machine gaydar

The [recent study](#) on detecting if a person is gay or straight based on a photograph is a clear example of how the choice of label categories imposes a binary view of sexuality. The aim of the study was to show that faces contain subtle information about sexual orientation that can be perceived and interpreted by deep neural networks (a class of AI).

In order to obtain the large [data sets](#) required for this type of machine learning, they harvested 130,741 [facial images](#) from public profiles

posted on a US dating website. The data set contained gay and heterosexual people in equal numbers, with [sexual orientation](#) established based on the gender of the partners that they were looking for according to their profiles.

Even though the use of a dating site is probably a good indicator of sexual interest in a person, the use of this data to train a binary classifier contradicts the reality of a wide [spectrum of human sexuality](#), ranging from asexual to various degrees of bisexual.

The problem is that once an automated system is shown to be capable of making such a reductionist classification with a high degree of reliability, it becomes a tool that can easily be applied at scale. Categorisation based on this simplified socio-psychological feature becomes an attractive new element to add to all kinds of service personalisation. There is therefore a real danger that such a simplified perspective of people will be further entrenched.

The discussion section of the gay/straight face categorisation paper indicates that the researchers are aware of the larger implications of this kind of work. They go so far as to state that one of the driving motivations for the research was to make "policymakers, the general public, and gay communities aware of the risks that they might be facing already" due to work that "is possible, and likely being done behind closed doors at corporations and government organizations".

Sourcing feedback

Unfortunately, despite this social awareness, the methodology used followed common practice in this field of research, which is to treat any publicly accessible data as "fair game", no matter that the data subjects likely never intended their data to get used for these research purposes. Of course, it might have been difficult to contact the people whose

images were used. But at the very least, spokespeople for the gay community should have been consulted.

In order to truly express their concerns about the impact of this kind of research on peoples' rights, the researchers should have allowed the affected community, in this case LGBT, to directly express their views as part of the discussion section of the paper. This would have changed the way in which the research gets reported in the media and the way in which it is received by the affected community. Such direct stakeholder engagement is one of the key principles of [responsible research and innovation](#), which aims to ensure the sustainability, acceptability and desirability of research processes and outputs.

In order to address the issues of unchecked use of AI for corporate gain it is important to promote a culture of broad stakeholder engagement and ethics within the AI research and development community. The good news is that this is already underway, with ethical [guidelines](#), [initiatives](#) and the development of [ethics based industry standards](#) that aim to provide a means to certify ethical use of AI, similar to food safety standards, on the rise.

This article was originally published on [The Conversation](#). Read the [original article](#).

Provided by The Conversation

Citation: Machine gaydar—AI is reinforcing stereotypes that liberal societies are trying to get rid of (2017, September 13) retrieved 26 April 2024 from <https://phys.org/news/2017-09-machine-gaydarai-stereotypes-liberal-societies.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.