1. Describe 1 potential risk of AI and how we can mitigate it (you can choose any potential risk except singularity/artificial general intelligence/superintelligence). You should describe how AI may bring about that risk, with evidence or reasoning.

The first concern around AI that comes to my mind is the Black boxes problem. Nowadays, AI models are often pretty complex and operate as “black boxes”. These models are created without special designs so that it is extremely difficult for people to know how they actually process data and make outputs. Therefore, they may make unpredictable or even dangerous behaviors in some special cases. Let’s take an autonomous AI as an example. After a model is trained, though it may perform well in tests, how do we prove that the AI will make rational decisions in any case? If we just ignore these concerns and let it drive vehicles, there is a great chance that it may threaten people's lives and property like a timed bomb.

2. (a) what is sentient AI, (b) your take on whether sentient AI is possible and why？

(a) After reading this article, my understanding of sentient AI is that with feelings and subjective will. It has emotions like happiness, anger, sadness, embarrassment, and so on. These feelings, willingness, and emotions may keep changing in the process of it feel the surrounding environment.

(b) I believe sentient AI is possible. The GNW theory in the article emphasizes the importance of human brain function in explaining consciousness. On the other hand, IIT theory infers that the intrinsic causal powers of the human brain really matter. Hence, the former theory believes mimicking the functionality of the brain meaning creates consciousness. However, IIT theory points out an opposite idea, that simulation is not good enough at all. On this question, I have the same idea as the author, which is that science and technology progress can always be exceeded people's imagination. In my opinion, people can make enough detailed simulation brains someday in the future.

3. Mention 5 AI systems that are already here – you can mention models you have interacted with or have read about.

(a) Face recognition

Facial recognition is a method of identifying people's identities using images of their faces.
I have built a small software taking advantage of face recognition. The software takes several photos and requires the email address of the laptop owner. While the computer is being used, the software will take photos periodically and try to recognize the face in them. it would send an email to the previously set email address if it could not find the owner's face to alert him or her that the laptop may be used by other people. Although this is a simple and immature program, I realize the convenience comes with face recognition.

(b) Object classification

Object classification is the task of identifying what an image represents.

(c) Object detection

Object detection refers to the task of locating objects in an image and identify each object.

(d) Object segmentation

Object segmentation is the process of partitioning an image into several segments which changes the representation of the image to something more meaningful and easier to analyze.

(e) Natural Language Processing (NLP)

NLP refers to giving computers the ability to understand the text and spoken words as same as human beings can.

4. What AI system do you envision will be possible in 10 years' time and why. Note that there is no right/wrong answer for this. We are interested in what you think is possible and why.

In my opinion, there is a great chance for autonomous vehicles to be part of people's daily life in the next decade. There have been a great number of achievements being made in the research of autonomous driving. Many major car manufacturers are actively following up and launching the related functions in their products. Various related technologies, like centimeter-scale positioning systems, sensors on vehicles, objects recognition systems, are becoming more mature as well.

5. Read https://nickbostrom.com/ethics/ai.html and discuss (a) what are some of the things we can do to mitigate the risk of a superintelligence system? (b) based on the article, do you think an “AI for good” system has the potential to be bad and why?

(a) Based on the content in the article, a superintelligence system is superior to human beings in every respect. Besides that, a superintelligence system may have extremely fast running/thinking speed, multitasking ability, and other things that people are not capable of. In my point of view, to mitigate the risk of it, people may sacrifice some performance of the system. For example, limiting its intelligence, computing speed, information resource, and so on.

(b) I think an “AI for good” system has the potential to be bad. Like I said in (a), such a system is beyond human. We should treat it as a “super-smart human with unlimited memory” instead of a tool. In this case, people’s desire to control and exploit the system is like monkeys are thinking about exploiting humans. It's doomed to fail because it's against the rules of nature.

No Classification without Representation: Assessing Geodiversity Issues in Open Data Sets for the eveloping World

Shreya Shankar, Yoni Halpern, Eric Breck, James Atwood, Jimbo Wilson, D. Sculley

Summary Presentation

Core idea

Standard open source image data sets may not have sufficient geo-diversity for broad representation across the developing world.

Major approach

In the paper, they analyze

(a) the geo-diversity of 2 popular public data sets: ImageNet and Open Images

(b) the differences that models trained on them exhibit when classifying images from varying geographical locations.

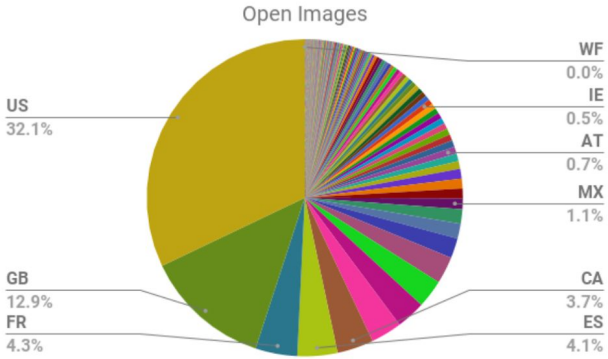
Major approach: Analyzing Geo-Diversity

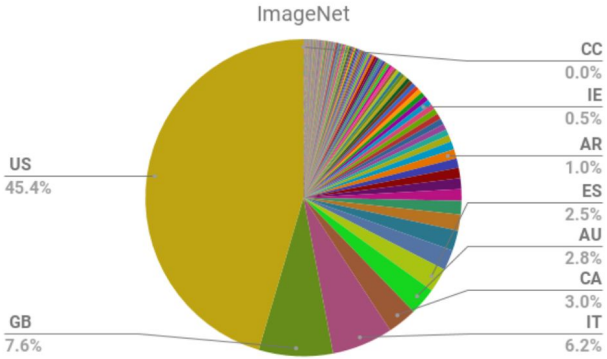
To get location information

(a) Textual / contextual information

(b) URL metadata

Major approach: Analyzing Geo-Diversity





Open Images:

32% images are from the US

60% images are from North America and Europe

1% images are from China

2% images are from India

ImageNet:

45% images are from the US

1% images are from China

2.1% images are from India

PS:China and India – the two most populous countries

Major approach: Analyzing Classification Behavior Based on Geo-Location

The paper collected image data for specific geographical regions using two separate methods

(a) Crowdsourced Data

They asked crowdsourced raters to find and return URLs of images on the internet that matched particular labels, specifically from a community that they identified with in an effort to avoid amerocentric or eurocentric bias.

(b) Geo-located web images

They identified 15 countries to target and joined the per-country location proxy with inferred labels related to “people”, such as bridegroom, police officer, and greengrocer.

Major approach: Analyzing Classification Behavior Based on Geo-Location

Classifier performance on localized data

The paper use two models, one trained on ImageNet and another trained on Open Images to test the difference in their performances between data come from the standard evaluation data split and rater-supplied images.

Result and Discussion

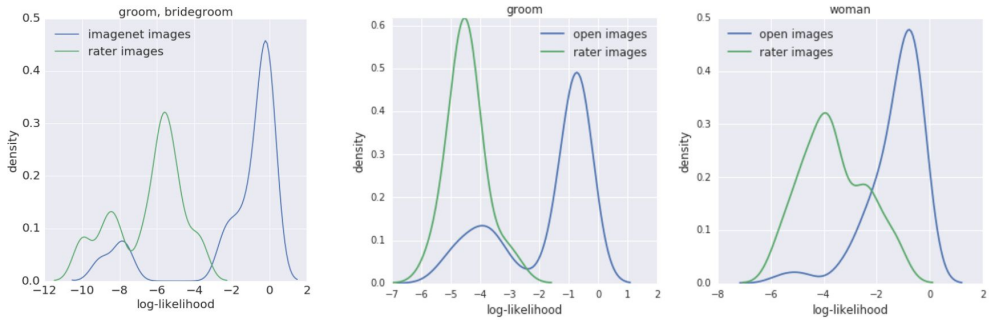
There is an observable amerocentric and eurocentric bias shown in both forms of assessement.

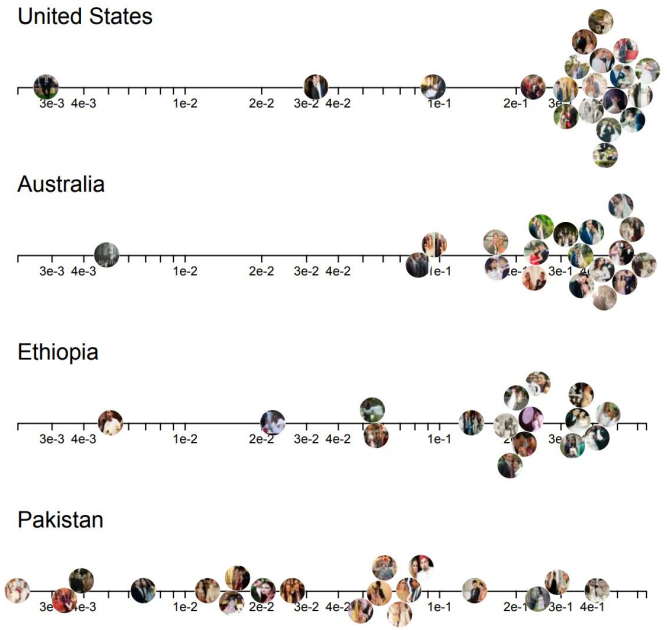
The figure shows some categories that showed noticeable differences in performance.

These differences appear in both classifiers, suggesting that this problem is not particular to a single data set.

The US-based images are clustered to the far right, showing high confidence, while images from Ethiopia and Pakistan are much more uniformly distributed, showing poorer classifier performance.

This trend across several other countries in different regions of the world.





Conclusion

It is clear that standard open source data sets such as ImageNet or Open Images may not have sufficient geo-diversity for broad representation across the developing world.

This study highlights the importance of assessing the appropriateness of a given data set before using it to learn models for use in the developing world.

Equally, this work emphasizes the importance of creating new data sets that prioritize broad geo-representation as first class goals, in order to aid ML in the developing world.

My Opinion

Good

A common problem

Sufficient reasons and evidence

Clear argumentation

Good to have

What should we to do avoid the problem?

How to mitigate the affection of the problem on existing model?