1. This video discusses the statistical measures TPR ("true positive rate"; the video calls this "sensitivity") and PPV ("positive predictive value"), which are commonly used to assess the accuracy of facial recognition AI. Please watch the video and answer the questions below.

(A) Apply this analysis to a (made-up) scenario where AI examines 125 actual M (male) faces and of these finds 110 to be M and 15 to be F (female), and it examines 95 actual F faces and of these finds 65 to be F and 30 to be M. (hint: "+" in video □ "F"). Write the appropriate matrix for this scenario, and calculate the TPR and PPV for F faces.

See attached spreadsheet. Briefly, TPR = 68.42% (predicted female/actual female, PPV = 81.25% (actual female/predicted female)

(B) Explain, in a non-technical way, the conceptual difference between TPR and PPV in the scenario described in Problem 1 (you should understand this, as you'll need to use this in answering a subsequent question).

TPR is the amount (as a percentage) of actual females in the sample that were correctly identified as female, while PPV is the amount (also as a percentage) of identified females that are actually female. An important distinction is that TPR only analyzes the correct identification given that the person is female (the person being female is a **requirement**), while PPV analyzes the correct identification given that the test believes the person to be female while they may not be. In other words, TPR asks: "Out of all the actual females in the sample, how many did the AI correctly recognize as female?", and PPV asks: "Out of all the faces the AI predicted as female, how many were actually female?".

2. This journal article presents the results of one of the most well-known and impactful studies on the accuracy of facial recognition AI. Please read sections 3.2 and 3.3 and page 10, and examine Tables 3, 4 and 5.

(A) What was the profession of the subjects chosen, and why was this choice made? What countries were chosen, and why was this choice made?

They chose "images of parliamentarians", or members of parliament as they are public figures with known identities, and they have "photos available under non-restrictive licenses posted on government websites". In other words, they could legally use the images without violating people's privacy. They chose 3 African Countries (Rwanda, Senegal, and South Africa) as well as 3 European Countries (Iceland, Finland, and Sweden). The countries were chosen due to their ranking in gender parity. Rwanda has the highest proportion of women in parliament, the Nordic regions (Iceland, Finland, and Sweden) are in the top 10 of the same proportion while having lighter skin, and next two highest ranking (in terms of gender parity) African countries (Senegal and South Africa) were chosen to balance for darker skin.

(B) What does Table 3 show, and what is the significance of this data?

Table 3 shows the proportion (distributions) of lighter and darker skin subjects in the 3 datasets (PPB - their's, IJB-A - NIST's, and Adience - a 2014 set). The table shows the skewed nature of the tests that aren't the author's, suggesting more bias/worse results from the IJB-A and Adience studies. The point of this table is to show that their dataset is more evenly distributed between lighter and darker skins at a roughly 50:50 split.

(C) From Table 4, what is the probability that the IBM algorithm correctly identifies a dark-skinned female as female? a light-skinned male as male? Note you must figure out which is the correct data in the table to use... only one type of data is correct to use (see Question 1).

We are investigating the probability that the model identifies F given that the subject is DF, this is the True Positive Rate. Table 4 reports this as 82.3% for the IBM model. For a light-skinned male, the TPR is 94.8%.

(D) In Table 5, the analysis is carried out for only the South African subjects. Why did they do this analysis for only one country, and why did they choose South Africa as this country?

They did a closer look at images from one country to see if differences in the algorithm's performance is mainly due to the image quality from each country's parliament. They chose South Africa as the European countries seemed to have higher resolution images, while the African country's parliament had lower image quality. Specifically, South Africa had similar image quality to the other African Countries while also maintaining relative skin-type diversity within their parliament.

(E) What is the statement the authors make in the last sentence of the conclusion, and do you agree with this statement and do you think engineers working with AI have an ethical obligation to do as they suggest?

The authors say that their data supports the idea that artificial intelligence facial recognition algorithms require increased “demographic and phenotypic transparency and accountability”. What they mean by this is that AI models need to be upfront about their performance and how it varies with different demographics, and that these models should be specific about different physical traits they struggle with. I agree with this, and believe that engineers have an ethical obligation to do this. This is because these models are becoming more and more mainstream, and have been used for more than just presenting data. As facial recognition becomes used in ways that can impact people's lives greatly, developers should take care to ensure that parties making use of the technology are aware of its shortcomings and adapt accordingly.

3. This video features interviews with scientists about detecting people's emotions from their facial expressions. Please watch the video and answer the questions below.

(A) What was the hypothesis developed by the researcher named Sylvan?

His hypothesis was that everyone around the world scowls when angry, smiles when happy, and that everyone around the world can recognize these facial expressions of emotion.

(B) Do the scientists in the video agree with this hypothesis? Explain why or why not.

Not necessarily. They say that it is highly variable, highly contextual, and ever changing patterns. They say that facial expressions are a result of everything going on around them (i.e., who they are with, where they are, what is happening around them). They also emphasize the need for context when it comes to deciphering facial expressions, whereas earlier scientists were just focusing on the face itself. Another reason for the doubt is that there is a distinction between Western and Eastern faces, where people in the west express themselves with their mouths while people in the east express themselves with their eyes. This cultural difference is what makes people present their facial expressions and decipher others differently.

(C) What are the specific reasons given by scientists in this video for why doubt that AI exists that can detect emotion?

They say that people often make facial expressions to show an emotion without actually feeling that emotion inside. An example given is smiling out of politeness or welcoming, but it doesn't mean you're experiencing the emotion itself. The other reason is that people can use false expressions to trick the detection systems, tricking the system into believing they are sad or happy. Regarding this, they say it'd be very easy for a person to generate a realistic, but false expression with their face (which could then trick the system).

4. This journal article presents a study by researchers at George Mason University and the University of Nebraska. Please read the abstract and the associated retraction note, and then answer the questions below.

(A) What is the "new level of image understanding" the researchers are exploring?

This "new level of image understanding" is inferring criminal tendency from facial images via deep learning (neural networks). They are researching using these models to reach the first milestone in inferring personality traits from facial images.

(B) Why was the paper retracted?

The paper was retracted because the authors did not seek approval from their ethics committee before undertaking the study as it uses human biometric data. It is stated that both authors of the paper agree with the retraction.

(C) Do you consider the research topic to be ethical? E.g., should CWRU allow this type of research to be carried out by our faculty, even if all applicable rules are followed?

No, I do not find this to be ethical as there is such variation in human facial expressions. As discussed in previous parts of this week's homework, people's facial expressions vary across regions of the world, and these facial expressions are not necessarily the actual emotion the person is experiencing. Moreover, most models do not assess certain demographics accurately, so certain groups may be targeted incorrectly as the models have unintentional underlying biases. While I wouldn;t riot if I found CWRU to be conducting this type of research, as I think the only way to make it better and improve it is to test it, I do not find it ethical in the grand scope of the topic. AI is notoriously biased, specifically towards its creators, so using it to identify potential criminals off of something as variable as facial expressions is very unethical in my opinion.

5. This journal article presents a study by researchers at Stanford University. Please read the abstract, the Method section for Study 1a, and the last two paragraphs of the article, and then answer the questions below.
(A) How were the images of the subject's faces obtained? How was the "correct answer" for their sexuality known?
They obtained the images from profiles posted on a US dating website, and the sexuality of the subjects was determined by what their profiles reported as the gender they were looking for in a partner.

(B) How successful is the AI for determining sexual orientation? And how successful were humans?
Given one image of the subject's face, the AI could determine whether or not a man was gay 81% of the time, and could discern if a woman was lesbian 71% of the time. Humans were able to predict 61% for men and 54% for women, less than the AI's proportion. They note that this proportion (success rate) increases when given more images to use per person, but only quantitatively describe this for the AI (91% for men, 83% for women).

(C) How might the authors respond to the accusation that their findings would help governments and companies build accurate sexual orientation classifiers? Do you agree with the authors' justifications for making their findings public, and why?
They would respond by saying, "Yes, that is a real concern and we agree with you, but making the public and affected communities aware of these risks is more important than allowing these governments to continue these studies behind closed doors and out of the public eye". I agree with this justification for making this public, especially as a member of the affected community and as someone with many people around me who would also be affected. While it is very well known that the government is doing hundreds of things behind closed doors, seeing this study go public makes the threat of something like this becoming public much more real. Putting people in the know is much more important than hiding something already hidden from them. The authors also note how AI is degrading the privacy of intimate traits, and making this public seems entirely reasonable. I would find myself outraged if I had found something like this was deliberately kept from the public eye, as it would seem to be done out of malicious intent in my opinion.

6. This journal article presents a study by researchers at Dalian University (China). Please read the abstract and the first paragraph of Section 5.4, and examine Fig. 8 and Table 2, and then answer the questions below.

(A) The goal of this research is AI ethnicity recognition. What ethnicities does this study focus on?

This study focuses on (1) Chinese Uyghur, (2) Tibetan, and (3) Korean ethnic groups.

(B) The authors define an "O" region and three "T" regions. What are these regions and which region do the authors find works best for ethnicity recognition?

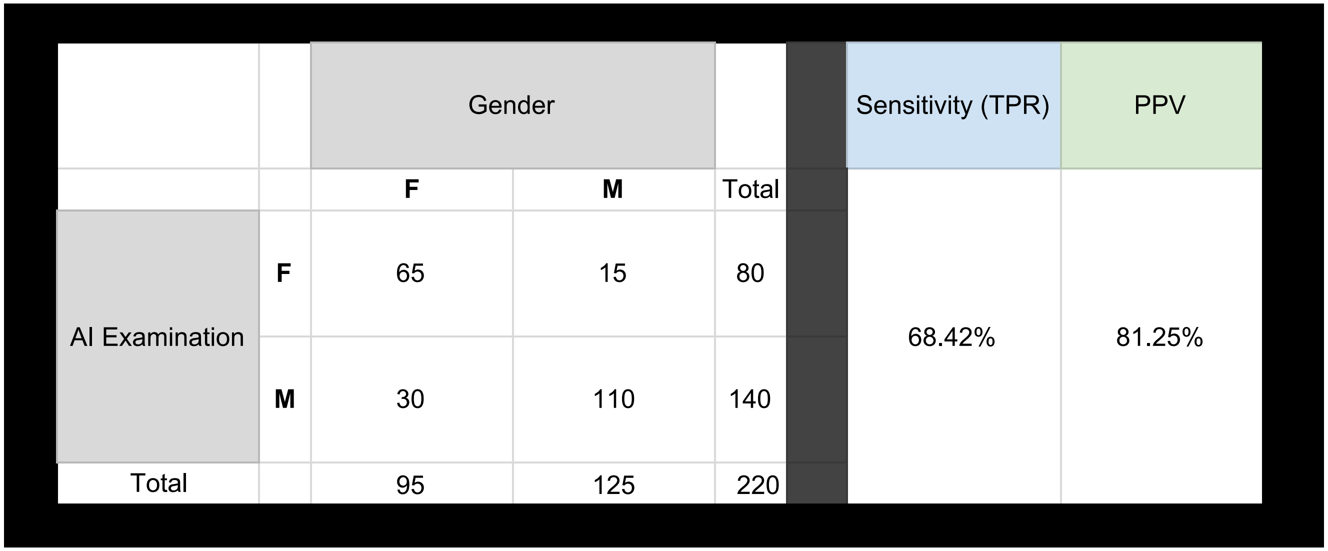
The T region has 3 types of regions denoted T1, T2, and T3. The T1 region includes the eyes and nose; the T2 region contains the eyebrows, eyes, and nose; and T3 contains eyebrows, eyes, nose, and mouth. The O region is defined as the entire face. They found that the T3 region is the best among all regions for ethnicity recognition.

(C) What is the TPR for determining ethnicity (in the best implementation of their method)?

The best implementation of their method was with the T3 region. This region yielded the highest TPR of 0.865, being around 0.10 higher than the T1 and T2 regions, and over double the TPR of the O region.

(D) Do you consider this research to be ethical, and why?

I take a deontological stance here, believing this research to not be ethical. This is because conducting this research violates the privacy of not only the subjects but also those affected by any future technology developed that makes use of this. If this research leads to something mainstream, then the people originally in the study will be part of a baseline for identification, where their facial features will be used to identify others in potentially malicious ways. Ultimately, the ethical obligation of those conducting this research comes down to respecting people's rights, and I think that, if these ethical considerations are taken to heart, then the research should not be conducted at all. Performing this research solidifies peoples facial image in the 'minds' of AI systems that will prevail for years to come.



		Gender			Sensitivity (TPR)	PPV
		F	**M**	Total	68.42%	81.25%
AI Examination	**F**	65	15	80		
	M	30	110	140		
Total		95	125	220		