

**IN THE HIGH COURT OF JUSTICE**  
**QUEEN'S BENCH DIVISION**  
**ADMINISTRATIVE COURT**  
**(CARDIFF DISTRICT REGISTRY)**

BETWEEN

THE QUEEN  
(on the application of EDWARD BRIDGES)

**Claimant**

-and-

THE CHIEF CONSTABLE OF SOUTH WALES POLICE

**Defendant**

- and -

THE SECRETARY OF STATE FOR THE HOME DEPARTMENT

**Interested Party**

---

SECOND EXPERT REPORT OF DR ANIL JAIN

---

I, Dr Anil Jain of [REDACTED]  
[REDACTED], will say as follows:

1. I previously made an expert report on behalf of the Claimant dated 30 September 2018 (my "First Report"). My position as a University Distinguished Professor in the Department of Computer Science & Engineering at Michigan State University and my qualifications remain as set out in my First Report.
2. I make this Report in order to assist the Court by providing factual information with which I am familiar due to my experience. In particular in this Report I respond to statements made by the Defendant in its Detailed Grounds for Contesting the Claim dated 26 November 2018 ("SWP DGC"), and by two of the Defendant's witnesses, Mr Roberts and Mr Edgell, in their witness statements submitted with the SWP DGC.
3. I have been asked by the Claimant to produce this Report. I am not authorised to, and I do

not by saying anything below, waive any privilege on behalf of the Claimant.

4. I have not been paid in any way by the Claimant or anyone else for this Report, my First Report or my expert advice.
5. The contents of this Report are within my knowledge, save where I indicate otherwise. Where the contents of this Report are within my knowledge, I confirm that they are true. Where they are not, I have identified the source of the relevant information, and they are true to the best of my knowledge and belief.
6. I confirm that I understand that I owe a duty to assist the court in matters within my expertise, and that this duty overrides any obligation to those by whom I was instructed. I confirm that I have complied with this duty, and will continue to comply with this duty.
7. I have been made aware of the requirements of Part 35 of the Civil Procedure Rules, Practice Direction 35 of the Civil Procedure Rules, and the Guidance for the Instruction of Experts in Civil Claims 2014, and I confirm that I have complied with these requirements, and will continue to comply with relevant requirements under these rules and guidance.
8. I exhibit to this Report a bundle of documents labelled "AJ 2/x", where 'x' is the exhibit number.
9. References in the form (C/x) and (D/x) are to documents in the Claimant's and Defendant's bundles respectively, where 'x' is the tab number. I adopt definitions in this Report from my First Report.

#### **AFR Locate**

10. In this Report, as in my First Report, I do not comment directly on AFR Locate as I do not know the exact design of that AFR system. However, I am confident that my comments on general AFR systems will also apply to AFR Locate.<sup>1</sup>

#### **Defendant's Grounds**

##### **Face detection**

11. The Defendant states that AFR Locate can only detect up to five faces in any frame (SWP DGC, §12(3)).
12. The processing capabilities of AFR systems vary; the number of faces that are detected in any given frame is generally a variable that is set by agreement between a customer and vendor. It is not a fact that only five faces can be detected in a frame in AFR systems.

---

<sup>1</sup> As explained in my First Report, I draw this conclusion because all AFR systems follow the same general framework (which I set out in my First Report) and because I understand that NEC has developed multiple AFR software solutions around a 'core' face recognition engine which remains the same for all of their solutions.

### Bias and discrimination

13. The Defendant says that two US academic papers relied on by the Claimant (by Buolamwini, and Klare and others) do not indicate that the use of AFR by the Defendant has had any discriminatory impact (SWP DGC, §85).
14. In my First Report I explained the issue of training bias (C/C1, §49). In particular I explained that the under or over representation of a particular demographic in a training dataset will affect the accuracy of an AFR system in identifying faces from that demographic. This issue of training bias is well known, both in the academic community as well as government agencies.
15. I cannot comment on whether AFR Locate has a discriminatory impact as I do not have access to the datasets on which the system is trained and therefore cannot analyse the biases in those datasets. For the same reason, the Defendant is not in a position to evaluate the discriminatory impact of AFR Locate. However, bias has been found to be a feature of common AFR systems.
16. The two US academic papers relied on by the Claimant (C/E3 and C/E18) demonstrate gender and race/ethnicity bias in AFR systems generally. Buolamwini's paper (C/E18) explains that algorithms that classify images by gender and race are poor at correctly classifying darker-skinned female faces (with error rates of up to 34.7%) and are much better at correctly classifying lighter-skinned male faces (with error rates of up to 0.8%). While the paper looks at algorithms that classify images by gender and race, the findings of the paper can be applied to AFR systems in general and imply that such systems will be worse at correctly identifying female and darker-skinned faces (even when they are not classifying by gender or race). Thus when generating matches between faces detected and images on a watchlist, AFR systems will have a higher error rate for women and people from black and ethnic minority groups.

### User-specific thresholds

17. The Defendant notes that in certain situations its AFR software may be set to a "*slightly lower sensitivity setting*" for a particular individual (SWP DGC, §87).
18. What the Defendant is describing is known as using "user-specific thresholds", i.e. setting a particular Threshold Value<sup>2</sup> for an individual on a watch list. As I said in my First Report, using user-specific thresholds is not common in operational settings.
19. It is the end user of an AFR system, not the manufacturer, who sets a user-specific threshold. The end user may choose to set a user-specific threshold if the False Alarm

---

<sup>2</sup> As explained in my First Report, a Threshold Value is a chosen Similarity Score which distinguishes between a potential match and a non-match. The Threshold Value is often chosen in order to fix a too high False Alarm Rate or a too high False Reject Rate.

Rate<sup>3</sup> or False Reject Rate<sup>4</sup> is too high for that particular individual and thus out of sync with the False Alarm Rate or False Reject Rate of others on the watchlist and that which is usually recommended by the manufacturer.

20. To set a user-specific threshold, the end user requires a large number of images of the individual it is being set for. Every combination of those images is then compared by the system (image 1 against image 2, image 1 against image 3, image 2 against image 3, etc., for all the images), and the system generates a Similarity Score<sup>5</sup> for each comparison. The end user uses these Similarity Scores to determine where to set the Threshold Value for the individual so that a particular False Alarm Rate or False Reject Rate is attained.
21. As I stated in my First Report, user-specific thresholds can lead to an AFR system having a discriminatory impact on individuals or certain groups. This can happen from end users purposefully lowering the Threshold Value for, for example, a darker-skinned individual(s) so that the False Alarm Rate for that individual(s) is higher and they are matched more often.
22. However, it can also happen unintentionally. When an end user is adjusting Threshold Values for individuals, it may unintentionally end up making its AFR system particularly sensitive for those individuals, i.e. those individuals will have a higher False Alarm Rate than others on a watchlist. This is the case despite the fact that the point of using user-specific thresholds is to bring the False Alarm Rate and False Reject Rate in line with others on the watchlist. It results from the fact that AFR systems are very sensitive and adjusting Threshold Values can have a significant impact on a False Alarm Rate and False Reject Rate that end users might not fully understand. If a system is particularly sensitive to an individual, that individual will be wrongly matched more often i.e. more people who are not in fact that individual will be wrongly identified as a match.
23. Further, if the Threshold Value for an individual from a black or ethnic minority group is adjusted with the result that the system becomes particularly sensitive to that person, the system may also become particularly sensitive to others from the same ethnic group. As a result, people from that ethnic group will all be wrongly matched more often.

### **First Witness Statement of Mr Paul Roberts**

#### **Training dataset and bias**

24. Mr Roberts notes that, typically on an annual basis, a new version of the NeoFace Algorithm (used by AFR Locate) is released and made available to customers. The new

---

<sup>3</sup> As explained in my First Report, this means the rate at which images of different people can wrongly be considered to be matches.

<sup>4</sup> As explained in my First Report, this means the rate at which images of the same person may wrongly be considered not to be matches.

<sup>5</sup> As explained in my First Report, the Similarity Score is a score that estimates the likelihood that the faces in both images belong to the same person. The Similarity Score is typically in a range 0 to 1, where 0 indicates a perfect non-match, and 1 indicates a perfect match.

version contains improvements such as additional training and no further training is carried out in any customer environments (D/C1, §10).

25. As explained in my First Report and above, the demographic composition of the dataset used to train AFR systems is important in that it affects the accuracy of those systems in identifying people from particular demographic groups (C/C1, §49). If the demographic composition of the training dataset does not match the demographic composition of the population of the place where the system is deployed, then concerns of bias and discrimination arise.
26. Mr Roberts states that the training dataset for the NeoFace Algorithm contains “*roughly equal quantities of male and female faces*”, a “*wide spectrum of ages*” and a “*wide spectrum of different ethnicities*” in order to minimise any impact of bias (D/C1, §§22-24).
27. What Mr Roberts says in paragraphs 22 to 24 of his statement is not sufficient to be able to determine that the NeoFace Algorithm is not biased towards a particular demographic group and thus AFR Locate not discriminatory. To make this determination, a thorough evaluation needs to be done of the demographic composition of the NeoFace Algorithm training dataset.
28. The Claimant’s solicitor has informed me that the Defendant has stated that NEC/Northgate are not prepared to disclose summary statistics (in terms of gender, race/ethnicity, age) or any empirical evaluation of the NeoFace Algorithm training dataset. This is because such information is considered commercially sensitive. However, without this information I cannot assess whether the training dataset is biased. Neither can the Defendant if it also does not have access to this information.

#### Use of biometric codes to identify individuals

29. Mr Roberts states that it is “*questionable*” to suggest that biometric code is capable of identifying an individual not on a watchlist. He says that the biometric template for even the same individual will vary in two different images taken seconds apart (D/C1, §15).
30. It is correct that biometric templates for the same individual may differ depending on variations in pose, expression and illumination, for example. However, if an AFR system is robust, the biometric code should still be capable of identifying the individual.
31. The biometric templates of two different individuals can be very similar, for example with identical twins or for someone wearing a mask, or between two other people with similar facial characteristics and pose, illumination, expression, age, etc.; this is why AFR systems wrongly match different people. However, even with identical twins, the biometric templates of the two twins are only highly similar, they are not identical.<sup>6</sup>

---

<sup>6</sup> For more on identical twins, see Brendan Klare, Alessandra A. Paulino, Anil K. Jain, "Analysis of facial features in identical twins", Biometrics (IJCB), 2011 International Joint Conference on IEEE, 2011 AJ 2/1.

32. In addition, Mr Roberts states that it is impossible to reverse-engineer NEC biometric templates back into facial images (D/C1, §16). However, it has been shown that with general AFR systems it is possible to reverse-engineer facial images from biometric templates.<sup>7</sup>

33. Mr Roberts also mentions that NEC biometric templates are encrypted (D/C1, §16), presumably to make the point that they cannot be stolen and accessed by a hacker. However, during the process of matching facial images, biometric templates will need to be decrypted and are accordingly as vulnerable as any other electronic data.

**First Witness Statement of Mr Dominic Edgell**

34. Mr Edgell concludes that he has seen no gender bias when using AFR technology. Despite there being proportionately more false positive female alerts than false positive male alerts, he explains this as being due to the presence of two “lamb” (D/C4, §§21-23, 26).

35. Before it is possible to draw conclusions on the existence of gender bias, an extensive study needs to be conducted where match scores are thoroughly analysed for both males and females, regardless of whether they generate alerts or not. Mr Edgell does not carry out this study; he considers only alert statistics.

I confirm that I have made clear which facts and matters referred to in this Report are within my own knowledge and which are not. Those that are within my own knowledge I confirm to be true. The opinions I have expressed represent my true and complete professional opinions on the matters to which they refer.

DATED THIS ..... DAY OF JANUARY 2019

SIGNED

DR ANIL JAIN

---

<sup>7</sup> Guangcan Mai et al., "On the Reconstruction of Face Images from Deep Face Templates", IEEE Transactions on Pattern Analysis and Machine Intelligence (2018) AJ 2/2.