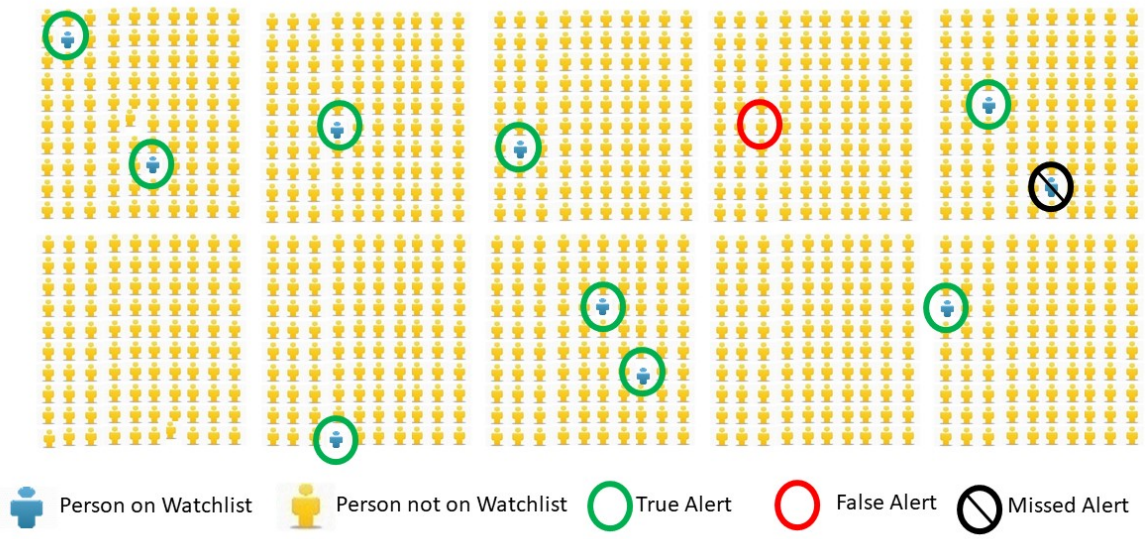
THE METROPOLITAN POLICE SERVICE LIVE FACIAL RECOGNITION SYSTEM; UNDERSTANDING ACCURACY AND BIAS

To assist officers in identifying persons of interest, the Met LFR (Live Facial Recognition) views individuals as they pass the system and alert an officer when the detected facial features closely match those of a person of interest on the Watchlist. An accurate system requires that (i) whenever an individual on the Watchlist passes the system the LFR system should generate an alert and (ii) whenever an individual who is not on the Watchlist passes the system, the LFR system should not generate an alert and should automatically delete all biometric data relating to that individual. In addition to having good accuracy, it is also important that the LFR system performs well in terms of minimising algorithmic / system biases, e.g. any performance differentials between different demographic groups that would disadvantage one demographic group in comparison to another. The Court of Appeal in the *Bridges* case identifies race and gender as being particularly relevant considerations for LFR systems.[1]

A. HOW TO UNDERSTAND LFR SYSTEM ACCURACY

1. It is incorrect to describe the 'accuracy' of a Live Facial Recognition system by a single figure (e.g 98% (in)accurate). Instead, the internationally accepted standards to assess overall system accuracy are determined based on two measures (i) the True-Positive Identification Rate (**TPIR**) and (ii) the False-Positive Identification Rate (**FPIR**).

	True Positive Identification Rate	False Positive Identification Rate
What is it?	Describes: • the total number of times an individual(s) on a Watchlist who is known to have passed the LFR system and correctly generate an Alert; • *as a proportion of* the total number of times those individuals[2] pass the LFR system, regardless of whether an Alert is generated by the LFR system or not. The TPIR is also known as the **True Recognition Rate**.	Describes: • the number of individuals who pass the LFR system, but who are not on the Watchlist and who incorrectly generate an Alert • *as a proportion of* the total number of occasions people[3] pass the LFR system. The FPIR is also known as the **False Alert Rate**.
Worked Example *		

[1] *R (on the application of Bridges) v the Chief Constable of South Wales Police* [2020] EWCA Civ 1058 at (amongst others) paras 164, 176, 181

[2] Given the need to know that an individual has passed the LFR system, this needs to be established using a controlled Watchlist of known individuals and a record kept of when they pass the LFR system. This list is known as a 'bluelist' and those on it are seeded into the passing crowd to establish if the LFR system generates an Alert against them or not.

[3] This is a separate measure to the TPIR and does not need to use a 'bluelist' of known individuals. This is a measure of how often the LFR system incorrectly generates an Alert against an individual who passes the LFR system and that individual is not on the Watchlist. If a person is on a Watchlist or not can be established following an Alert.

	The True Positive Identification Rate would be 90% if 10 people on the Watchlist each pass the LFR system, and a Correct Alert is generated for 9 out of 10 of those people (with no Alert being generated against the 10th person – Missed Alert).	The False Positive Identification Rate would be 0.1%, if for every 1,000 people that passed the LFR system, an Alert was generated against one person who was not on the Watchlist.
	*Simplified to demonstrate the concept of TPIR & FPIR	

B. HOW ACCURATE IS THE MET'S LFR SYSTEM?

2. The National Institute of Standards and Technology (**NIST**) is a physical sciences laboratory and a non-regulatory agency of the United States Department of Commerce. NIST have run open, large-scale Face Recognition Vendor Tests (the **NIST Tests**) to assess the accuracy of facial recognition algorithms since 2004. NIST Tests are normally run on very large 'data sets' of still images (typically between 1.6 and 12 million). These tests allow the Met to compare the baseline accuracy of different algorithms from different vendors. They have also allowed the Met to monitor the improvement of facial recognition accuracy over time.

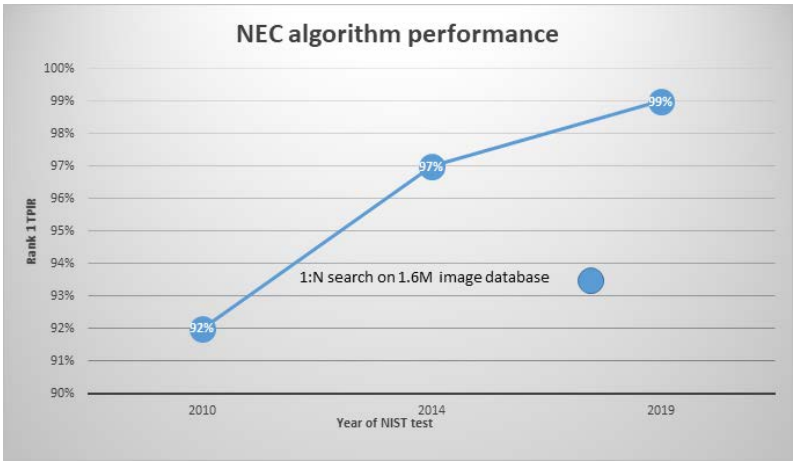
What does NIST tell us about the Met's LFR system?

3. The Met's facial recognition system uses face recognition algorithms supplied by NEC, one of the top performing vendors in NIST FRVT evaluations.

4. The NIST Test report published in 2019[5] evaluated over 200 algorithms for their accuracy. Its findings state that:

> *"NEC, which had produced broadly the most accurate algorithms in 2010, 2013, submitted algorithms that are substantially more accurate than their June 2018 versions and on many measures are now the most accurate".*

5. The results of the NIST Tests for identification[6] since 2010 are shown in the graph. This confirms a Rank 1 - True Positive Identification Rate for the NEC algorithm of 92% in 2010; this increases over time to 99% in the most recent NIST Test. To set this in context, if a search is run against a data set of 1.6 million images, this is the percentage of searches where the correct result comes back as 'Number 1'.



6. In March 2017, NIST also published a Face In Video Evaluation (FIVE) report.[7] Unlike the other NIST Tests, the FIVE test involved the use of video footage as opposed to static images. This is of particular interest to the Met because this aligns more closely to the Met's use of facial recognition in a 'live' - video context. The NEC algorithm was found to be the most accurate across the different measures with a True Positive Identification rate of over 80% at a fixed number of False Positive Alerts.

C. WHAT ABOUT BIAS IN THE MET'S LFR SYSTEM?

7. Consideration of 'bias', i.e. biometric system performance variation across demographic groups, involves measurement of the overall system accuracy to establish the statistical significance of any differences in performance for different demographics groups. The *Bridges* decision identifies gender and race as being particularly relevant to LFR.

8. From 2019, the NIST face recognition vendor tests have started to assess whether demographics such as gender or ethnicity cause FR Identification system accuracy to vary.[8] Tests were run on a 2.6 million image dataset where

[4] https://www.nist.gov/programs-projects/face-recognition-vendor-test-frvt
[5] https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8271.pdf
[6] 1:N Identification
[7] https://nvlpubs.nist.gov/nistpubs/ir/2017/NIST.IR.8173.pdf
[8] https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8280.pdf

images were balanced with respect to representation of gender and ethnicity. The NIST results demonstrate that not all algorithms show uniform accuracy levels across the different demographics. However, NEC, the vendor used by the Met was found to perform well, with NIST saying that NEC had:

> *"provided an algorithm for which the false positive differential was undetectable"* and the NEC-3 algorithm "*is on many measures, the most accurate [NIST] have evaluated*".

9. This provides assurance to the Met as the test showed that the variation in accuracy between male, female, black and white individuals in the NEC face comparison algorithm is imperceptible.

How do the algorithms tested by NIST relate to the Met's LFR system?

10. The algorithm tested by NIST in December 2019 was their NEC-3 variant, which was developed using the same Neo Face technology and the same training data set as the algorithm used by the Met.

11. The NIST Tests provide the Met with robust, transparent, independent and comparable information on how different algorithms including the NEC algorithms perform both in terms of accuracy and bias. However, the NIST Tests do not directly replicate the conditions found when deployed to support law enforcement use cases given differences when using facial recognition in a live operational environment. These include factors such as environmental conditions (lighting, camera positioning), the number and density of subjects passing the LFR system, how subjects behave when passing the LFR system and occlusion given the uncontrolled environment. NIST recognises this and recommends that end users should 'know their algorithm' [in the context of their system and Concept of Operation].

D. WHAT HAS THE MET DONE TO UNDERSTAND THE NEC ALGORITHM IN AN OPERATIONAL ENVIRONMENT?

12. The NIST Tests can only take the Met so far, and by their nature, factors relevant to an operational environment can only be realistically tested with real-life operational use. Further controlled testing would not accurately reflect operational conditions, particularly the numbers of people who need to pass the LFR system in a way that is necessary to provide the Met with further assurance. To that end the Met have tested and continue to test NEC algorithms under operational conditions.

What about accuracy?

13. The performance of the S.17 and M20 versions of the NEC algorithm were evaluated as part of the Met's trials of Live Facial Recognition between 2016 and 2019[9]. The average True Positive Identification Rate (across all trials) was 74% at a corresponding maximum False Positive Identification Rate of 0.1% (or 1 in 1000). This result was in line with the NIST Test results at the time and allows the Met to take confidence in the accuracy of the NEC algorithm, not least as these tests included deployments in some particularly challenging conditions for LFR systems.

What about bias?

14. In addition to the Facial Recognition algorithm, there is a potential for other operational factors, such as Watchlist selection and environmental factors to introduce or contribute to bias. The Met considered the M20 version of the NEC algorithm with respect to accuracy across gender and ethnicity demographics[10]. This crucially used data collected as part of the Met's operational trials as well as a supporting controlled trial – both are necessary to measure the FPIR and TPIR. No statistically significant differences were observed in either the FPIR or TPIR with respect to ethnicity. Differences in FPIR and TPIR recognition rates with respect to gender were statistically significant, with females having a lower FPIR, and lower TPIR than males. The results relate to a single test that showed that the M20 LFR algorithm appeared less likely to trigger alerts in relation to women who pass the LFR system. It also confirmed that the LFR system was not likely to disproportionally generate false alerts on the basis of ethnicity.

15. In line with its commitment to use the latest, most accurate technology, the Met has now upgraded to the latest version of the NEC algorithm. The NIST Tests confirm that the accuracy of the NEC algorithm has continued to improve over time and "the most accurate algorithms produce many fewer errors and these algorithms can

[9] https://www.met.police.uk/SysSiteAssets/media/downloads/central/advice/met/facial-recognition/met-evaluation-report.pdf
[10] https://www.met.police.uk/SysSiteAssets/media/downloads/central/advice/met/facial-recognition/met-evaluation-report.pdf

therefore be expected to have smaller demographic differentials"[11]. Reflective of the need to undertake testing in realistic operational conditions, the Met will continue to monitor its LFR system's performance, both in terms of overall system accuracy and demographic differential performance going forwards.

E. WHAT ABOUT THE VIEWS OF OTHER EXPERTS BEYOND THE MET?

16. The Met's Technology, Research and Innovation unit has considerable expertise in relation to facial recognition technology. Met personnel are members of the Organisation for Scientific Area Committees for Forensic Science Facial Identification Subcommittee[12], sit on and hold an executive position of the Facial Identification Scientific Working Group[13,] are members of the British Standards Institution committee for standards development in biometrics[14] and the International Organisation for Standardisation Committee for Biometrics[15]. Additionally, they advise the National Police Chiefs Council Facial Recognition Technology & Visual Voice ID System Board.

17. To support public confidence and to ensure the Met undertakes resilient analysis in discharging its Public Sector Equality Duty, the Met also recognises the value in seeking the views of others with recognised expertise. The Met's approach outlined in this document has therefore benefited from peer review from the National Physical Laboratory, the UK's national metrology institute responsible for developing and maintaining the national primary measurement standards with recognised expertise in testing biometric systems. It has also sought input from the Defence Science and Technology Laboratory (DSTL), an executive agency of the Ministry of Defence whose stated purpose is to deliver high impact science and technology for the UK's defence, security and prosperity.

Protective marking:	Official
Publication scheme Y/N:	No
Title:	Understanding the Metropolitan Police Service LFR System's Accuracy and Bias Position
Version:	Version 2.0
Summary:	A published summary to assist the public understanding the steps the Metropolitan Police Service has taken to understand the accuracy of its algorithm its performance in relation to bias issues.
Branch/ OCU:	MPS LFR
Review date:	29th November 2022

[11] https://www.nist.gov/speech-testimony/facial-recognition-technology-frt-0

[12] https://www.nist.gov/organization-scientific-area-committees-forensic-science/facial-identification-subcommittee

[13] https://fiswg.org/index.htm

[14] https://standardsdevelopment.bsigroup.com/committees/50087978

[15] https://www.iso.org/committee/313770.html