



Amazon Web Services

Amazon Rekognition - DetectFaces, CompareFaces API

FMR/FNMR

The purpose of this document is to provide an overview of the offline matching technology testing and findings performed by iBeta Quality Assurance ("iBeta") and Amazon Web Services ("Vendor").

iBeta conducted testing on the Amazon Web Services Rekognition DetectFaces and CompareFaces API from 23 February to 11 April 2024, using the biometric data of ~8000 subjects pulled from the Meta Casual Conversations Dataset versions 1.0 and 2.0. Testing was performed for False Match Rate (FMR) and False Non-Match Rate (FNMR) for 6 different demographic bins.

To prepare the dataset for testing, the following filters were applied to the existing biometric samples during readiness review:

- Use Rekognition DetectFaces API and select images with only 1 face.
- Use DetectFaces API and select images where yaw is within the bounds of -30 to 30 degrees, pitch is within the bounds of -30 to 30 degrees (ensures a front facing pose)
- Use DetectFaces API to ensure sharpness > 25 and brightness > 25.
- Use DetectFaces API to ensure OccludedFace = false
- If an image fails the above filters, retry sampling another random image from the video. If no valid image can be sampled after 10 tries, move to the next one.

iBeta created genuine and imposter pairs for groups selected across genders and all six designations of the Fitzpatrick scale for human skin color. Comparisons were only conducted between same gender and within bins of skin tone 1+2, skin tone 3+4, and skin tone 5+6. Error rates were calculated at score thresholds of 80,90,95, and 99 with the CompareFaces API using a 0-100 score system. Bootstrapping methods were applied at a 95% Confidence Interval. Multiple unlabeled duplicate subjects had to be adjusted in the dataset in order to reach the presented results.

The following tables represent the findings for FMR and FNMR at the labeled score thresholds.

Table 1: Observed Errors for False Matches

	Observed Errors – False Matches				
Demographic Bins	Total Matches (imposter)	80%	90%	95%	99%
Female Skin Tone 1 + 2	32752869	440	66	14	14
Female Skin Tone 3 + 4	84910169	1352	262	100	62
Female Skin Tone 5 + 6	27485291	321	34	0	0
Male Skin Tone 1 + 2	15468093	29	0	0	0
Male Skin Tone 3 + 4	46555549	248	34	6	0
Male Skin Tone 5 + 6	21401845	325	39	6	6
Shaded cells have had Rule of 3 applied due to 0 observed errors					

Table 2: False Match Rate Based on Observed Errors

	Observed Error Rates – False Match Rate			
Demographic Bins	80%	90%	95%	99%
Female Skin Tone 1 + 2	0.00134%	0.00020%	0.00004%	0.00004%
Female Skin Tone 3 + 4	0.00159%	0.00031%	0.00012%	0.00007%
Female Skin Tone 5 + 6	0.00117%	0.00012%	0.00001%	0.00001%
Male Skin Tone 1 + 2	0.00019%	0.00002%	0.00002%	0.00002%
Male Skin Tone 3 + 4	0.00053%	0.00007%	0.00001%	0.00001%
Male Skin Tone 5 + 6	0.00152%	0.00018%	0.00003%	0.00003%

	Shaded cells have had Rule of 3 applied due to 0 observed errors
--	--

Table 3: False Match Rate Bootstrapping Results

	95% Confidence Interval (bootstrap method)			
	False Match Rate			
Demographic Bins	80%	90%	95%	99%
Female Skin Tone 1 + 2	0.00146%	0.00025%	0.00006%	0.00006%
Female Skin Tone 3 + 4	0.00168%	0.00035%	0.00014%	0.00014%
Female Skin Tone 5 + 6	0.00130%	0.00017%	0.00001%	0.00001%
Male Skin Tone 1 + 2	0.00003%	0.00002%	0.00002%	0.00002%
Male Skin Tone 3 + 4	0.00060%	0.00010%	0.00002%	0.00001%
Male Skin Tone 5 + 6	0.00168%	0.00024%	0.00500%	0.00500%
	Shaded cells have had Rule of 3 applied due to 0 observed errors			

Table 4: Observed Errors for False Non-Matches

	Observed Errors FNM				
		Errors based on Score threshold			
Demographic Bins	Total Matches (genuine)	80%	90%	95%	99%
Female Skin Tone 1 + 2	31964	146	146	147	160
Female Skin Tone 3 + 4	39190	214	222	225	274
Female Skin Tone 5 + 6	32048	215	218	220	281
Male Skin Tone 1 + 2	24964	94	106	107	120

Male Skin Tone 3 + 4					
	29893	84	85	86	89
Male Skin Tone 5 + 6					
	27306	230	237	238	248

Table 5: False Non-Match Rate Based on Observed Errors

	False Non-Match Rate			
Demographic Bins	80%	90%	95%	99%
Female Skin Tone 1 + 2	0.45676%	0.45676%	0.45989%	0.50056%
Female Skin Tone 3 + 4	0.54606%	0.56647%	0.57413%	0.69916%
Female Skin Tone 5 + 6	0.67087%	0.68023%	0.68647%	0.87681%
Male Skin Tone 1 + 2	0.37654%	0.42461%	0.42862%	0.48069%
Male Skin Tone 3 + 4	0.28100%	0.28435%	0.28769%	0.29773%
Male Skin Tone 5 + 6	0.84231%	0.86794%	0.87160%	0.90823%

Table 6: False Non-Match Rate Bootstrapping Results

	95% Confidence Interval (bootstrap method)			
Demographic Bins	80%	90%	95%	99%
Female Skin Tone 1 + 2	0.53000%	0.53000%	0.54000%	0.58000%
Female Skin Tone 3 + 4	0.62000%	0.64000%	0.65000%	0.78000%
Female Skin Tone 5 + 6	0.76000%	0.77000%	0.78000%	0.98000%
Male Skin Tone 1 + 2	0.45000%	0.51000%	0.51000%	0.57000%

Male Skin Tone 3 + 4	0.34000%	0.34000%	0.35000%	0.36000%
Male Skin Tone 5 + 6	0.95000%	0.98000%	0.98000%	1.02000%

Beyond the above results, the False Non-Match Rate was calculated for the overall dataset at a False Match Rate of 1:1 million and 1:10 million. At the overall limit of False Match Rate at 1:1 million, the False Non-Match Rate is 0.55% at a Score Threshold of ~92.525. At the overall limit of False Match Rate at 1:10 million, the False Non-Match Rate is 0.81% at a Score Threshold of ~99.675.



Ryan Borgstrom
iBeta Director of Biometrics
RBorgstrom@ibeta.com



David Yambay
iBeta Deputy Director of Biometrics
DYambay@ibeta.com