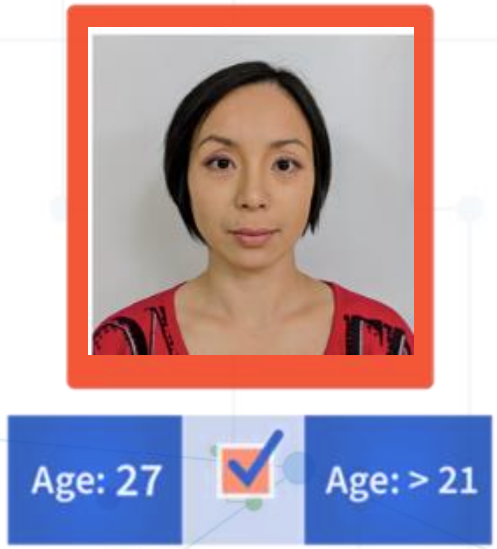
FATE: Age Estimation & Verification Evaluation



Kayee Hanaoka, NIST

IFPC April 1, 2025

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

NIST Face Technology Benchmarks

NIST NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY U.S. DEPARTMENT OF COMMERCE

FRTE

FACE **RECOGNITION** TECHNOLOGY EVALUATION

RECOGNITION: **WHO** IS IN AN IMAGE

1:1 VERIFICATION	SAME PERSON OR NOT?
1:N SEARCH	WHO? WHERE? WHEN?
FACE IN VIDEO 2024	1:N ON NON-COOP PEOPLE
TWINS DISAMBIGUATION	SAME PERSON, OR TWIN?

FATE

FACE **ANALYSIS** TECHNOLOGY EVALUATION

ANALYSIS: **ABOUT** IN AN IMAGE

MORPH DETECTION	TWO PEOPLE IN ONE FACE?
QUALITY + DIAGNOSTICS	HOW BAD IS THIS PHOTO
PAD	SUBVERSIVE PHOTO?
AGE ESTIMATION	HOW OLD? OLD ENOUGH?

Benchmarks are:

- Independent
- Public
- Free
- Open globally
- Regular
- Fast
- Repeatable
- Fair
- Black box
- IP-protecting
- Large-scale
- Sequestered datasets
- Statistically robust
- Public
- Transparent
- Extensible
- **ABSOLUTE ACCU**
- **RELATIVE ACCU**

2014-03-20:
NISTIR 7995 Performance of Automated Age Estimation

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

Face Recognition Vendor Test (FRVT)

Performance of Automated Age Estimation Algorithms

NIST Interagency Report 7995

M. Ngan and P. Grother

Information Access Division
National Institute of Standards and Technology

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

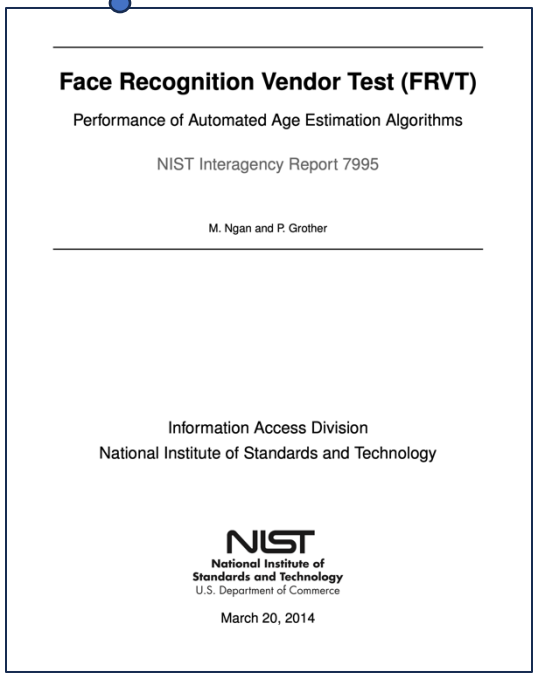
March 20, 2014

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

2014-03-20:
NISTIR 7995 Performance of Automated Age Estimation

2023-08-14:
Publish v1 API

API functions



Face Recognition Vendor Test (FRVT)

Performance of Automated Age Estimation Algorithms

NIST Interagency Report 7995

M. Ngan and P. Grother

Information Access Division
National Institute of Standards and Technology

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

March 20, 2014

#1

#1: ESTIMATE_AGE

FROM SINGLE IMAGE

⟶ AGE = 36.4

#2

#2: VERIFY_AGE > 25

FROM SINGLE IMAGE

⟶ TRUE

#3

#3: ESTIMATE_AGE

GIVEN

+

⟶ AGE = 36.4

BEYOND SCOPE: AGE ASSURANCE ⟶ TRUSTED AV including LIVENESS + PAD + FALLBACKS

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

2014-03-20: NISTIR 7995 Performance of Automated Age Estimation

2023-08-14: Publish v1 API

2023-09-05: Open for submission

API functions

Face Recognition Vendor Test (FRVT)

Performance of Automated Age Estimation Algorithms

NIST Interagency Report 7995

M. Ngan and P. Grother

Information Access Division
National Institute of Standards and Technology

NIST National Institute of Standards and Technology U.S. Department of Commerce

March 20, 2014

#1

#1: ESTIMATE_AGE

FROM SINGLE IMAGE

⟶ AGE = 36.4

#2

#2: VERIFY_AGE > 25

FROM SINGLE IMAGE

⟶ TRUE

#3

#3: ESTIMATE_AGE

GIVEN

+

⟶ AGE = 36.4

BEYOND SCOPE: AGE ASSURANCE ⟶ TRUSTED AV including LIVENESS + PAD + FALLBACKS

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

2014-03-20: NISTIR 7995 Performance of Automated Age Estimation

2023-08-14: Publish v1 API

2023-09-05: Open for submission

2024-05-29: NISTIR 8525 FATE-AEV

API functions

Face Recognition Vendor Test (FRVT)

Performance of Automated Age Estimation Algorithms

NIST Interagency Report 7995

M. Ngan and P. Grother

Information Access Division
National Institute of Standards and Technology

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

March 20, 2014

#1

#1: ESTIMATE_AGE

FROM SINGLE IMAGE

⟶ AGE = 36.4

#2

#2: VERIFY_AGE > 25

FROM SINGLE IMAGE

⟶ TRUE

#3

#3: ESTIMATE_AGE

GIVEN

+

⟶ AGE = 36.4

NIST Interagency Report
NISTIR 8525

Face Analysis Technology Evaluation: Age Estimation and Verification

Kayee Hanaoka
Mei Ngan
Joyce Yang
George W. Quinn
Austin Hom
Patrick Grother
Information Access Division
Information Technology Laboratory

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8525

NIST IR May 2024
INCLUDES UPDATES AS OF 2025-02-28

U.S. Department of Commerce
Gina M. Raimondo, Secretary

National Institute of Standards and Technology
Laurie E. Locascio, NIST Director and Under Secretary of Commerce for Standards and Technology

BEYOND SCOPE: AGE ASSURANCE ⟶ TRUSTED AV including LIVENESS + PAD + FALLBACKS

2014-03-20: NISTIR 7995 Performance of Automated Age Estimation

2023-08-14: Publish v1 API

2023-09-05: Open for submission

2024-05-29: NISTIR 8525 FATE-AEV

2024-06 to 2025-03: 8 report updates

22 Algorithms

- cvut_001
- cvut_002
- cyberlink_000
- dermalog_001
- dermalog_002
- hzailu_001
- idemia_001
- inbut_000
- incode_000
- innovatrics_000
- jumio_000
- neurotechnology_000
- neurotechnology_001
- nominder_000
- ondato_000
- ondato_001
- roc_000
- roc_001
- unissey_001
- viante_000
- yoti_001
- yoti_002

16 Developers

- Brno University of Technology-FIT DCGM(1)
- Cyberlink Corp(1)
- CzechTechnical University in Prague (2)
- Dermalog(2)
- Hangzhuo AIlu Network (1)
- Idemia(1)
- Incode Technologies Inc(1)
- Innovatrics(1)
- Jumio(1)
- Neurotechnology(2)
- Nominder(2)
- Ondato(2)
- ROC(2)
- Unissey(1)
- Viante.AI(1)
- Yoti(2)

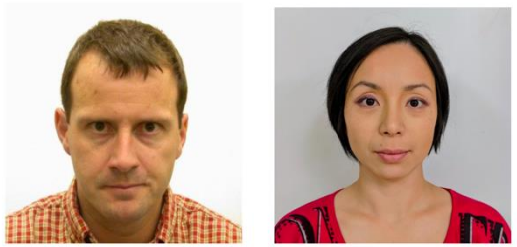
10 Countries

- United States
- United Kingdom
- France
- Germany
- Czech Republic
- Lithuania
- Slovakia
- Taiwan
- United Arab Emirates
- China

FATE AEV Datasets

Name	Age Precision	Num. of Images	Num. People	Purpose
Visa	Day	6,249,294	5,738,091	Exact repeat of 2014 study
Application	Day	1,054,704	802,332	Challenge-T and country of birth
Border	Day	2,715,230	632,520	Analysis of effect of quality
Mugshots	Year	1,482,667	1,482,667	AE accuracy on standardized photos
Kalina Everyday	Day	1991	1	Longitudinal ageing

High
Uniform Quality
Global Population



Application and Visa Photos

Border Photos

Mugshot Photos

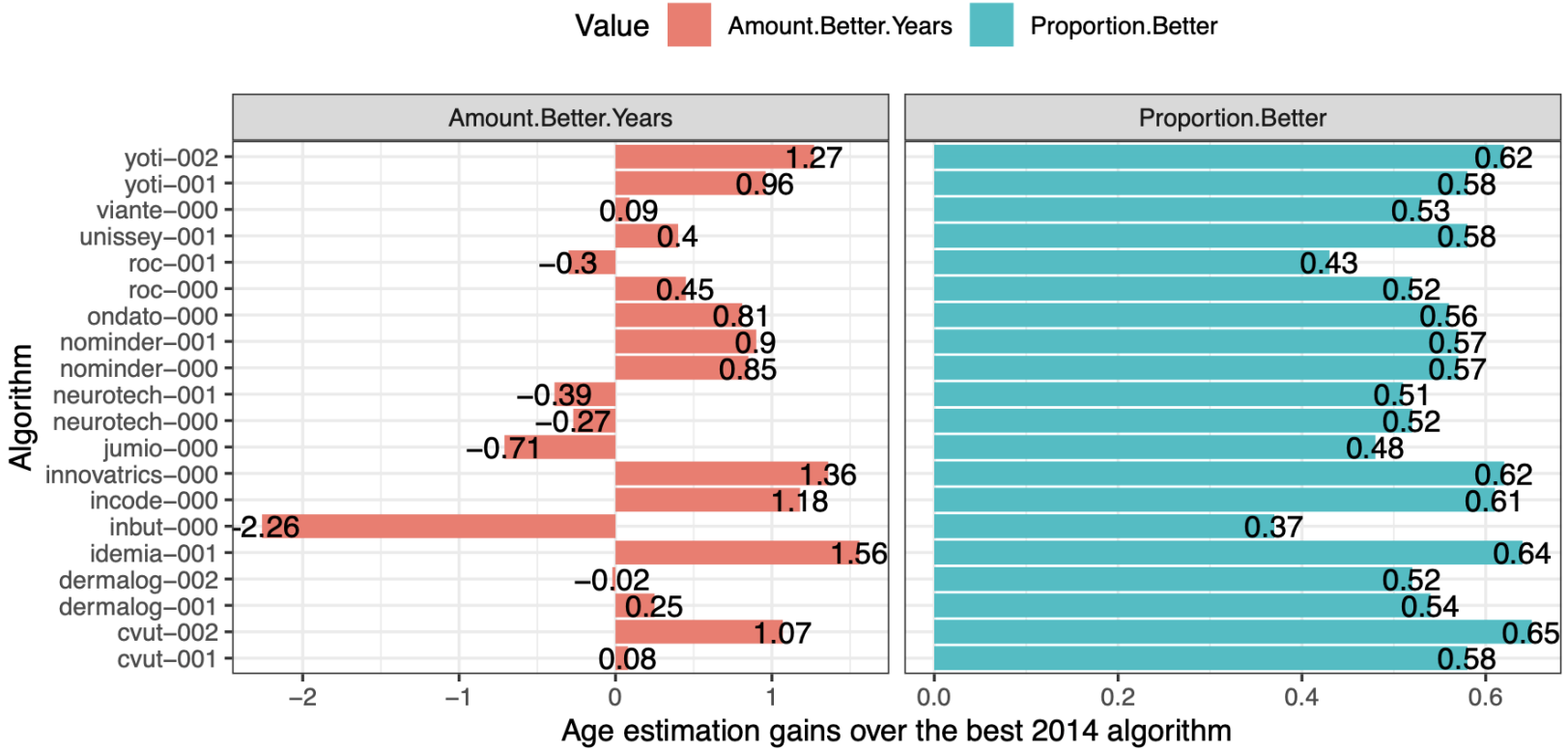
New Dataset	Num. of Videos	Num. of Frames	Num. People
Frontal Videos	482	434,230	482
Yaw Videos	29	26,056	29

AE Progress Since 2014

Visa	6,249,294 images	5,738,091 people	Exact repeat of 2014 study

Gains since 2014: Mean and median absolute estimation error for the new algorithms + two most accurate from 2014.

Algorithm	MAE	MDAE	ME	SD
idemia-001	2.70	2.07	1.23	2.37
yoti-002	2.99	2.19	1.43	2.79
incode-000	3.08	2.33	1.48	2.74
innovatrics-000	2.90	2.36	1.12	2.39
yoti-001	3.30	2.46	1.74	2.95
cvut-002	3.19	2.53	1.05	2.72
nominder-001	3.36	2.71	1.65	2.79
nominder-000	3.41	2.75	1.26	2.83
ondato-000	3.46	2.84	1.06	2.82
cvut-001	4.18	3.00	0.01	5.08
unissey-001	3.87	3.00	0.87	3.35
roc-000	3.81	3.12	2.64	3.05
dermalog-001	4.01	3.21	2.46	3.29
cognitec2013Oct-001	4.27	3.25	0.30	3.86
viante-000	4.17	3.39	3.20	3.29
dermalog-002	4.29	3.42	2.27	3.55
neurotechnology-000	4.54	3.52	-0.19	3.96
neurotechnology-001	4.65	3.69	-0.09	3.98
jumio-000	4.98	3.77	3.61	5.05
nec2013Oct-002	5.32	4.19	0.31	4.61
roc-001	4.55	4.27	4.00	2.88
inbut-000	6.49	5.55	5.82	5.03



Mean reduction in absolute age estimation error - negative values show worse performance than 2014.

Fraction of images that have better estimates than the best 2013-10 algorithm, Cognitec-001

FATE AEV Leaderboard/Webpage

NIST NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY U.S. DEPARTMENT OF COMMERCE

Resource Cost | MAE By Dataset and Variability | Challenge-25 & Child Online Safety (13-16) | MAE Across Geography

Resource Usage:

- AE Time (msec) is the median duration of the AE function call.
- Config (MB) is the total size of the neural network model and configuration data of the algorithm.
- Lib (MB) is the total size of the implemenation's compiled libraries.

Show 30 entries

Search: ex. 2027|alphanum

Algorithm	Submission Date	Resources		
		Estimation Duration (msec)	Config Model Size (MB)	Libraries Size (MB)
incode-000	2023-09-05	236	457	40
unissey-001	2023-09-19	113	0	952
yoti-001	2023-09-19	268	227	429
neurotechnology-000	2023-10-17	153	113	96
dermalog-001	2023-11-02	27	0	188
roc-000	2023-11-03	134	0	474
cvut-001	2024-06-14	813	2,887	958
neurotechnology-001	2024-07-19	97	53	120

FATE AEV Leaderboard/Webpage

NIST NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY U.S. DEPARTMENT OF COMMERCE

Resource Cost | MAE By Dataset and Variability | Challenge-25 & Child Online Safety (13-16) | MAE Across Geography

Mean Absolute Error (MAE) values are averages of 26 MAE estimates from two sexes and 13 ages (18 to 30).

- Algorithm name
- Submission data
- MAE for frontal view, no-glasses, office application photos.
- MAE for airport concourse border crossing photos featuring variable pose, illumination and eye glasses.
- MAE for good quality mugshot portraits collected with two diffuse light sources and a uniform background.

Age Estimation Noise

- Proportion of frontal video frames for age estimate was not produced
- Mean absolute deviation of frontal age estimates about their mean - see section 2.7 in the report
- Proportion of left-right yaw video frames for which age estimate was not produced
- Modelled increase in age estimation error for 90 degrees of head rotation - see section 2.8 in the report

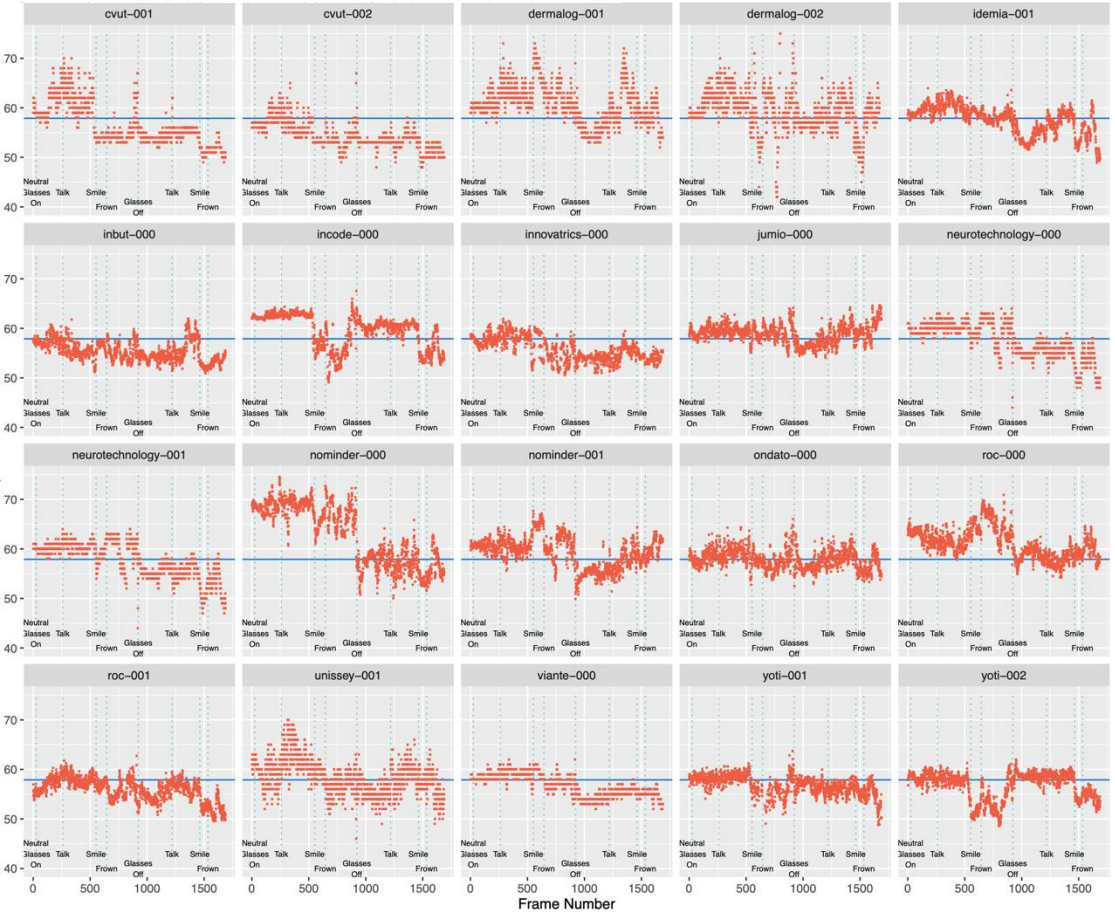
Show 30 entries

Search: ex. 2023|insight

Algorithm	Submission Date	MAE			Variability in video and attributable to non-frontal pose			
		Application	Border	Mugshots	FTE Frontal	MAE Noise Frontal	FTE Yaw	sin(Yaw) Dependence
roc-000	2023-11-03	3.564(13)	4.015(16)	2.998(7)				
viante-000	2024-11-25	3.969(21)	3.458(11)	4.028(18)	0.00	0.37(1)	0.00	2.24 ± 0.01(1)
idemia-001	2024-08-26	2.846(3)	2.89(2)	2.539(2)	0.00	0.41(2)	0.00	3.12 ± 0.02(5)

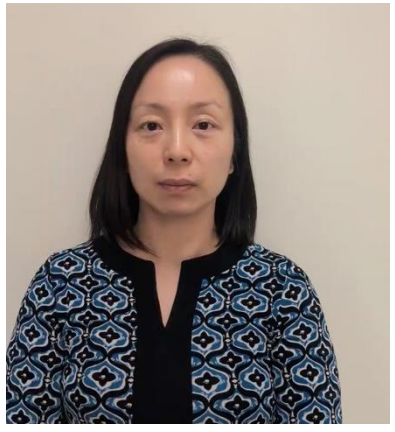
Nuisance Factors Effects

New Dataset	Num. of Videos	Num. of Frames	Num. People
Frontal Videos	482	434,230	482
Yaw Videos	29	26,056	29



Frontal Videos

Yaw Videos



Algorithm	Submission Date	Variability in video and attributable to non-frontal pose			
		FTE Frontal	MAE Noise Frontal	FTE Yaw	sin(Yaw) Dependence
roc-000	2023-11-03				
viante-000	2024-11-25	0.00	0.37(1)	0.00	2.24 ± 0.01(1)
idemia-001	2024-08-26	0.00	0.41(2)	0.00	3.12 ± 0.02(5)
inbut-000	2024-09-11	0.00	0.44(3)	0.00	2.80 ± 0.02(2)
innovatrics-000	2025-02-05	0.00	0.45(4)	0.00	2.84 ± 0.01(3)
nominder-001	2024-11-25	0.00	0.52(5)	0.00	3.62 ± 0.03(6)
roc-001	2024-10-29	0.00	0.53(6)	0.00	3.00 ± 0.02(4)
yoti-002	2024-09-17	0.00	0.56(7)	0.00	5.10 ± 0.05(10)
incode-000	2023-09-05	0.00	0.63(8)	0.00	4.83 ± 0.03(8)
nominder-000	2024-07-22	0.00	0.64(9)	0.00	3.91 ± 0.02(7)
jumio-000	2024-08-30	0.00	0.72(10)	0.16	28.04 ± 0.11(19)
neurotechnology-000	2023-10-17	0.00	0.75(11)	0.00	8.29 ± 0.04(17)
neurotechnology-001	2024-07-19	0.00	0.79(12)	0.00	7.80 ± 0.04(16)
yoti-001	2023-09-19	0.00	0.82(13)	0.00	5.65 ± 0.03(13)
ondato-000	2024-11-08	0.00	0.91(14)	0.00	5.20 ± 0.03(11)
dermalog-001	2023-11-02	0.00	0.97(15)	0.00	5.41 ± 0.03(12)
dermalog-002	2025-01-13	0.00	1.09(16)	0.00	13.43 ± 0.07(18)
unissey-001	2023-09-19	0.00	1.32(17)	0.00	4.84 ± 0.03(9)
cvut-002	2024-11-04	0.18	2.05(18)	0.01	6.01 ± 0.10(14)
cvut-001	2024-06-14	0.18	4.14(19)	0.01	7.18 ± 0.12(15)

FATE AEV Leaderboard/Webpage

NIST NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY U.S. DEPARTMENT OF COMMERCE

Challenge-25 and Child Online Safety (13-16)

The table shows accuracy measures for 2 common applications, Challenge-25 age-restriction and children age 13-16 online chatroom access. False positive rates quantify how often people outside the allowed age range are estimated to be within required category. True positive rates quantify how often children within age 13-16 are estimated to be within that age range.

Challenge 25:

- [Dataset: Application (good quality), Border (medium quality)]
- Application - *False Positive Rate (FPR)* values are averages of 12 FPR estimates from two sexes and six regions-of-birth for subjects aged 17.
- Border - *False Positive Rate (FPR)* values are averages of 12 FPR estimates from two sexes and six regions-of-birth for subjects aged 17.

Child Online Safety (13-16):

- [Dataset: Visa (good quality)]
- Age 13-16 - *Mean Absolute Errors (MAE)* are averages of four MAE values for age group 13, 14, 15, and 16 (below 17).
- Age 13-16 - *True Positive Rates (TPR)* are proportions of subjects aged 13 to 16 whose age is correctly estimated as 13 to 16 (below 17).
- Age 13-16 - *False Negative Rates (FNR↓)* are proportions of subjects aged 13 to 16 whose age is incorrectly estimated as 8 to 12.
- Age 13-16 - *False Negative Rates (FNR↑)* are proportions of subjects aged 13 to 16 whose age is incorrectly estimated as 17 to 22.
- Age 8-12 - *False Positive Rates (FPR)* are proportions of subjects aged 8 to 12 but whose age is incorrectly estimated as 13 to 16 (below 17).
- Age 17-22 - *False Positive Rates (FPR)* are proportions of subjects aged 17 to 22 but whose age is incorrectly estimated as 13 to 16 (below 17).

FATE AEV Leaderboard/Webpage

NIST NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

Resource Cost | MAE By Dataset and Variability | Challenge-25 & Child Online Safety (13-16) | MAE Across Geography

Challenge-25 and Child Online Safety (13-16)

The table shows accuracy measures for 2 common applications, Challenge-25 age-restriction and children age 13-16 online chatroom access. False positive rates quantify how often people outside the allowed age range are estimated to be within required category. True positive rates quantify how often children within age 13-16 are estimated to be within that age range.

Show 30 entries

Search: ex. 2023|insight

Algorithm	Challenge 25		Child Online Safety (13-16)					
	Application	Border	Age 13-16				Age 8-12	Age 17-22
	FPR	FPR	MAE	TPR	FNR↓	FNR↑	FPR	FPR
yoti-002	0.072±0.018$^{(7)}$	0.190±0.029$^{(9)}$	2.0629$^{(1)}$	0.5690$^{(23)}$	0.1008$^{(21)}$	0.3303$^{(1)}$	0.2154$^{(3)}$	0.0685$^{(23)}$
idemia-001	0.024±0.011$^{(1)}$	0.044±0.015$^{(1)}$	2.4601$^{(2)}$	0.4954$^{(22)}$	0.1475$^{(23)}$	0.3572$^{(2)}$	0.1213$^{(1)}$	0.0609$^{(22)}$
incode-000	0.047±0.017$^{(3)}$	0.254±0.033$^{(16)}$	2.7267$^{(3)}$	0.3678$^{(16)}$	0.1027$^{(22)}$	0.5295$^{(4)}$	0.1409$^{(2)}$	0.0309$^{(18)}$
yoti-001	0.110±0.024$^{(12)}$	0.223±0.032$^{(14)}$	2.8515$^{(4)}$	0.4774$^{(21)}$	0.0952$^{(20)}$	0.4273$^{(3)}$	0.2342$^{(4)}$	0.0434$^{(21)}$

Challenge 25:

- [Dataset: App
- Application - cts aged 17.
- Border - *False* ged 17.

Child Online Safety

- [Dataset: Vis
- Age 13-16 - *F*
- Age 13-16 - low 17).
- Age 13-16 - *False Negative Rates (FNR↓)* are proportions of subjects aged 13 to 16 whose age is incorrectly estimated as 8 to 12.
- Age 13-16 - *False Negative Rates (FNR↑)* are proportions of subjects aged 13 to 16 whose age is incorrectly estimated as 17 to 22.
- Age 8-12 - *False Positive Rates (FPR)* are proportions of subjects aged 8 to 12 but whose age is incorrectly estimated as 13 to 16 (below 17).
- Age 17-22 - *False Positive Rates (FPR)* are proportions of subjects aged 17 to 22 but whose age is incorrectly estimated as 13 to 16 (below 17).

FATE AEV Leaderboard/Webpage

NIST NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY U.S. DEPARTMENT OF COMMERCE

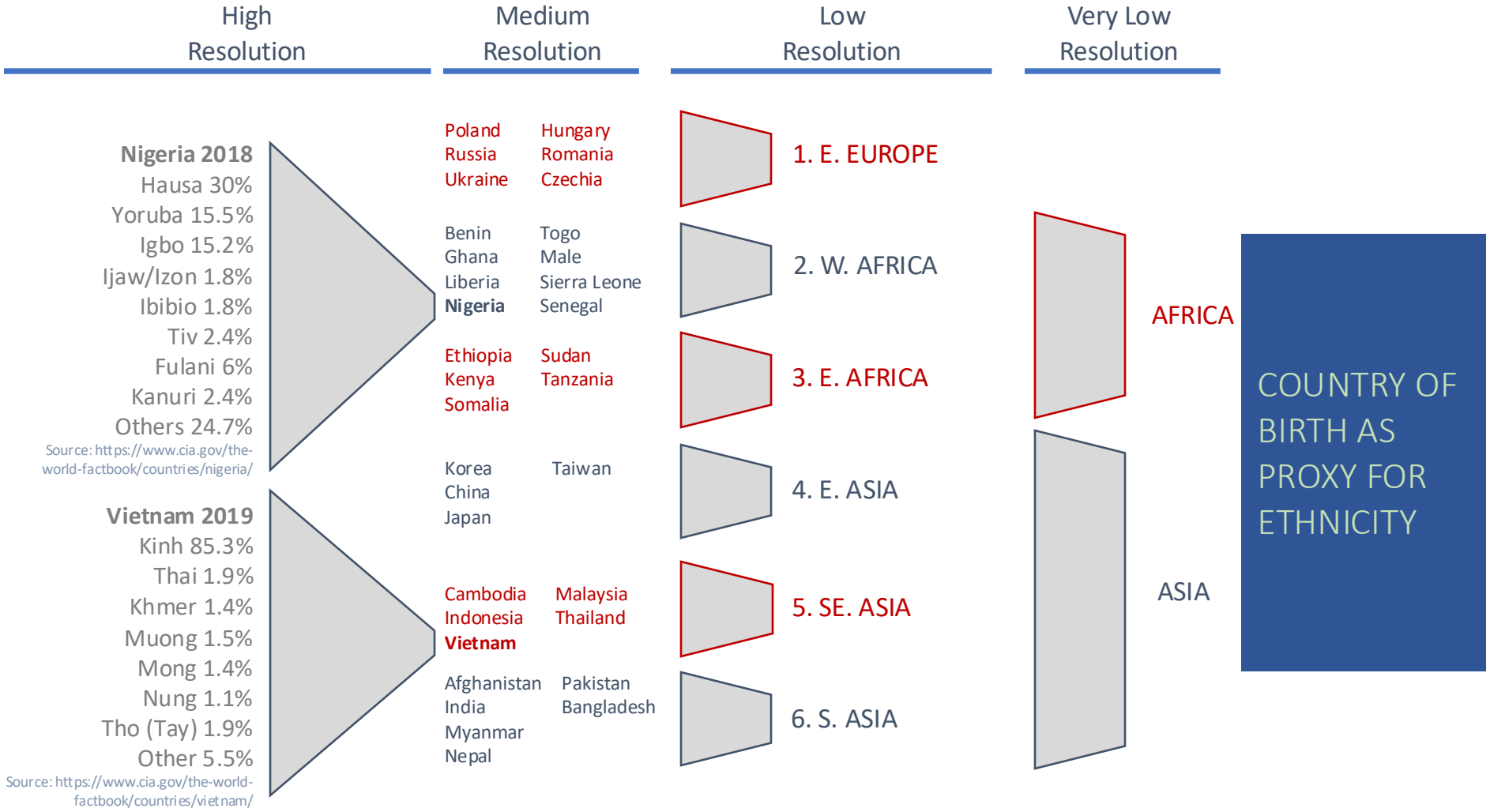
Resource Cost | MAE By Dataset and Variability | Challenge-25 & Child Online Safety (13-16) | MAE Across Geography

The table shows mean absolute error (MAE) for men and women born in six regions of the world. See the AEV report for discussion of how these regions were selected and grouped. Lower values of MAE indicate better accuracy. Low variation across columns indicates level accuracy across countries of birth. MAE is estimated over high quality visa-like immigration office application photos. The uncertainty estimates span 95% of bootstrap estimates of the mean. The table can be sorted on any column by clicking the small arrows in its header.

Show 30 entries

Search: ex. 2023|insight

Algorithm	Region of Birth and Sex											
	E Africa		E Asia		E Europe		S Asia		S E Asia		W Africa	
	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
idemia-001	3.3 ± 0.2$^{(1)}$	3.2 ± 0.2$^{(3)}$	2.45 ± 0.07$^{(2)}$	2.19 ± 0.08$^{(2)}$	2.9 ± 0.1$^{(10)}$	2.1 ± 0.1$^{(7)}$	2.73 ± 0.09$^{(1)}$	2.60 ± 0.09$^{(6)}$	2.7 ± 0.1$^{(2)}$	2.3 ± 0.1$^{(3)}$	3.3 ± 0.2$^{(1)}$	3.0 ± 0.2$^{(1)}$
nominder-001	3.4 ± 0.2$^{(2)}$	3.0 ± 0.2$^{(1)}$	2.80 ± 0.07$^{(5)}$	2.52 ± 0.08$^{(6)}$	2.6 ± 0.1$^{(4)}$	2.1 ± 0.1$^{(6)}$	2.89 ± 0.08$^{(3)}$	2.49 ± 0.08$^{(2)}$	2.8 ± 0.1$^{(3)}$	2.4 ± 0.1$^{(4)}$	4.2 ± 0.2$^{(9)}$	3.1 ± 0.2$^{(3)}$
neurotechnology-001	3.8 ± 0.2$^{(3)}$	4.0 ± 0.2$^{(15)}$	4.1 ± 0.1$^{(21)}$	3.85 ± 0.09$^{(20)}$	3.5 ± 0.2$^{(16)}$	3.0 ± 0.2$^{(19)}$	3.8 ± 0.1$^{(17)}$	3.7 ± 0.1$^{(19)}$	3.9 ± 0.2$^{(16)}$	3.7 ± 0.2$^{(18)}$	4.0 ± 0.2$^{(5)}$	3.8 ± 0.2$^{(15)}$
cvut-002	3.8 ± 0.2$^{(4)}$	3.3 ± 0.2$^{(4)}$	3.06 ± 0.08$^{(7)}$	2.60 ± 0.08$^{(8)}$	2.7 ± 0.1$^{(5)}$	2.3 ± 0.1$^{(11)}$	2.97 ± 0.10$^{(4)}$	2.68 ± 0.10$^{(8)}$	3.1 ± 0.1$^{(7)}$	2.5 ± 0.1$^{(5)}$	3.9 ± 0.2$^{(4)}$	3.3 ± 0.2$^{(4)}$
nominder-000	4.0 ± 0.2$^{(5)}$	3.5 ± 0.2$^{(9)}$	3.09 ± 0.09$^{(8)}$	2.92 ± 0.10$^{(14)}$	2.7 ± 0.1$^{(6)}$	2.5 ± 0.2$^{(13)}$	3.1 ± 0.1$^{(6)}$	3.2 ± 0.1$^{(14)}$	3.0 ± 0.1$^{(6)}$	2.8 ± 0.2$^{(10)}$	4.8 ± 0.3$^{(13)}$	3.6 ± 0.2$^{(14)}$


High Resolution
Medium Resolution
Low Resolution
Very Low Resolution
Nigeria 2018
Hausa 30%
Yoruba 15.5%
Igbo 15.2%
Ijaw/Izon 1.8%
Ibibio 1.8%
Tiv 2.4%
Fulani 6%
Kanuri 2.4%
Others 24.7%
Source: https://www.cia.gov/the-world-factbook/countries/nigeria/
Vietnam 2019
Kinh 85.3%
Thai 1.9%
Khmer 1.4%
Muong 1.5%
Mong 1.4%
Nung 1.1%
Tho (Tay) 1.9%
Other 5.5%
Source: https://www.cia.gov/the-world-factbook/countries/vietnam/
Poland Hungary
Russia Romania
Ukraine Czechia
1. E. EUROPE
Benin Togo
Ghana Male
Liberia Sierra Leone
Nigeria Senegal
2. W. AFRICA
Ethiopia Sudan
Kenya Tanzania
Somalia
3. E. AFRICA
Korea Taiwan
China
Japan
4. E. ASIA
Cambodia Malaysia
Indonesia Thailand
Vietnam
5. SE. ASIA
Afghanistan Pakistan
India Bangladesh
Myanmar
Nepal
6. S. ASIA
AFRICA
ASIA
COUNTRY OF BIRTH AS PROXY FOR ETHNICITY

MAE Across Countries of Birth

NIST NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY U.S. DEPARTMENT OF COMMERCE

Algorithm	Region of Birth and Sex											
	E Africa		E Asia		E Europe		S Asia		S E Asia		W Africa	
	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
innovatrics-000	4.1 ± 0.2 (7)	3.5 ± 0.2 (8)	2.43 ± 0.07 (1)	2.05 ± 0.07 (1)	2.4 ± 0.1 (1)	2.0 ± 0.1 (3)	2.84 ± 0.09 (2)	2.54 ± 0.08 (4)	2.4 ± 0.1 (1)	2.0 ± 0.1 (1)	4.3 ± 0.2 (10)	3.4 ± 0.2 (7)
yoti-002	4.0 ± 0.2 (6)	3.2 ± 0.2 (2)	3.44 ± 0.10 (10)	2.43 ± 0.08 (4)	2.5 ± 0.1 (3)	1.8 ± 0.1 (1)	3.5 ± 0.1 (11)	2.52 ± 0.09 (3)	3.9 ± 0.2 (15)	2.5 ± 0.1 (6)	4.2 ± 0.3 (7)	3.3 ± 0.2 (5)
nominder-001	3.4 ± 0.2 (2)	3.0 ± 0.2 (1)	2.80 ± 0.07 (5)	2.52 ± 0.08 (5)	2.6 ± 0.1 (4)	2.1 ± 0.1 (5)	2.89 ± 0.08 (3)	2.49 ± 0.08 (2)	2.8 ± 0.1 (3)	2.4 ± 0.1 (4)	4.2 ± 0.2 (9)	3.1 ± 0.2 (3)
idemia-001	3.3 ± 0.2 (1)	3.2 ± 0.2 (3)	2.45 ± 0.07 (2)	2.19 ± 0.08 (2)	2.9 ± 0.1 (9)	2.1 ± 0.1 (6)	2.73 ± 0.09 (1)	2.60 ± 0.09 (6)	2.7 ± 0.1 (2)	2.3 ± 0.1 (3)	3.3 ± 0.2 (1)	3.0 ± 0.2 (1)
incode-000	4.2 ± 0.2 (8)	3.3 ± 0.2 (6)	2.80 ± 0.09 (4)	2.24 ± 0.07 (3)	3.0 ± 0.2 (11)	2.1 ± 0.1 (7)	3.0 ± 0.1 (5)	2.43 ± 0.09 (1)	2.9 ± 0.1 (4)	2.2 ± 0.1 (2)	3.8 ± 0.2 (2)	3.0 ± 0.2 (2)

A: Application Photos

B: NOTABLE:

- BEST MAE IN E. EUROPEAN MEN, 1.8 YRS
- WORST MAE IS E. AFRICA WOMEN, 3.3 YRS
- MAE IN WOMEN > MAE IN MEN
- INNOVATRICS 2025 OVERALL LOWEST MAE
- IDEMIA 2024-08 LEADER, NOW #3

MAE values are averages of MAE measured over persons **14 … 30** estimated in good quality a visa-like application photos

Function #1 and #2

Estimate Age Image

Function #1
estimateAge()

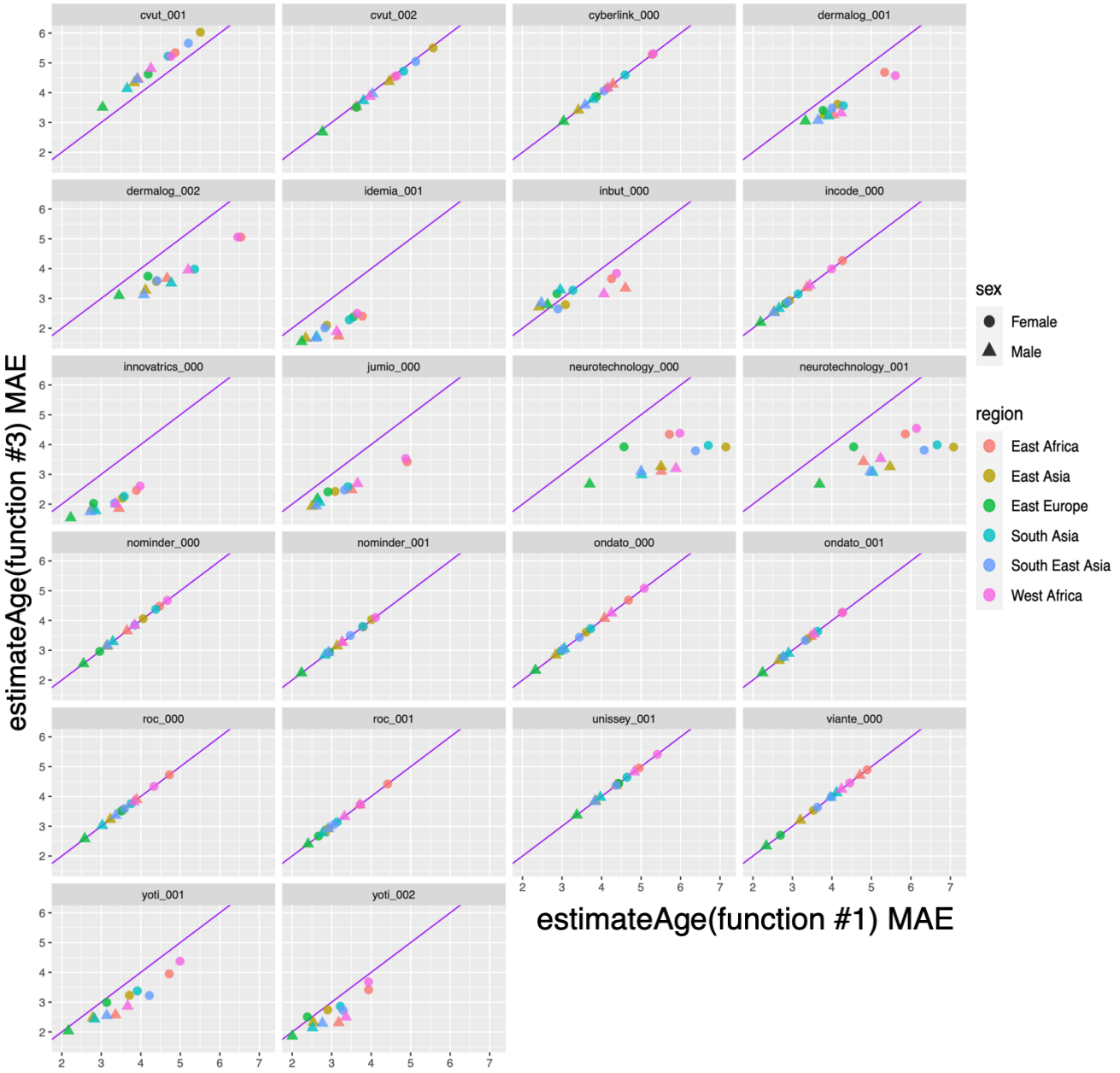
Age = 42.43

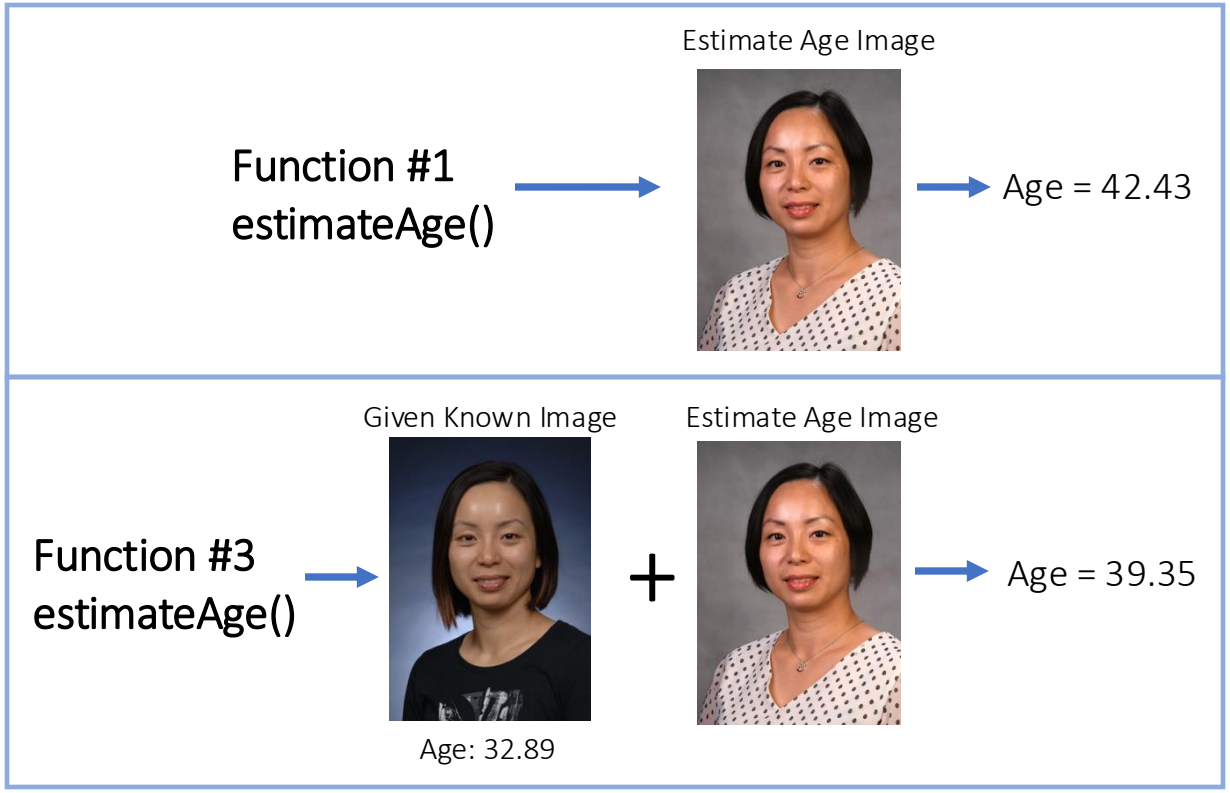
Verify Age > 25

Function #2
verifyAge()

> 25 = True

What is Function #3?





	MAE	
	Function #1	Function #3
Idemia-001	3.02	1.99
Dermalog-002	4.80	3.80
Yoti-002	3.00	2.61
Jumio-000	3.34	2.51
Neurotechnology-001	5.49	3.63

API Change :

#1

#1: ESTIMATE_AGE

FROM SINGLE IMAGE

⟶ AGE = 36.4

CHANGES:
1. None
2. MUST implement function

#2

#2: VERIFY_AGE > 25

FROM SINGLE IMAGE

⟶ SCORE

CHANGES:
1. Optional function
2. Specify limited set of key ages instead of any integer e.g. {13,16,18,21,25}
3. Return a score instead of a true/false
4. Only implement IF adopts a more sophisticated method than simply comparing AE to the specified threshold.

#3

#3: ESTIMATE_AGE

GIVEN

+

⟶ AGE = 36.4

CHANGES:
1. Optional function
2. Only implement IF a) Effectively leverage the known-age image for gain, b) Articulate a real-world case to us

BEYOND SCOPE: AGE ASSURANCE ⟶ TRUSTED AV including LIVENESS + PAD + FALLBACKS

Summary

Performance

- Mean absolute error is a generic performance indicator
  - Error distributions can be asymmetric
- Application-specific performance metrics:
  - Challenge-T FPR
  - Chat room FPR rates.
- Accuracy varies across algorithm
  - Most algorithms superior to the 2014 best
- Operationally Presentation Attack Detection is usually required
  - Coupled to the AE system

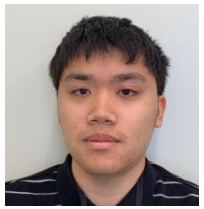
Next steps

- Accuracy today != accuracy tomorrow.
- Work needed on
  - Nuisance variables
  - Reduced quality photos
- FATE AEV is an ongoing resource available to developers + purchasers + policy makers
- FATE AEV status
  - We're open
  - New API
  - New global dataset 0 – 99+ expected 2025 Q3.
  - New high-resolution dataset 14 – 99 expected 2025 Q2

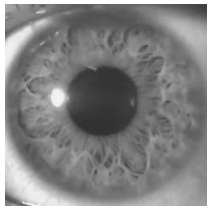
Thank you!


Patrick Grother (?)


Kayee Hanaoka (morphed)


Austin Hom


(not) Mei Ngan


Jim Matey


Joyce Yang

Thank you!!

Kayee.Hanaoka@nist.gov

frvt@nist.gov