



Yoti Facial Age Estimation

White Paper | Full version

March 2023



Executive summary

Yoti's facial age estimation can determine a person's age by using an image of their face, with no need for a physical check of documents or human intervention. It is very accurate across all skin types and gender. It is so accurate that any business globally could accept estimated age as a reliable proof of age.

The technology is built in accordance with the GDPR principle of 'privacy by design'. An individual cannot be identified by the model and it encourages data minimisation by only needing a facial image - Yoti immediately deletes all images of users. The model cannot infer anything else about a person nor can it uniquely identify a person.

Our True Positive Rate¹ (TPR) for 13-17 year olds being correctly estimated as under 25 is 99.93% and there is no discernible bias across gender or skin tone. The TPRs for female and male 13-17 year olds are 99.90% and 99.94% respectively. The TPRs for skin tone 1, 2 and 3 are 99.93%, 99.89% and 99.92% respectively. This gives regulators globally a very high level of confidence that children will not be able to access adult content.

Our TPR for 6-11 year olds being correctly estimated as under 13 is 98.35%. The TPRs for female and male 6-11 year olds are 98.00% and 98.71% respectively. The TPRs for skin tone 1, 2 and 3 are 97.88%, 99.24% and 98.18% respectively so there is no material bias in this age group either.

Yoti's facial age estimation is performed by a 'neural network', trained to be able to estimate human age by analysing a person's face. Our technology is accurate for 6 to 12 year olds, with a mean absolute error (MAE) of 1.3 years, and of 1.4 years for 13 to 17 year olds. These are the two age ranges regulators focus upon to ensure that under 13s and 18s do not have access to age restricted goods and services.

At Yoti, we take our ethical responsibilities as a company developing new technology very seriously. The millions of faces used to train the algorithm (face image, month and year of birth only) are obtained by Yoti during the onboarding process for our apps and in accordance with GDPR or using consented data collection exercises. See page 24, Appendix.

This iteration of the algorithm continues to show improvements in accuracy. We have invested significant time and resources to reduce bias and we are now able to show that there is no material bias for age ranges between 6 and 17. We have also made significant progress on reducing bias for older females and those with a darker skin tone to ensure facial age estimation is as equitable as possible. Additionally, to meet regulatory best practice, we are beginning to retire old training and testing data.

1. True Positive Rate - the probability that an actual positive will test positive, such as an 18 year old is correctly estimated to be under 25.

Expanding the data set & improving accuracy

Our first white paper, published in December 2018, contained accuracy across the age ranges of 13-60. Since September 2021, we have published our 6-12 age range data, and from May 2022 included data for age range 6-70, published by year, gender and skin tone.

We are pleased to report the algorithm continues to show improvements in accuracy. We continue to see some small deviations in this trend, best explained by demographic changes in the underlying training and testing data (see page 24 for a detailed discussion).

Key takeaways

- TPR for 13-17 year olds correctly estimated as under 25 is 99.93%.
- TPR for 6-11 year olds correctly estimated as under 13 is 98.35%.
- Mean Absolute Errors (in years) are 2.9 for 6-70, 1.4 for 13-17 & 1.3 for 6-12.
- Users are not individually identifiable
- Helps organisations to meet Children's Codes or Age Appropriate Design Codes
- Does not result in the processing of special category data
- Gender and skin tone bias minimised.
- Training data collected in accordance with GDPR.
- Independently tested and certified.
- A secure, privacy respecting solution that protects individuals.
- Yoti liveness and age estimation is very hard to 'fool'.
- Over 593 million checks performed worldwide.
- Solution is fast and scales to over 25 million checks per day, or 300 checks per second.
- We are continuously improving both accuracy and usability.


Skin tones

For skin tone, our research team tagged the images using a scheme based on the widely used Fitzpatrick dermatological scale (see p27 for further information). Fitzpatrick uses six bands, from Type I (lightest) to Type VI (darkest). For the present, we have presented our data in three bands (based on Fitzpatrick Types I & II, Types III & IV, and Types V & VI).

Skin tone scale



Mean Absolute Error by age band

 Yoti facial age estimation accuracy					Mean estimation error in years split by gender, skin tone and age band				
Gender	Female				Male				All
Skintone	Tone 1	Tone 2	Tone 3	All	Tone 1	Tone 2	Tone 3	All	
6-12	1.3	1.4	1.6	1.4	1.2	1.4	1.3	1.3	1.3
13-17	1.3	1.5	1.6	1.5	1.0	1.3	1.5	1.3	1.4
18-24	2.4	2.5	2.6	2.5	2.0	2.0	2.2	2.0	2.3
25-70	2.9	3.3	4.6	3.6	2.6	3.3	3.6	3.2	3.4
6-70	2.5	2.9	3.8	3.1	2.3	2.8	3.0	2.7	2.9

With age estimation, once you know you're dealing with a child, you can...

-  Turn off excessive notifications.
-  Set geolocation to off but give the child the ability to turn it on if needed.
-  Provide age-appropriate content.
-  Be certain the online community is within the same age threshold.
-  Minimise the data you collect - don't store it.
-  Shield their data. It shouldn't be used for things not in their interest.
-  Use child-friendly language to explain platforms.
-  Always be sure to treat a child like a child.

About 'Mean Absolute Error'

Yoti facial age estimation can make both positive and negative errors when estimating age (that is, it can estimate too high, or it can estimate too low). By taking 'absolute' values of each error, we mean ignoring whether the error is positive or negative, simply taking the numerical size of the error. We then take the average (or 'arithmetic mean') of all those absolute error values, producing an overall 'MAE'. A table of MAE by year can be found in the appendix on pages 28-30.

Mean Absolute Error by year

Age	Gender								All
	Female				Male				
	Skin Tone (Fitzpatrick Scale)								
	Type I & II	Type III & IV	Type V & VI	All	Type I & II	Type III & IV	Type V & VI	All	
	MAE	MAE	MAE	Average MAE	MAE	MAE	MAE	Average MAE	
6	1.3	2.0	1.6	1.6	1.3	2.2	1.7	1.8	1.7
7	1.4	1.4	1.3	1.3	1.0	1.6	2.2	1.6	1.5
8	1.6	1.3	1.4	1.4	1.5	1.7	1.4	1.5	1.5
9	1.3	1.2	1.6	1.4	1.5	1.2	1.0	1.2	1.3
10	1.2	1.2	1.5	1.3	1.0	1.1	1.0	1.0	1.2
11	0.9	1.2	1.9	1.3	0.9	0.9	0.9	0.9	1.1
12	1.2	1.5	1.6	1.4	0.9	1.1	0.9	1.0	1.2
13	1.9	2.5	2.3	2.2	1.4	1.5	1.9	1.6	1.9
14	1.4	1.7	2.1	1.7	1.1	1.5	2.0	1.5	1.6
15	1.2	1.3	1.5	1.3	1.0	1.2	1.4	1.2	1.3
16	0.9	1.0	1.2	1.0	0.9	1.1	1.0	1.0	1.0
17	1.0	0.9	1.0	1.0	0.9	1.1	1.0	1.0	1.0
18	1.3	1.2	0.9	1.1	1.2	1.4	1.2	1.3	1.2
19	1.8	1.9	1.6	1.8	1.6	1.7	1.7	1.7	1.7
20	2.3	2.3	2.2	2.3	2.1	2.0	2.2	2.1	2.2
21	2.8	2.8	2.4	2.7	2.1	2.1	2.2	2.1	2.4
22	2.9	2.9	2.9	2.9	2.3	2.1	2.6	2.3	2.6
23	3.0	3.2	3.8	3.3	2.5	2.2	2.4	2.3	2.8
24	2.9	2.9	4.7	3.5	2.5	2.3	3.0	2.6	3.1
25	2.8	3.0	4.2	3.3	2.0	2.1	2.7	2.3	2.8
26	2.3	3.0	4.3	3.2	1.9	2.1	2.5	2.2	2.7
27	2.5	2.8	4.6	3.3	1.9	1.9	2.5	2.1	2.7
28	2.3	3.0	4.4	3.2	2.0	2.2	2.8	2.3	2.8
29	2.4	2.8	4.6	3.3	2.0	2.3	2.8	2.4	2.8
30	2.5	2.9	4.1	3.2	2.0	2.3	2.7	2.3	2.8

Mean Absolute Error by year

Age	Gender								All
	Female				Male				
	Skin Tone (Fitzpatrick Scale)								
	Type I & II	Type III & IV	Type V & VI	All	Type I & II	Type III & IV	Type V & VI	All	
	MAE	MAE	MAE	Average MAE	MAE	MAE	MAE	Average MAE	
31	2.9	3.1	4.6	3.5	2.2	2.4	2.4	2.3	2.9
32	2.8	3.1	5.3	3.8	2.3	2.8	2.9	2.7	3.2
33	3.2	3.7	4.2	3.7	2.6	3.1	2.9	2.9	3.3
34	2.9	3.4	4.7	3.7	2.4	3.0	3.0	2.8	3.3
35	3.2	3.3	5.7	4.1	2.6	2.9	2.6	2.7	3.4
36	2.5	2.6	4.7	3.3	2.6	2.9	3.1	2.9	3.1
37	3.1	3.9	3.7	3.5	2.6	3.2	3.3	3.0	3.3
38	2.4	2.8	4.0	3.1	2.2	3.1	3.0	2.8	2.9
39	2.8	3.9	3.6	3.4	2.6	2.7	3.6	2.9	3.2
40	2.6	2.8	3.8	3.0	2.5	2.5	2.5	2.5	2.8
41	2.6	3.2	3.5	3.1	2.3	2.4	2.7	2.5	2.8
42	2.6	3.0	3.1	2.9	2.4	2.9	3.1	2.8	2.9
43	3.1	3.3	4.3	3.5	2.5	2.6	3.1	2.7	3.1
44	2.5	3.4	3.5	3.2	2.4	2.9	3.1	2.8	3.0
45	2.6	3.7	3.6	3.3	2.3	3.4	2.7	2.8	3.0
46	2.8	3.2	3.2	3.1	3.0	4.2	3.8	3.7	3.4
47	3.1	3.2	3.2	3.2	2.7	3.5	3.1	3.1	3.1
48	3.0	4.0	2.8	3.3	2.7	2.9	3.6	3.1	3.2
49	3.1	2.9	3.7	3.2	3.1	3.2	3.7	3.3	3.3
50	3.0	2.8	5.2	3.7	2.9	3.4	3.5	3.3	3.5
51	3.3	3.2	6.7	4.4	3.1	3.1	3.9	3.4	3.9
52	3.1	3.0	4.9	3.7	3.2	3.1	5.1	3.8	3.7
53	3.2	3.5	4.2	3.7	3.1	3.0	4.0	3.4	3.5
54	2.9	3.5	9.7	5.4	3.2	4.1	4.6	4.0	4.7
55	3.3	3.5	4.7	3.8	2.5	4.2	4.5	3.7	3.8

Mean Absolute Error by year

Age	Gender								All
	Female				Male				
	Skin Tone (Fitzpatrick Scale)								Average MAE
	Type I & II	Type III & IV	Type V & VI	All	Type I & II	Type III & IV	Type V & VI	All	
	MAE	MAE	MAE	Average MAE	MAE	MAE	MAE	Average MAE	
56	2.6	4.0	4.7	3.8	2.8	3.5	3.9	3.4	3.6
57	3.1	3.4	5.4	4.0	2.8	3.2	4.2	3.4	3.7
58	2.9	3.2	6.4	4.2	2.7	3.8	3.9	3.5	3.8
59	2.7	3.6	4.5	3.6	2.6	3.4	4.1	3.4	3.5
60	2.5	3.3	4.7	3.5	2.7	3.5	4.7	3.6	3.6
61	2.8	3.7	5.1	3.8	2.6	3.5	4.2	3.4	3.6
62	2.1	4.5	6.8	4.5	2.6	3.1	3.7	3.2	3.8
63	2.3	2.9	4.7	3.3	2.6	3.1	5.1	3.6	3.4
64	2.5	3.2	4.0	3.2	2.8	3.2	3.9	3.3	3.3
65	2.2	2.3	2.5	2.3	2.4	3.7	3.9	3.4	2.9
66	3.2	3.0	4.4	3.5	3.1	4.8	4.2	4.0	3.8
67	3.5	2.8	8.6	4.9	3.6	3.8	3.8	3.7	4.3
68	3.4	3.1	2.5	3.0	3.5	6.1	4.9	4.8	3.9
69	3.8	3.9		3.9	3.7	4.5	2.9	3.7	3.8
70	4.0	5.4	5.3	4.9	3.3	6.5	10.2	6.6	5.8
Avg	2.5	2.9	3.8	3.1	2.3	2.8	3.0	2.7	2.9

Contents

What is facial age estimation and what can it do?	6	Legal compliance	20
Data privacy and network security	7	Fair standardised measurement	21
How does it actually work?	7	Yoti's commitment to ethical use of AI technologies	22
Tackling the challenge of age determination	8		
Human ability to determine age	9	Appendix	23
More on how it works	10	Data used to build the model ('training data')	24
How accurate is facial age estimation?	11	Data used for testing	26
About 'Mean Absolute Error'	12	Accuracy across the entire data set	26
Safety barriers	13	Accuracy by age, gender and skin tone	27
Public acceptance of AI technologies	14	Mean Absolute Error by year	28
Learnings from practical use	15	Standard deviation of absolute error	31
UK government live trials	16	Absolute versus percentage errors	32
Supporting Children's Codes	17	Improvement in accuracy as the training data set changes	33
Ensuring accurate and genuine assurance	18	False positives	35
Product development	19	False positive improvement over time	38
		Trade-off between false negatives and false positives	39

What is facial age estimation and what can it do?

Yoti facial age estimation is a secure, effective age-checking service that can estimate a person's age by looking at their face. We consider it to have very wide application in the provision of any age-restricted goods and services, both online and in-person. It is also a means to combat social exclusion for the significant numbers of individuals around the world who do not possess a state-issued photo ID document.

Our facial age estimation is designed with user privacy and data minimisation in mind. It does not require users to register with Yoti, or provide any documents to prove their identity. It is unable to personally identify an individual, it simply estimates a person's age from analysing their face.

The images are not stored, not re-shared, not re-used and not sold on. Images are immediately, permanently deleted according to GDPR best practice, and we do not use them for our own learning or training purposes. This is externally reviewed as part of our SOC2 / PAS1296 assessment under control PAS-2.

In a retail setting, facial age estimation can be used at a point-of-sale terminal with a dedicated camera, letting a consumer choose to prove age at a self-checkout without the need for staff assistance. This is not only quicker and less of a nuisance for shoppers, but can greatly reduce friction between hundreds of millions of shoppers and hundreds of thousands of retail staff each day.

For general online use, it can be embedded into web pages or incorporated into apps, and receive an image of the user's face from a webcam connected to their computer or the camera in their mobile device. This is ideal for controlling access to age-restricted gaming, social media, e-commerce, online dating, gambling and also adult content (pornography).

Facial age estimation can play an important role in safeguarding children online. As well as preventing minors from accessing adult content, it can prevent predatory adults from entering social media spaces designed for children and teenagers. This is illustrated by Yoti's partnership with the Yubo social networking platform. Yubo uses facial age estimation within its app to help identify user profiles where there is suspicion or doubt about the user's age, and flags these cases to its moderation team.

Deployment on premise and on device

Facial age estimation can also be deployed on premise by law enforcement to assess ages of victims and perpetrators in child abuse images. We have also developed a more efficient and lightweight age estimation model that can run on platforms with limited or low computational resources and mobile devices. This model provides much faster results, has no reliance on internet connectivity, and is just 8.4% less accurate than our production model.

A further potential use is at the entrances to age-restricted premises such as bars, nightclubs and casinos. In this kind of application, facial age estimation offers clear advantages – it does not get fatigued on a long shift,¹ and it cannot show favour to personal friends, or bias against individual customers. It is very hard for under 18s to 'fool'. It also reduces the burden on staff to try and estimate customer ages and it reduces abuse to staff.

1. Studies have shown that the objectivity of human judgement of this kind can be significantly affected by hunger and fatigue – see for instance Danziger, Levav, Avnaim-Passo (2011) *Extraneous factors in judicial decisions*, PNAS April 26, 2011 108 (17) 6889-6892; <https://doi.org/10.1073/pnas.1018033108>

Data privacy and network security

Yoti's facial age estimation has been designed with data privacy and security as primary considerations.

The user does not have to register to use the service, and does not have to provide any information about themselves - they simply present their face in front of the camera. Their image is not stored locally, for example on a point-of-sale terminal, but securely transmitted to the Yoti backend server (currently hosted in the United Kingdom) and secured by TLS 1.2 encryption. After the age estimate is performed, the captured facial image is deleted from these servers. The photograph is not viewed by any Yoti staff.

Although Yoti's facial age estimation works by processing a facial photograph, under the GDPR definition of biometric data, it is not a 'biometric' method of age checking. Our means of processing does not allow the "unique identification or authentication of a natural person", instead it merely returns an age estimate and deletes the captured photograph.

In GDPR terms, Yoti is a data processor for the facial age estimation service. The relying party (Yoti's customer) is the data controller. As such, they will decide the lawful basis for their use of facial age estimation (if required under EU or UK privacy law). In many jurisdictions, the individual will need to provide consent. The facial age estimation user interface is configurable so that relying parties can build in this request for consent. This feature is enabled by default for our US customers.

How does it actually work?

Facial age estimation is based on a computing technique known as a 'neural network', which we have trained to be able to estimate human age using a process of 'machine learning'. This is a form of artificial intelligence (AI), and is increasingly used in a wide variety of applications, from driverless cars to medical diagnosis, from tailoring online advertising to detecting credit card fraud. We discuss machine learning in more detail on page 10, but first some context on the problem we are using it to solve.

3 minute video explanation of Facial Age Estimation, delivered by Yoti partner Be in Touch



<https://www.youtube.com/watch?v=6KCUO2vIn3M>

Tackling the challenge of age determination

In a medical sense, 'ageing' refers to the physiological changes we experience as we grow from children to adults. The important point is that the rate at which human bodies 'age' in this way is influenced by numerous external factors other than simply the passage of time. Quality of diet and nutrition, exposure to disease, adverse environmental conditions, use of narcotics, physical labour, stress, lack of sleep - these can all have a contributing effect on how we 'age'.

The important point is that the rate at which human bodies 'age' in this way is influenced by numerous external factors other than simple passage of time. Clearly, there are large variations throughout populations as to how different individuals are exposed to these ageing factors. The more extensively we look through different countries, ethnicities, and socio-economic groups, the wider these variations in exposure to ageing factors become.

It may be surprising to learn that there are currently no entirely reliable medical or forensic methods to determine human age. Two of the more commonly attempted medical techniques focus on trying to ascertain whether the subject is above or below the legal age of maturity. These are X-ray or Magnetic Resonance Imaging of bone structure² in the wrists (the degree to which the cartilage between the carpal bones has ossified) and dental X-rays (examining the maturity of wisdom teeth). However, both are expensive to carry out and have a typical margin of error of at least two or three years, and for individuals with an atypical history to the general population, the error can be significantly worse. Due to this unreliability, their use has proved controversial – for instance, immigration authorities attempt to differentiate between child and adult refugees without documentation.

Other medical techniques examine 'biomarkers' taken from blood or tissue samples. Examples include measuring the degree of DNA methylation present, the length of the 'telomere' portion of chromosomes, or the serum levels of the metabolite C-glycosyl tryptophan. Whilst these techniques tend to provide good indicators of ageing processes in an individual, they do not correlate reliably with their chronological age from date of birth.

Ultimately, it could be argued that much of the difficulty in trying to measure 'age' (that is, a person's chronological age from their date of birth) arises because defined in this way it's an arbitrary quantity that does not mean anything definite in physiological terms. Science can accurately measure the extent to which a person's body has aged (that is, to what extent it has developed, grown, matured and decayed), but cannot always reliably determine how many years it took for their body to arrive at that state.

2. Biological evaluation methods to assist in assessing the age of unaccompanied asylum-seeking children;
<https://www.gov.uk/government/publications/methods-to-assess-the-age-of-unaccompanied-asylum-seeking-children/biological-evaluation-methods-to-assist-in-assessing-the-age-of-unaccompanied-asylum-seeking-children-accessible>

Human ability to determine age

Notwithstanding the difficulty in devising an accurate forensic test for age, people still possess a reasonably good ability to guess someone's age simply by looking at them. People have been estimating other people's ages for hundreds of years. Regulators globally often rely on individuals working for businesses to estimate the age of customers to prevent harms to children. How do we manage to estimate age? In terms of facial features, what are the tell-tale signs we look for?

The most obvious visual cues include bone structure (bones grow and develop as we pass from child to adulthood), skin (wrinkles, elasticity) and hair colour (greyness), male baldness or facial hair after puberty. We could add many more cues to this list. However, whatever the detailed nature of the visual cues, the more general point is this: as humans, we simply learn "that's what people of a particular age look like". As we go through life, we encounter other people, we see what they look like and we sometimes learn how old they are, with varying degrees of precision (e.g. "a baby", "14", "mid-40s", "79" and so on). We accumulate this information and experience throughout our lives, and our brains can use it to make quick intuitive judgements. The extent of our previous experiences - familiarity - will be an important factor in how good our guesses are. We will be more accurate at guessing the age of someone from our own familiar peer group than from one we've not encountered.

However, whilst some people are good at estimating age, others struggle, and this variability can frustrate teenagers and young adults who are often age estimated and asked to provide physical proof of age. There are also many other practical issues; the need to prove the age of children, particularly online; or in a law enforcement setting, where the requirement could be to estimate the age of a person in a image; or at scale for global networks.

A study in this area³ reported an MAE in human guesses of 4.7 years across an age range of 0 to 70; across an age range of 16–70, this rose to an MAE of 7.4 years.

It is worth emphasising that, although we might be able to retrospectively rationalise or refine our guess at someone's age, our initial judgement is more or less intuitive. We are not consciously following some step-by-step, rule-based method (for instance "add five years if there are wrinkles", or "add ten years for grey hair"). In effect, we don't 'know how we do it' – generally, our brains process the image and form an instinctive judgement, in line with what we've learnt from past experience, faster than any conscious deliberation or systematic evaluation of facial features. It turns out that this 'black box' approach to describing our cognitive process (that is, simply training our brain with data, without worrying too much about how it works) can actually be employed as a successful technique in machine learning too.

3. H. Han, C. Otto, X. Liu and A. K. Jain (2015) *Demographic Estimation from Face Images: Human vs. Machine Performance*, IEEE Trans. PAMI, Vol. 37, No. 6, 1148–1161 <https://doi.org/10.1109/TPAMI.2014.2362759>. See also Clifford CWG, Watson TL, White D. (2018) *Two sources of bias explain errors in facial age estimation*. R. Soc. open sci. 5:180841. <http://dx.doi.org/10.1098/rsos.180841> and Voelkle, Ebner, Lindenberger & Riediger (2012) *Let Me Guess How Old You Are: Effects of Age, Gender and Facial Expression on Perceptions of Age*. Psychology & Aging, 27 No.2 265–277. <https://doi.org/10.1037/a0025065>

More on how it works

The first challenge for facial age estimation is 'face detection'. It has to examine the image it gets from the camera, and work out which bit of it is an actual human face. Only this portion of the image is then fed into the neural network to get an age estimate. This stage also allows for basic error checking: if the system can't find a face in the image (perhaps because a customer didn't position themselves properly in front of the camera, or some inappropriate object is put there) then the system can return an explanatory error message instead. This is also the stage when Yoti can check to be sure the face is real in front of the camera.



Note that this is not 'facial recognition' (where a computer system is trying to match a particular face against a database, to confirm that person's identity). It is simply detecting whether or not there is anything in the captured image that looks like a human face.

We now come to the interesting bit. The facial image is made up of pixels, which to the computer, is just a set of numbers. These numbers are fed into the artificial neural network - this is a network of mathematical processing nodes, arranged in layers, that is roughly analogous to the connections in the human brain.

Whilst a typical brain has around 100 billion neurons, this network has just hundreds of thousands of nodes. We feed numbers (pixel data) in, and they percolate through the neural net. Each node performs a mathematical function on the pixel data, and passes the result on to nodes in the next layer, until a number finally emerges out the other side. This number is an age estimate.

It's an obvious question to ask 'how is the neural network processing the data? What is it looking for - wrinkles? grey hairs?' and so on. However, this is a human way of thinking about it. To the computer, it's just being fed numbers. It doesn't 'know' what the numbers represent or what they mean and we don't try to tell it that. In the training phase of facial age estimation, we told it the right answers and fed it millions of diverse facial images, for which we knew the subject's age with confidence.

The neural network keeps digesting the pixel data from each image, processing the numbers and trying to get a result which matches the right answer. By repeating and adjusting the process, keeping the variations which bring it closer to the right answer, rejecting the variations which don't, it is 'learning'.

After repeating the process a huge number of times, it arrives at sets of processing formulae which work best. To a human, these formulae would be bafflingly long and complex, and next to meaningless. However, it has effectively created a very complex model of age determination that is far superior to relying on a set of handcrafted instructions that a human programmer might supply.

The quality of the training data is crucial to any machine learning process. To train our facial age estimation algorithm, we use millions of images from Yoti users who have opted in to this use of their data. The process is explained to them at onboarding, and is discussed in more detail on p23 in the Appendix to this paper. They are free to opt out of this research at any time simply by selecting to Opt Out in the Yoti app's settings. Most Yoti users want to make their lives safer and simpler, and understand that Yoti using their data for internal research purposes is how we are able to improve and develop the products and technology to achieve this. We publish white papers that demonstrate such applications. Yoti does not scrape any websites such as school yearbooks for unconsented use of such images and approximated age to train its age estimation model.

For facial age estimation, these research images are tagged with only two attributes taken from a verified ID document that they have uploaded: their gender and their month and year of birth. Supported documents include passports, driving licences and national ID cards. We believe the size, diversity and verified age accuracy of this data set gives Yoti's technology an advantage over competing solutions.

How accurate is facial age estimation?

When presented with a clear facial image, our technology compares very favourably with human abilities. Anyone who has used a complicated computer spreadsheet will recognise that in some areas, computers are better than humans at doing some things.

Humans tend to systematically underestimate the ages of older people, and overestimate the age of younger, and as we ourselves get older, our ability to estimate accurately tends to decrease. When viewing a succession of faces, a person's judgement tends to be influenced by the faces they have just seen - this isn't a problem that affects facial age estimation. These problems clearly have particular implications for provision of age-restricted goods and services, where we need to check whether teenagers are above or below a required legal age.

Currently, the MAE across the entire data set, de-skewed to give equal weighting to male and female subjects for all 65 age years, is 2.9 years and just 1.4 for 13-17 year olds. Further detail on our algorithm's accuracy, broken down by gender, skin tone and each year of age, is presented in this paper's appendix.

The vast majority of organisations who need to check age need to check whether individuals are over the age of 13, 18 or 21. However, there are additional very important requirements for checking individuals are under an age of interest. We recognise that we still have further to go to reduce bias for older age groups, particularly individuals with skin tone V & VI. However, these older individuals are not materially disadvantaged when the age of interest is for example 18 or 21 and the thresholds are usually 25 or 30 respectively.

Yoti's facial age estimation has been certified by the Age Check Certification Scheme for use in a Challenge 25 policy area. The ACCS report is available at: <https://www.accscheme.com/registry>.



About 'Mean Absolute Error'

Yoti facial age estimation can make both positive and negative errors when estimating age (that is, it can estimate too high, or it can estimate too low). By taking 'absolute' values of each error we mean ignoring whether the error is positive or negative, simply taking the numerical size of the error. We then take the average (or 'arithmetic mean') of all those absolute error values, producing an overall 'MAE'.

The average MAE can be measured as;

- i) the average of each year's MAE - eg. there are 65 year MAEs in the 6-70 age range,
- ii) the average of each age MAE - all age ranges are shown on pages 5-7,
- iii) the average of all the images in the training data (but this data may be skewed towards certain ages with more training data).

There is a clear need for regulators to ensure MAEs are measured independently and consistently to ensure trust in the accuracy of models

Yoti facial age estimation accuracy									Mean estimation error in years split by gender, skin tone and age band	
Gender	Female				Male				All	
Skintone	Tone 1	Tone 2	Tone 3	All	Tone 1	Tone 2	Tone 3	All		
6-12	1.3	1.4	1.6	1.4	1.2	1.4	1.3	1.3	1.3	
13-17	1.3	1.5	1.6	1.5	1.0	1.3	1.5	1.3	1.4	
18-24	2.4	2.5	2.6	2.5	2.0	2.0	2.2	2.0	2.3	
25-70	2.9	3.3	4.6	3.6	2.6	3.3	3.6	3.2	3.4	
6-70	2.5	2.9	3.8	3.1	2.3	2.8	3.0	2.7	2.9	

4. H. Han, C. Otto, X. Liu and A. K. Jain (2015) *Demographic Estimation from Face Images: Human vs. Machine Performance*, IEEE Trans. PAMI, Vol. 37, No. 6, 1148–1161 <https://doi.org/10.1109/TPAMI.2014.2362759>. See also Clifford CWG, Watson TL, White D. (2018) *Two sources of bias explain errors in facial age estimation*. R. Soc. open sci. 5:180841. <http://dx.doi.org/10.1098/rsos.180841> and Voelkle, Ebner, Lindenberger & Riediger (2012) *Let Me Guess How Old You Are: Effects of Age, Gender and Facial Expression on Perceptions of Age*. Psychology & Aging, 27 No.2 265–277. <https://doi.org/10.1037/a0025065>

Safety buffers

To manage this potential for errors, we recommend using facial age estimation as part of a strategy such, as the British Beer & Pub Association's 'Challenge 21'⁵, which is already widely adopted by publicans and their bar staff in England and Wales. This works as follows: certain goods and services can only be sold to customers over a particular age (e.g. 18 years old). However, it is difficult for human staff to be sure whether someone is over 18 just by looking at them. Conversely though, it is fairly easy to tell if someone is significantly older than 18, and customers in this age range would find it an unjustifiable inconvenience to have to show ID to prove their age. Therefore, the store's policy is to only require customers to prove their age if they appear to be under 21. Most supermarkets in England use a Challenge 25 policy.

Facial age estimation can be configured to work with legal age thresholds in a similar way. Unlike human staff, facial age estimation's capacity for error is well quantified statistically. This makes it easier to choose a suitable buffer that is comfortably outside the technology's margin of error, and configure the system to estimate whether customers are above or below that threshold.

As an example, consider the situation in the USA, where the selling of alcohol is restricted to over 21s, and common practice today is for retailers to challenge people who appear to be under 40. In this case, a retailer using facial age estimation might choose to set an initial threshold of 30. If it estimates that the customer is at least 30 years old, then no further age checking is required. This would provide a positive error rate of 0.04% using our current algorithm, and a false negative rate of 0.03%. This compares very favourably with human staff.

If it estimates that the customer is below 30, then they will be directed into a user flow where they need to present documentary proof of their age. For example, a customer could use their Yoti app that is pre-verified to their passport, driving licence or national ID card, or in a retail setting, revert to an existing ID check by staff. Other, less privacy protecting, checks are frequently used online, such as a credit card check, mobile phone check or an identity verification check using ID documents.

Since early 2019, we have spent much time reviewing the appropriate size of buffer for a number of use cases. We have come to the conclusion that this depends on a number of variables. The primary one is the demographic of users. The under 18 age group is the chief area of concern for regulators globally in terms of age restricted goods and services. Given the improvements in accuracy of facial age estimation for this demographic, we now suggest a buffer of 3–5 years for highly regulated sectors (e.g. adult content, gambling, alcohol, tobacco) as most appropriate for the 13–25 age band. whereas no buffer may be deemed fine for social media or gaming use case. In some countries, more cautious regulators may initially look for a higher buffer. For a jurisdiction with legal age restriction of 18, and a threshold set to 28 (a 10 year buffer) we would currently have a 0.02% error rate (that is, only 1 in 5,000 14–17 year olds would be incorrectly let through)⁶. With a threshold set to 25 years, facial age estimation current error rate is 0.9%. For a threshold of 21 years, the error rate is again 0.09% assuming equal numbers of 14-17 year olds in the test sample⁷.

For a demographic of senior citizens, such as for a travel entitlement use case, a regulator may consider a buffer of five to seven years more appropriate.

5. See <https://beerandpub.com/campaigns/challenge-21/>

6. For more information see page 36

7. For more information see page 37

However, there is not currently a commercial demand from relying parties or regulators for age estimation of this demographic. This will always be discussed with the relying party and with the relevant sector and jurisdiction regulator. Over time, as the accuracy of age estimation technology improves, regulators will be able to set lower buffers with confidence.

In the appendix of this paper, more statistical detail on facial age estimation ‘false positive’ rates can be found by age, gender and skin tone.. It is also worth considering ‘false negatives’ too (where facial age estimation incorrectly estimates someone as being younger than the threshold age), as these can be a source of unwanted friction. These are also discussed in the appendix.

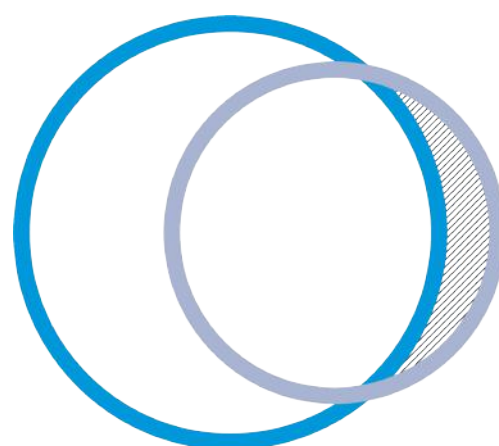
Public acceptance of AI technologies

When discussing the accuracy of facial age estimation, it is worth considering a general point about machine learning and the public’s attitude to AI technologies of all kinds: namely, how unforgiving humans tend to be in regard to mistakes made by AI.

Whilst we feel it is fair to claim that the accuracy of facial age estimation generally compares very favourably with human judgement in the broad majority of cases, there will inevitably be rare occasions where it ‘makes mistakes’. Of course, humans make mistakes too. However, sometimes machine learning systems make mistakes that no human would have made. This is illustrated in the Venn diagram.

As can be seen, typically, humans make errors, just as a well-trained machine learning system does. In most of the cases where the system gets it wrong, a human would make the same mistake. Humans tend to be much more bothered by the small percentage of cases on the right of the diagram – cases where the machine learning system makes a mistake, but a human would not have been fooled. It can be argued that this is an irrational reaction.

Nevertheless, the general public may often unduly focus their attention on the machine failings, until they become comfortable with the new technology. We believe that digital approaches can be harnessed to support age appropriate design of services, enabling data minimisation, improving online safety and countering certain online harms. One of the objectives of this white paper is to support the education of the public.



■ Errors made by humans
■ Errors made by machines
▨ Errors humans react more badly to

Learnings from practical use

Facial age estimation works quickly, returning an age estimate in around 1 second. 97.5% of phone users submitting a face image are successfully age estimated. The user needs to present their face to the camera, uncovered (although glasses do not usually present a problem). We recognise that in some areas, internet speed can be challenging; we can cater for small image sizes of 50-100KB. We have scaled to handle tens of millions of checks per day, and we are currently able to handle up to 300 checks per second, but could scale to do more if needed.

Dim lighting is not helpful; bright ambient light works best. Our research has found that the effect of beards and facial disfigurement can make a minor impact, but does not materially affect estimated ages. In response to the COVID-19 pandemic, we have been researching how facial age estimation copes when a person is wearing a mask covering the lower half of the face. Results suggest that whilst accuracy is reduced somewhat, acceptable performance can usually still be achieved as long as an appropriate safety buffer is used.



222

Countries and territories



4

Scripts



19

Languages



97.5%

Success globally

Yoti facial age estimation can scale to handle large volumes



300

Checks a second

18K

Checks a minute

1.1m

Checks an hour

26m

Checks a day

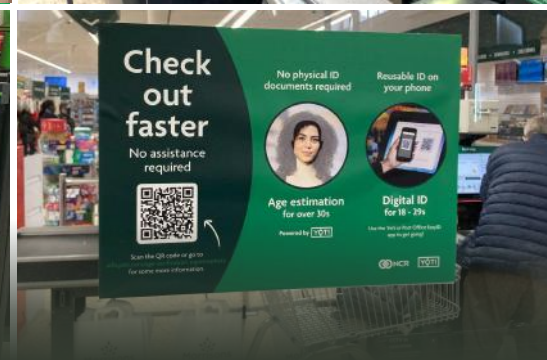
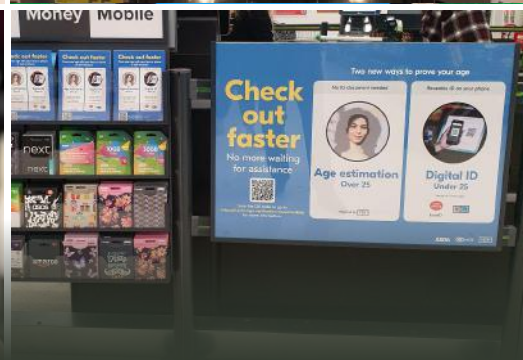
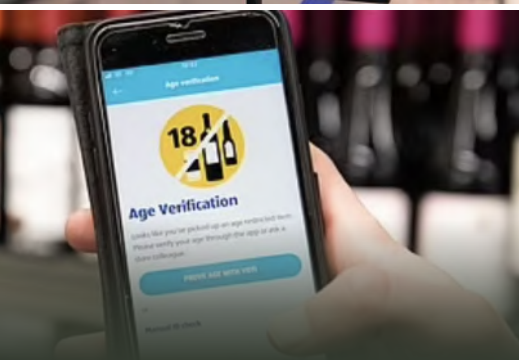
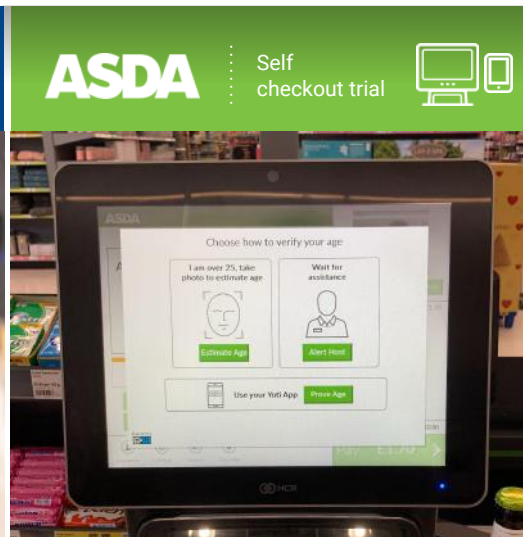
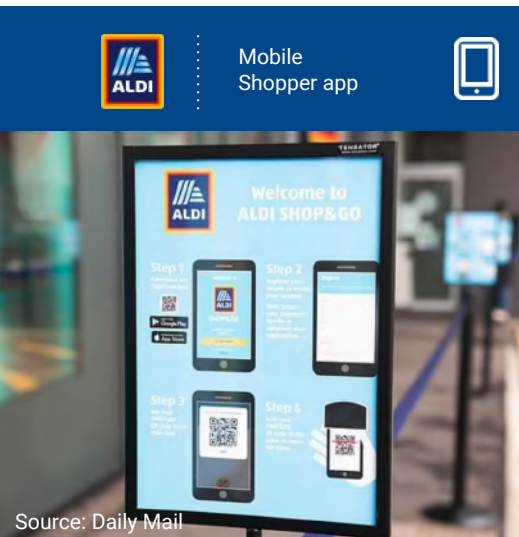
780m

Checks a month

UK government live trials

Last year, UK supermarkets – including Asda, Morrisons, Tesco and The Co-op – trialed our digital age verification at self-checkouts in a scheme [run by the Home Office](#). Key takeaways from the trials:

- Participating supermarkets confirmed they support digital age verification, and would welcome legislative change in this area.
- There were no reported sales of underage customers purchasing age restricted items when using our age verification technology.
- Informed consent was gathered from all customers, who were given a choice whether to use the technology or present an ID document to a member of staff.
- The majority of shoppers who used Yoti digital proof of age solutions liked the technology and would use it again, once available.
- Digital age verification technology provided an opportunity to reduce the number of physical age interventions, giving retail staff more time to monitor other activities, including spotting proxy sales.
- Yoti facial age estimation is more accurate than humans which reduces the risks of incorrectly estimating the age of shoppers.
- Yoti facial age estimation is more inclusive because anyone who looks over the required age threshold does not need to carry around a physical ID to prove their age.
- Digital age verification supports the ability for retailers to achieve the Licensing Objectives



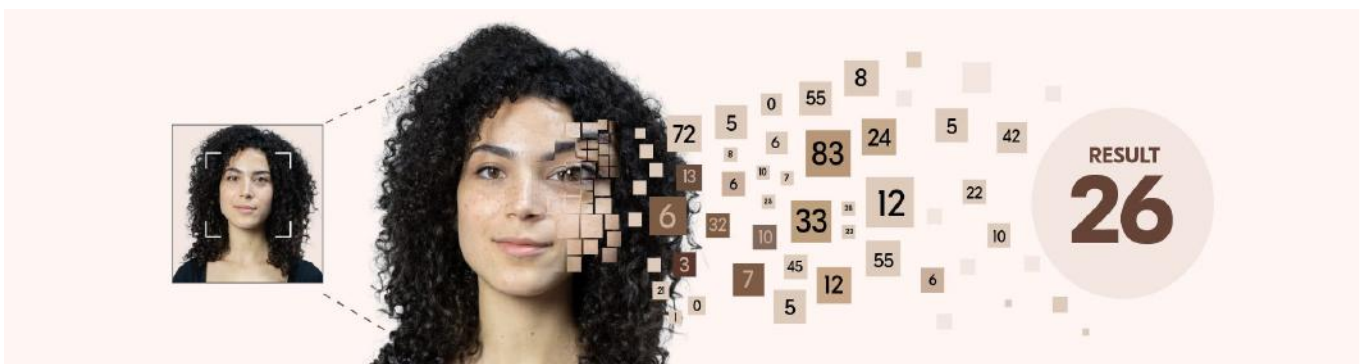
Supporting Children's Codes

Given the growing importance of age checking online for younger children and teenagers, we have introduced additional training data to enhance our algorithm to estimate the age of 6 to 12 year olds.

The Age Appropriate Design Code, originating in the UK, is driving a global movement to design online interaction 'age appropriately' across the 4 C's - be that content, conduct, contact or contract⁸. The challenge for designers and platforms is to enable young people to be supported to thrive online whilst also enabling age appropriate interaction, protecting against detrimental content, grooming and supporting age appropriate content moderation. We can support platforms to recognise child users and not employ nudge techniques, encourage children to provide unnecessary personal data, or make a child's real time location publicly available. Children should no longer be encouraged to stream to large groups of unknown adults. There are a growing number of countries and states around the world also reviewing legislation for a range of age restricted goods and services; in particular age assurance for access online. Notably, the Californian Age-Appropriate Design Code Act will become effective from 1st July 2024 with significant fines for non-compliance. There are also adult content sites already using Yoti age estimation successfully to prevent children from accessing their websites.

Obtaining consented data to develop our software to accurately estimate 6 to 12 year olds has been a significant challenge. We have worked hard to ethically obtain parental consent to use anonymous images of children in our training data; that is facial images with month and year of birth. We can now correctly estimate 63% of images of 16 year olds to be between 15 and 17. For 6-12 year olds, our first MAE results are 1.3 years, so could be used effectively for triaging access to 13+ apps.

We will continue to invest more to improve our accuracy to make the internet safer for young people.



Detect face

A face is detected in an image and reduced to pixels. Each pixel is assigned a number that the AI can understand.

Compute numbers

The numbers are computed by a neural network that has been trained to recognise age by looking at millions of images of faces.

Determine age

The AI finds a pattern in the numbers and produces an age.

8. Livingstone, S. and Stoilova, M. "The 4 Cs, Classifying Online Risk to Children." SSOAR, 2021. <https://doi.org/10.21241/ssoar.71817>.

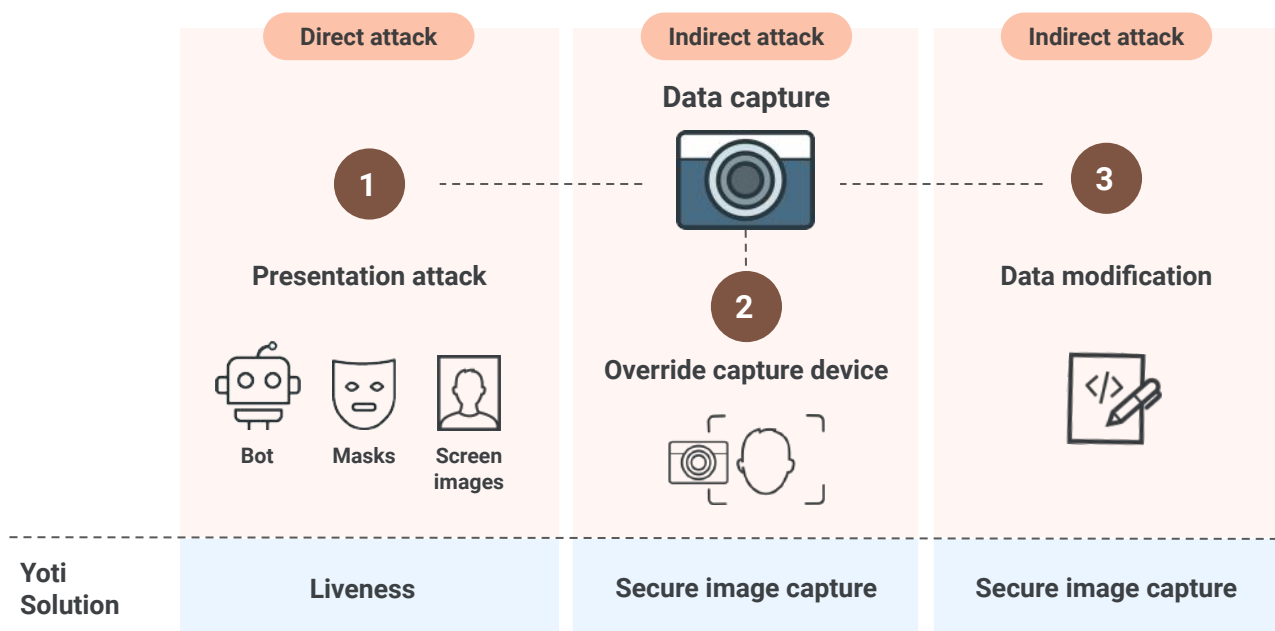
Ensuring accurate and genuine assurance

The first natural question around age estimation is 'how accurate is it'? As discussed, by using buffers according to your age of interest, facial age estimation can be used with a very high level of confidence.

The second, equally important question, is how secure is the process? Can it be spoofed, are bad actors able to hack into the system to override checks, images or results? There's little point estimating a face accurately if it's not the real face in front of the camera.

This is why it is important to use a combination of technologies to secure a high level of assurance. There are a number of threat vectors, as illustrated in the diagram below.

Data capture attack threats



Step 1 as a direct spoofing attack - an attempt to present an image, mask or video, often called a presentation attack. This is an attempt to spoof a check by appearing older or to be another person. To overcome this we use our NIST 2 certified liveness technology⁹. This ensure that the person undertaking the check is a real person and not someone wearing a mask, or presenting to the camera a picture or screen of another (older or younger) person.

Steps 2 & 3 is a newer, more sophisticated, but relatively easy way for technically competent individuals to spoof the system. They are called injection attacks. An injection attack involves injecting an image or video designed to pass authentication, rather than the one captured on the device camera. Using free software and some limited technical ability, a bad actor is able to overwrite the image or video of the camera with pre-prepared images.

Yoti has developed a solution, called SICAP (**Secure Image CAPture**) for which we have filed a patent, that makes injection attacks much more difficult for imposters. It is a new way of adding security at the point an image is being taken for a liveness check¹⁰.

9. [Yoti MyFace Liveness white paper](#)

10. [How Yoti can help combat injection attacks](#)

Product developments

Our R&D and product teams continue to improve our service, not just in terms of the accuracy of the age estimation algorithm, but also solving practical problems of deploying the service in different environments. We work very closely with our partners across many sectors and regulatory frameworks across many countries.

Anti-spoofing technology - Yoti MyFace® **LIVE**

We've developed proprietary anti-spoofing technology - Yoti MyFace - to prevent fake images being used for age estimation. Our proprietary passive liveness technology analyses the depth of an image to make sure it's a real person and not a photograph, video or bot. As one would expect we measure bias on our liveness detection technique. As of March 2023 Yoti's MyFace solution is NIST Level 2 approved with 100% attack detection. This means an experienced person with a budget of \$300 has presented hundreds of paper, screen and mask attacks and MyFace correctly detected every attack.

Injection attacks - SICAP - **LIVE**

An issue of using images taken by a laptop or mobile device is the risk of injection attacks. This is where bad actors are able to swap the image used for verification with a different image designed to pass. We have developed, called SICAP, a unique way to detect injection attacks, for which we have applied for a patent.

Image fidelity - **ONGOING**

We continue to develop our algorithm to work with lower quality cameras and smaller file size images. This is crucial to ensure the Yoti solution can be successfully used on millions of already in service terminals which use older, less advanced cameras.

Enhanced auto face capture (AFC) - **ONGOING**

As part of our ongoing efforts to improve usability and increase success rates we have introduced an auto face capture module. This reduces friction by removing the need for a user to fit their face into an outline face frame on a screen. Our enhanced AFC is able to recognise when a suitable picture will be captured making Yoti age estimation easier and quicker to use and improving success rates.

Running multiple models - **Q2 2023**

We will soon be deploying multiple models to our production environment. This will significantly add flexibility for regulators and customers to use whatever version they feel is better suited to their particular scenario. A business could choose a model optimised for estimating the age of younger children.

Legal compliance

Whilst legal compliance is a complex area, it is important to cover given there are understandable concerns about the potential unlawful use of personal and biometric data by governments and businesses.

Yoti's facial age estimation complies with EU GDPR, and also our own ethical approach to user data and privacy. KJM (Kommission für Jugendmedienschutz), the German regulator, approved Yoti facial age estimation for online age assurance for online age-restricted content in November 2021. When clients use facial age estimation to age verify their users, Yoti acts as the data processor, with clients as the data controllers. Our clients therefore need a legal basis to use facial age estimation according to their own jurisdiction.

The Yoti Age Portal has a consent option built in so clients can easily collect consent if that is the lawful basis under which they operate.

Yoti's facial age estimation does not involve the processing of special category data - this has been confirmed by the UK Information Commissioner's Office. This is because the age estimation model is unable to allow or confirm the unique identification of a person, and it has not been trained to, and is not being used for the purpose of identification (and that is the key test for special category data). Put simply, if you put the same face into the model several times, the model would have no idea it is the same face (and no way of working that out) and it would give slightly different age estimation results each time. The model was not trained to recognise any particular individual's face, but instead to categorise a presented face into an age.

Definition of special category data in Article 9 of the UK GDPR:

*Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, **biometric data for the purpose of uniquely identifying a natural person**, data concerning health or data concerning a natural person's sex life or sexual orientation*

Recital 51 of the UK GDPR further says that:

*The processing of photographs should not systematically be considered to be processing of special categories of personal data as they are covered by the definition of biometric data only when processed through a specific technical means allowing the **unique identification or authentication of a natural person**.*

For more information about why Yoti's facial age estimation does not process biometric data, please see our blog [here](#).



Fair, standardised measurement

In the May 2022 white paper we introduced accuracy levels (MAE) for each year of age, across gender and skin tone, from age 6 to age 70. This helps customers and regulators look at the age ranges that interest them, rather than only see an average from a group of years, where there may be imbalances in test set sample sizes, particularly within large age ranges. It also prevents Yoti, or future competitors cherry picking a single or small number of years of age to misrepresent a model's accuracy.

Mean Absolute Error and Absolute Error Standard Deviation

For the March 2023 white paper we are introducing AE standard deviation measures, alongside MAE. MAE demonstrates we have a good average performance, but this does not signify a guaranteed error range.

On feedback from regulators and academics, we are now adding AE standard deviation as an additional measurement of accuracy. Standard deviation can add qualification to the mean by saying, for example, we have a low deviation, ie - our error rate is consistent.

A higher standard deviation tells us that the errors are spread over a bigger range. A lower standard deviation indicates errors across the data tend to be of a similar range (or more standard).

By publishing both, we hope to help both relying parties and regulators to form a better view of, 'how accurate is age estimation'. For full standard deviation results by age band please refer to p31.

In addition, at the request of one of our clients, our May 2022 white paper has been independently verified by the ACCS as to the measurement methodology and accuracy of our results. The ACCS had this to say:

"The training, testing and results reporting presented in the Yoti white paper have been independently validated by ACCS, who have certified that Yoti have deployed appropriate methodologies to analyse the performance of their Facial Age Estimation algorithm, including ensuring appropriate separation of machine learning training data, testing data and validation data."

Yoti's commitment to ethical use of AI technologies

At Yoti, we take our ethical responsibilities as a company developing new technology very seriously.

Our Data Protection Officer has completed a formal Privacy and Ethics Impact Assessment for Yoti age-checking solutions, which is available on request to potential clients. It covers Yoti both as a data controller for our own use of agechecking solutions with our own users, and as a data processor when offering age-checking solutions to corporate customers.

We have set up an internal Ethics Committee with members from several different areas of our business, to consider ethical issues related to our technology and its use. We used frameworks such as 'Responsible 100' and 'Digital Catapult' as starting points for the scope of these considerations. Findings of the committee are shared with Yoti's senior management teams, Board of Directors and our Guardian Council.

External scrutiny

We have also obtained an ISAE 3000 assurance report from one of the top four global auditing firms, validating our age checking services as being in accordance with the British Standards Institution's PAS 1296 code of practice.¹¹

The German Association for Voluntary Self-Regulation of Digital Media Service Providers (FSM) awarded us its Seal of Approval for our age verification solutions¹².

We have hosted three roundtable sessions to get feedback from a range of industry practitioners on unintended consequences of our approach. Participants from the UK included the University of Warwick, the University of Keele, the Home Office Biometrics Ethics Committee, the Children's Commissioner for England, the NSPCC, the ICO, GCHQ, and groups such as Women Leading in AI, and techUK¹³.

We have also been actively reaching out to organisations representing various minority groups to seek their views and input, including the UK transgender charity, Sparkle.

We have asked the US Centre for Democracy & Technology to perform a deep dive with full access to our CTO and tech team, and we have sought comment from World Privacy Forum and Future of Privacy Forum. We have met with the New Zealand government and a Danish Member of Parliament.



11. PAS 1296: 2018 *Online age checking—Provision and use of online age check services—Code of Practice*. Available from the British Standards Institute shop.bsigroup.com.

12. <https://www.fsm.de/de/fsm.de/yoti>

13. <https://www.yoti.com/blog/age-estimation-technology-tackles-grooming-online/>

Appendix

This appendix provides further detail on the current accuracy of facial age estimation. Taking confidence from the trends we've seen in past months (illustrated below), we expect these figures to continue to improve as the volume and diversity of our dataset increases.

Data used to build the model ('training data')

Yoti has invested significantly in building a leading R&D team since early 2015, working on a variety of AI initiatives.

The current production model of facial age estimation (March 2023) was built using a training data set taken mainly from Yoti apps' users (though not US users). We provide information to users at onboarding about our use of biometrics with links to more details, including the Privacy Notice¹⁴ where the use of user data by our R&D team for internal research is extensively detailed. The screenshots overleaf show the current onboarding screen and the screen where users can opt out of their data being used for R&D activity.

Any user can go to the app settings at any time and opt out of R&D use of their data. This prevents further data from that user being sent to R&D, and it deletes all the data associated with that user that is on the R&D server and available for R&D to use. We have chosen to automatically delete the existing data when a user opts out or deletes their account, even though we do not legally have to under the research provision in GDPR article 17(3)(d).¹⁵ We employ a privacy-by-design approach (hashed numbering) so that although we can find data of a specific user to action the data deletion, there is no way to recreate a specific user's identity from that R&D data.

To enhance our coverage of particular demographics, further age-verified images were gathered by Yoti with consent in Nairobi, Kenya. Through the Share2Protect campaign, we have enabled parents and children to support the extension of the facial age estimation to extend to 6-13 year olds.¹⁶ We have also purchased further parent consented child facial images, with month and year of birth, and we undertook thorough due diligence on all our data sources.

In 2021, Yoti was part of the ICO Sandbox to extend Yoti facial age estimation AI programme to under 13 year olds without ID documents.¹⁷ Ending in May 2022 the ICO has since published their exit report. Our participation helped ICO gain insights into age assurance for young people; that facial age estimation does not process special category data; that ICO have updated their guidance on special category data as a result of the sandbox.¹⁸

14. <https://www.yoti.com/privacypolicy>

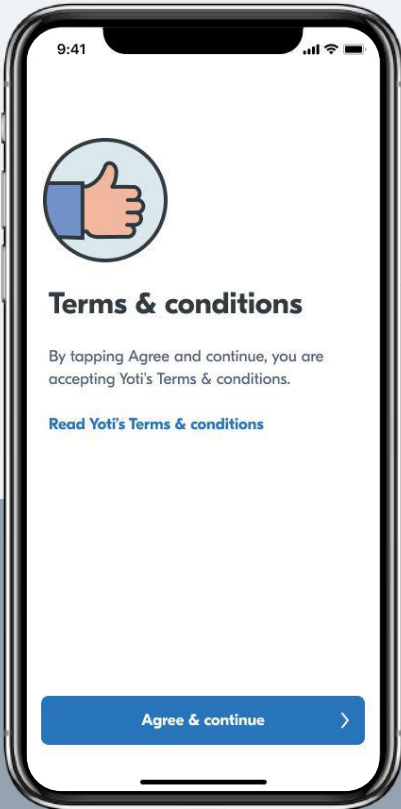
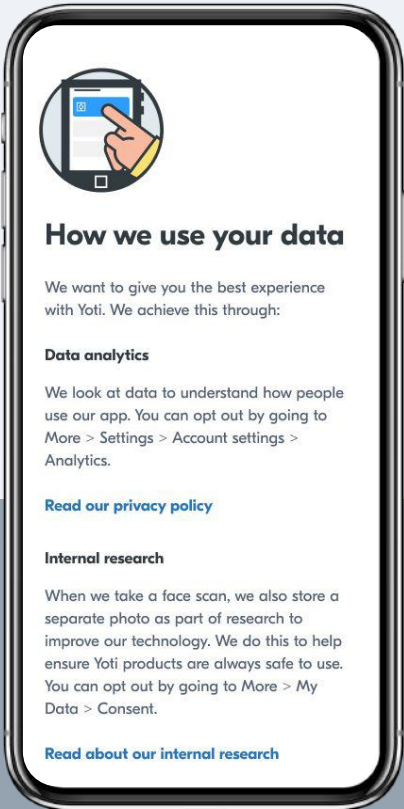
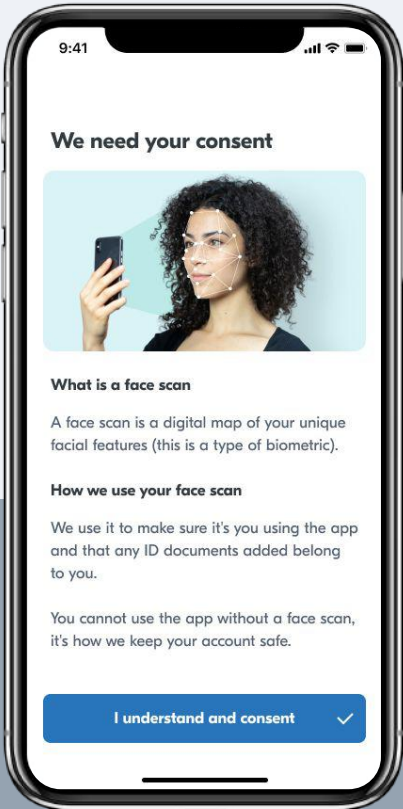
15. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) <https://eur-lex.europa.eu/eli/reg/2016/679/oj>

16. <https://www.yoti.com/blog/protecting-kids-safer-internet-day-2021/>

17. <https://ico-newsroom.prgloo.com/news/ico-supports-projects-to-strengthen-childrens-privacy-rights>

18. https://ico.org.uk/media/for-organisations/documents/4020427/yoti-sandbox-exit_report_20220522.pdf

On-boarding and R&D opt-out screens in the Yoti app



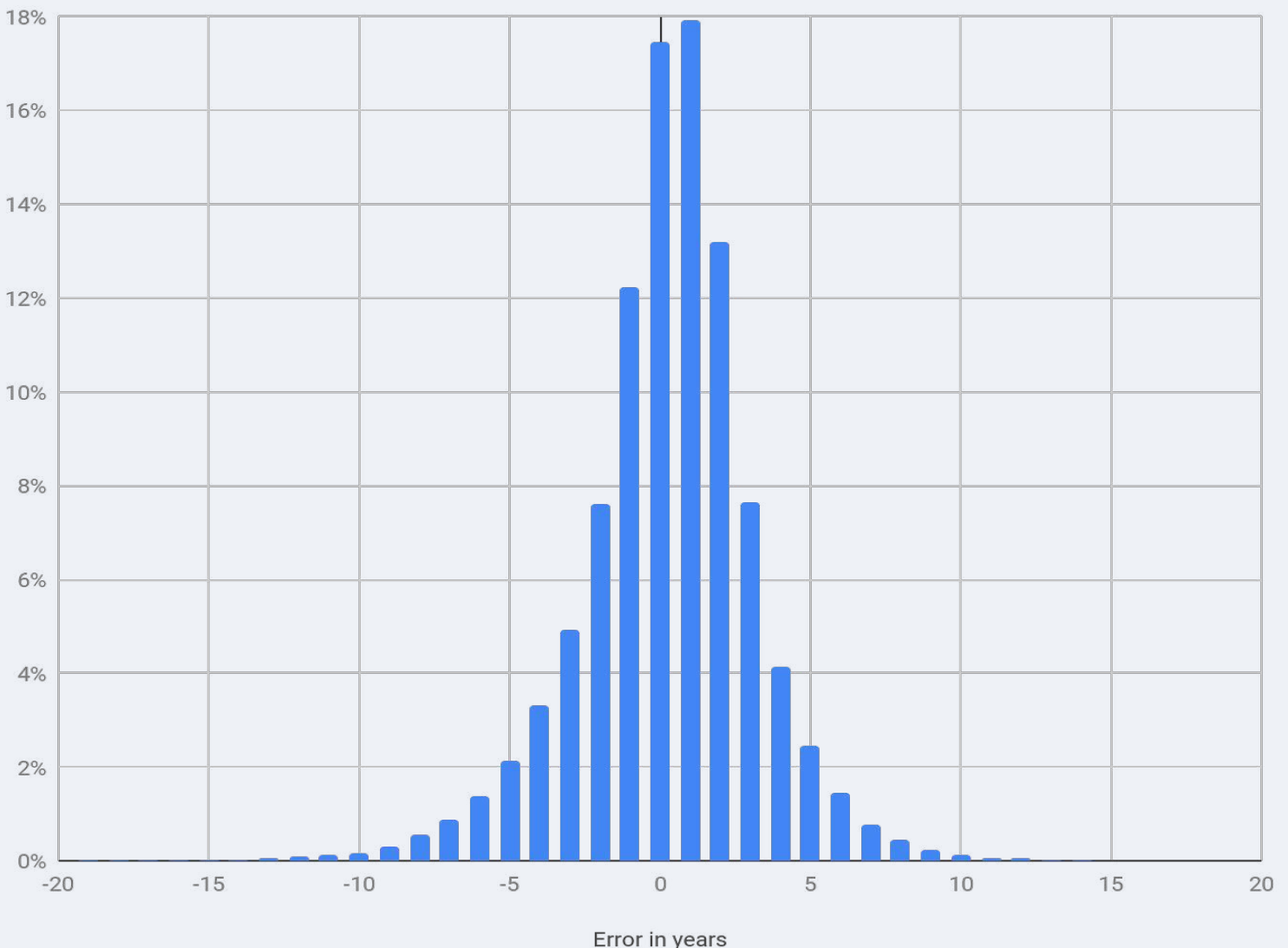
We provide information to users at onboarding about our use of biometrics with links to more details, including the full privacy notice, where the use of user data for R&D is extensively detailed. Users can opt out of their data being used for R&D activity at any time, via the settings on the app.

Data used for testing

Our testing data is also taken from Yoti users worldwide (not US users), in the same manner as the training data. We strive to ensure that it represents as broad a demographic as possible, considering age, gender and skin tone, giving us confidence that the results presented in this White Paper will be reproducible in a wide variety of real world situations.

Accuracy across the entire dataset

In our most recent testing of the model, (performed March 2023), we used test data comprising hundreds of thousands of images with a verified age. The MAE across all years is now 2.9 years; for females it is 3.16, for males it is 2.77. This reflects a higher number of males in the training data across most years. The range of errors tends towards a normal distribution, with a standard deviation of 2.94. The standard deviation is a measure of the variance of the data around the mean - this is illustrated in the chart below.



Accuracy by age, gender and skin tone

We have explored how the accuracy (mean absolute error) of facial age estimation varies with age, gender and skin tone. Hundreds of thousands facial images of verified age in our test set were tagged with the subject's gender and skin tone. Gender was taken from the subject's uploaded identity document, and for skin tone, our research team tagged the images using a scheme based on the widely used Fitzpatrick¹⁹ scale. Fitzpatrick is a dermatological test that involves grading skin tone at two different points in time, one before exposure to sun and then after a week's exposure to sun.

Effectively you are measuring the skin's ability to tan (or susceptibility to burn). You would not use a face for this test, you'd use a bit of the body that was normally always clothed. There is reasonably good correlation between ability to tan / burn and ethnicity.

However, we don't do a Fitzpatrick test, we instead take a single-point-in-time photo of someone's face. We have no idea to what extent they've been out in the sun before that photo was taken. This therefore gives much weaker correlation with ethnicity people with the ability to tan and who've been out in the sun can often have facial skin tone that is one or two scale degrees darker than their unexposed skin.

The majority of the tagging was performed using a manual process, with some data tagged automatically. We have put quality procedures in place to help ensure our manual tagging is reliable and free from bias.

For each age, we present the mean absolute error (MAE) in six classes: female (for three different skin tones), and male (for three different skin tones).

For each age, the table also displays:

- the average MAE for females (of all skin tones), calculated as $(\text{MAE for Type I \& II}) + (\text{MAE for Type III \& IV}) + (\text{MAE for Type V \& VI}) \div 3$
- the average MAE for males (of all skin tones), calculated as $(\text{MAE for Type I \& II}) + (\text{MAE for Type III \& IV}) + (\text{MAE for Type V \& VI}) \div 3$
- the overall average MAE, calculated as $(\text{weighted average MAE for females} + \text{weighted average MAE for males}) \div 2$

The average aims to deskew the test data set, so as to present equal contributions from the three skin tone groupings and both genders.

Skin tone scale



19. Fitzpatrick, T, (1988) *The Validity and Practicality of Sun-Reactive Skin Types I Through VI*. Archives of Dermatology 1988; 124 (6): 869–871

Standard deviation of absolute error

Age Band	Gender								All
	Female				Male				
	Skin Tone (Fitzpatrick Scale)								
	Type I & II	Type III & IV	Type V & VI	All	Type I & II	Type III & IV	Type V & VI	All	
	SD	SD	SD	Average SD	SD	SD	SD	Average SD	Average SD
6-9	1.2	1.1	1.1	1.1	1.0	1.2	1.3	1.2	1.1
10-12	1.0	1.2	1.5	1.2	0.8	0.9	0.7	0.8	1.0
13-15	1.0	1.1	1.1	1.1	0.8	1.0	1.0	0.9	1.0
16-17	0.9	1.1	1.1	1.0	0.8	1.3	1.1	1.1	1.1
18-24	1.6	1.6	1.6	1.6	1.4	1.7	1.9	1.7	1.6
25-29	2.1	2.4	3.1	2.6	1.7	1.8	2.3	2.0	2.3
30-39	2.3	3.0	3.7	3.0	1.9	2.2	2.3	2.1	2.6
40-49	2.4	3.6	3.5	3.2	2.1	2.8	2.4	2.4	2.8
50-60	2.4	3.0	5.8	3.7	2.3	3.3	3.2	2.9	3.3
60-70	2.7	3.5	3.6	3.3	2.3	4.3	3.1	3.2	3.3
All	2.0	2.5	3.0	2.5	1.7	2.4	2.4	2.2	2.3

Absolute versus percentage errors

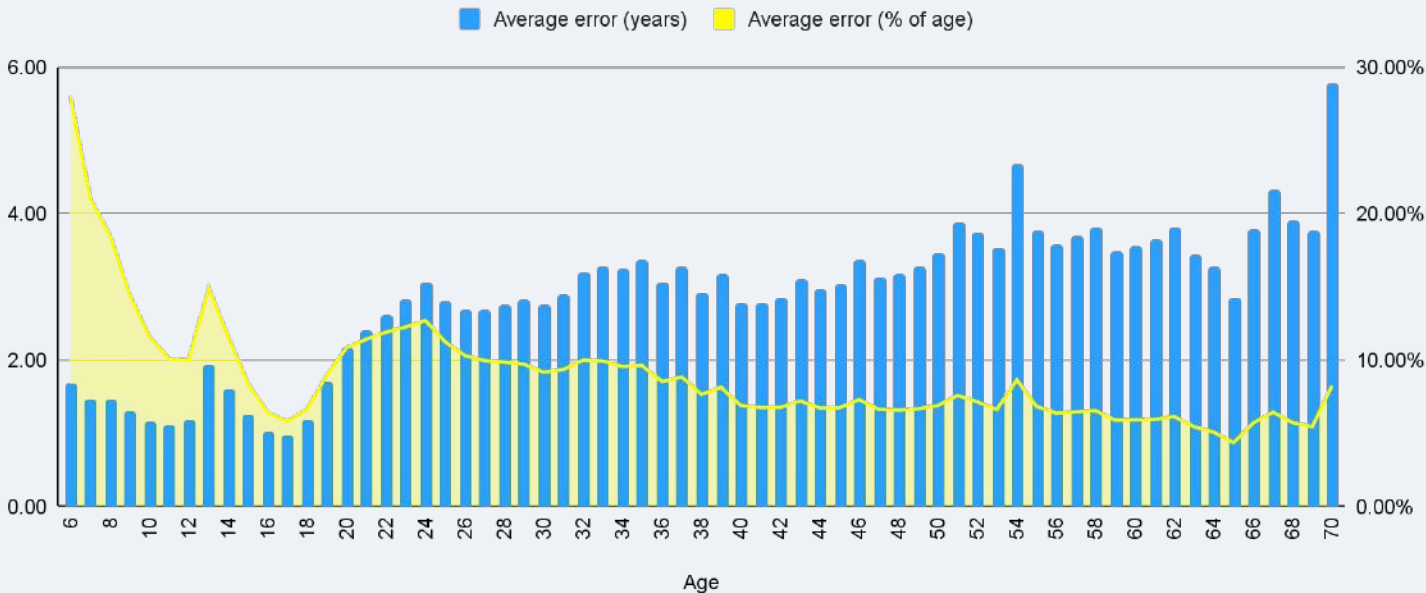
When we started publishing mean absolute error values for teenagers, a key age of regulatory interest, our MAE in April 2019 was 2.9. Some stakeholders felt it unlikely that it would improve sufficiently to become an efficient age assurance technique. However, our May 2022 MAE for teenagers is now 1.4 years.

For our first set of children aged 6-12 and after a much shorter period of research and smaller training data set, the average error is now 1.3 years. We believe it is very likely our MAE for 6-12 year olds will improve as our training set increases. The MAE of 1.3 years means Yoti can already offer a highly effective age estimation solution for businesses wishing to, or being required to, offer age appropriate design.

We have sufficiently high volumes of training data for males of all skin tones across the ages 6 to 29 - our average MAE for 6-29 year old males is 2.1. It is for I (lightest) skin tone males, 2.0 for skin tone II males, and 2.5 for III (darkest) skin tone males, meaning there is less than 5% difference between the highest and lowest skin tone accuracy across these 24 years of age. We have since even up the data between male and female, in particular, darkest skin tone females. Our corresponding MAE for 6-29 year old females is 1.9. It is 1.7 for I (lightest) skin tone females, 1.7 for II skin tone females and is highest at 1.9 for skin tone III (darkest) females.

It is also worth noting that although the magnitude of error may appear larger for older age bands, when considered as a percentage of the subject's age, it often is more accurate in relative terms. For instance, an error of 2 years for a 15 year old is a 13% error, whereas an error of 2 years for a 50 year old is an error of 4%. This is illustrated in the chart below.

Average Error and Error in % of Age



Improvement in accuracy as the training data set grows and changes

As mentioned above, we believe the differing mean absolute error shown for different groups (age, gender, skin tone) correlates strongly with how well-represented those groups are in the training data set. We periodically retrained our age estimation model on an ever-expanding data set of millions of face images, as millions more people create their Yoti app digital IDs. The charts below illustrate the significant improvements in accuracy that we have observed over time. The size and composition of our test data has itself changed (diversified) over this period too, so the comparisons from one model's results to the next are not absolute, however the overall trend is clear and encouraging. Where appropriate, we will endeavour to undertake further targeted fieldwork in this regard.

N.B. From September 2021, we have revised our approach to concentrate on achieving a reduction on bias, even where this may have a detrimental effect on accuracy.

Since our May 2022 update, we have removed some older images from both our training and testing data sets. This complies with our privacy policy on customer data retention, where if a user has been inactive for over 3 years, we delete their data. This will have two implications to note:

- **Training data** - where deleted data may have a skewed number of images in a certain subcategory, this may affect accuracy in that data range.
- **Testing data** - changes in this data set will mean results over time are not strictly 100% comparable as each model is not being tested against exactly the same set of test data.

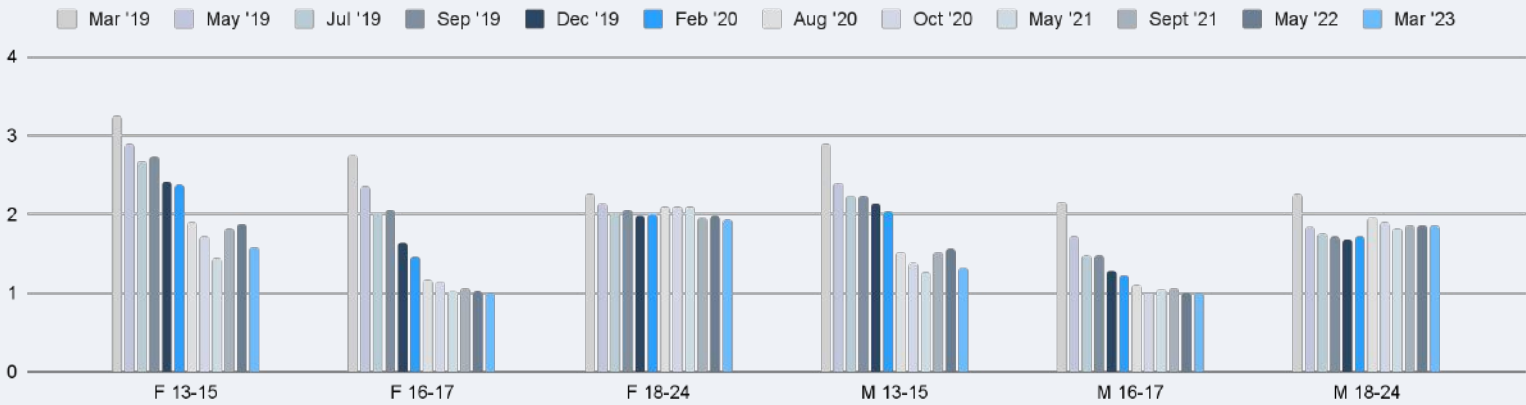
We do not believe the change in training or testing data will materially affect the accuracy of the model over time. We will also monitor churn of our data sets to ensure we replace data with the corresponding demographic that may have any significant effect on our accuracy or testing.

Summary of the performance of the new algorithm:

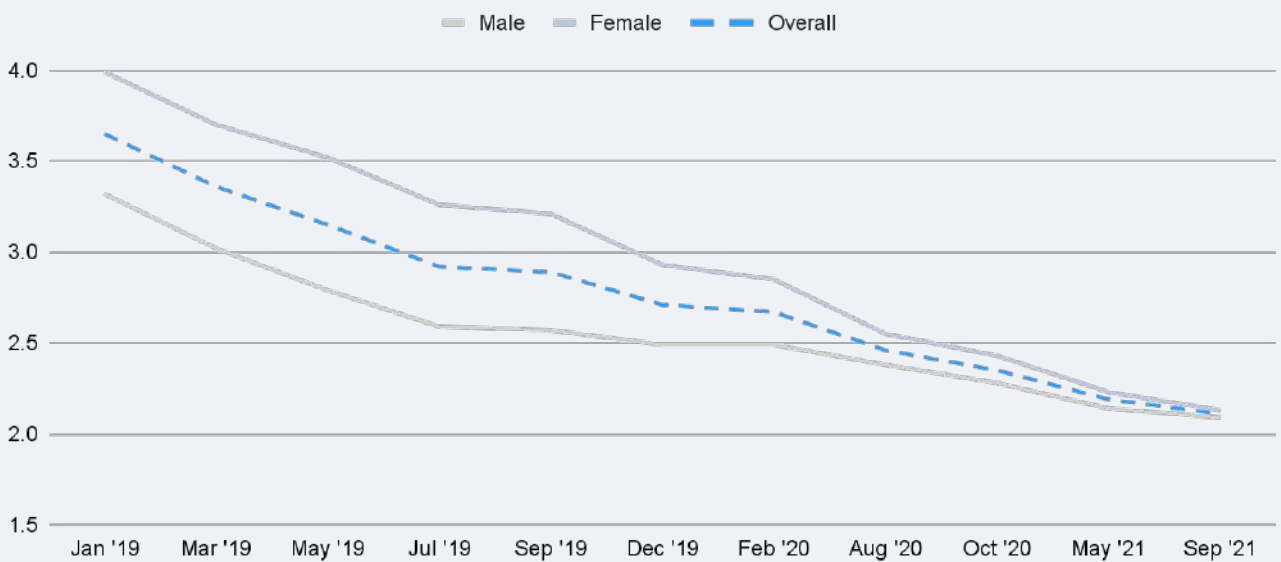
1. We have made improvements in MAE between the ages of 6-10 and 10-19. This was one of two key objectives.
2. The algorithm has reduced bias overall. It shows improvements in both (a) gender equality by improved performance for females and (b) skin tone equality by improved performance for skin tones 2 and 3. This was the second of our key objectives.

Improvement in accuracy as the training data set grows and changes

Average MAE changes between AI models (by Age Group & Gender)



Average MAE changes between AI models (by Gender)



False positives

'False positives' are when we ask a question with a yes/no answer, and the answer comes back as 'yes' when it should have been 'no'. So for example, when dealing with age-restricted goods or services, if we ask 'Is this person old enough to buy alcohol?' and facial age estimation tells us 'Yes they are', but actually they are not, then we have a 'false positive'. In this kind of use case, we can regard false positives as a measure of facial age estimation being too lenient.

Let's define some terms to help quantify things. When dealing with age-restricted goods and services, the **age of interest** is what we call the age stipulated in the relevant law or regulation. For example, in many jurisdictions, the age of interest for buying alcohol is 18, and in many use cases, we will ask 'is this person above the age of interest?' (e.g. 'are they over 18?'), and configure facial age estimation to simply return 'yes, they're 18+' or 'no they're not'.

However, as described earlier in this paper, facial age estimation has a margin of error, and we would expect some false positive replies when asking if a person was above the age of interest (particularly if their true age is close to it). For this reason, we recommend configuring a **threshold age** above the age of interest, to create a safety buffer. Instead of asking facial age estimation if the person is above the age of interest, we actually ask if they are above the threshold age instead. For an age of interest of 18, we might chose a threshold age of 23. We ask facial age estimation whether or not people are over 23. If the answer is 'yes, they are', we accept with confidence that they are over 18.

The challenge, therefore, is to pick an appropriate threshold for the given use case, which delivers an acceptably low false positive rate. The two tables below provide detailed statistics from our testing of facial age estimation, showing false positive rates for different ages of young people, for a succession of threshold ages. The first table considers a scenario where the age of interest is 18, the second table considers an age of interest of 21.

As is to be expected, the results show that it is much easier for facial age estimation to correctly estimate that young teenagers are below a threshold age than people who are only one year away from it. However, when considering the acceptability of false positive rates for any given use case, the risk involved should be considered too: for example, the potential harm in a 14 year old purchasing alcohol is likely to be greater than for a 20 year old.

In the tables below, we also present an average false positive rate for each threshold, weighting the value equally for each age's contribution (regardless of the number of test subjects for that age).

False Positive rates for a selection of thresholds, for an age of interest of 18

(March 2023)

						Average False Positive Rate (weighted equally for each age)
		14	15	16	17	
Test Sample Size		3,167	7,445	10,105	10,214	
Thresholds (years)	20	0.32%	0.67%	1.67%	4.91%	1.89%
	21	0.19%	0.38%	1.02%	2.90%	1.12%
	22	0.16%	0.27%	0.62%	1.57%	0.65%
	23	0.09%	0.13%	0.29%	0.98%	0.37%
	24	0.06%	0.08%	0.25%	0.46%	0.19%
	25	0.00%	0.04%	0.06%	0.27%	0.09%
	26	0.00%	0.00%	0.03%	0.17%	0.05%
	27	0.00%	0.00%	0.01%	0.06%	0.02%
	28	0.00%	0.00%	0.01%	0.06%	0.02%
	29	0.00%	0.00%	0.01%	0.04%	0.01%
	30	0.00%	0.00%	0.00%	0.03%	0.01%

False Positive rates for a selection of thresholds, for an age of interest of 21

(March 2023)

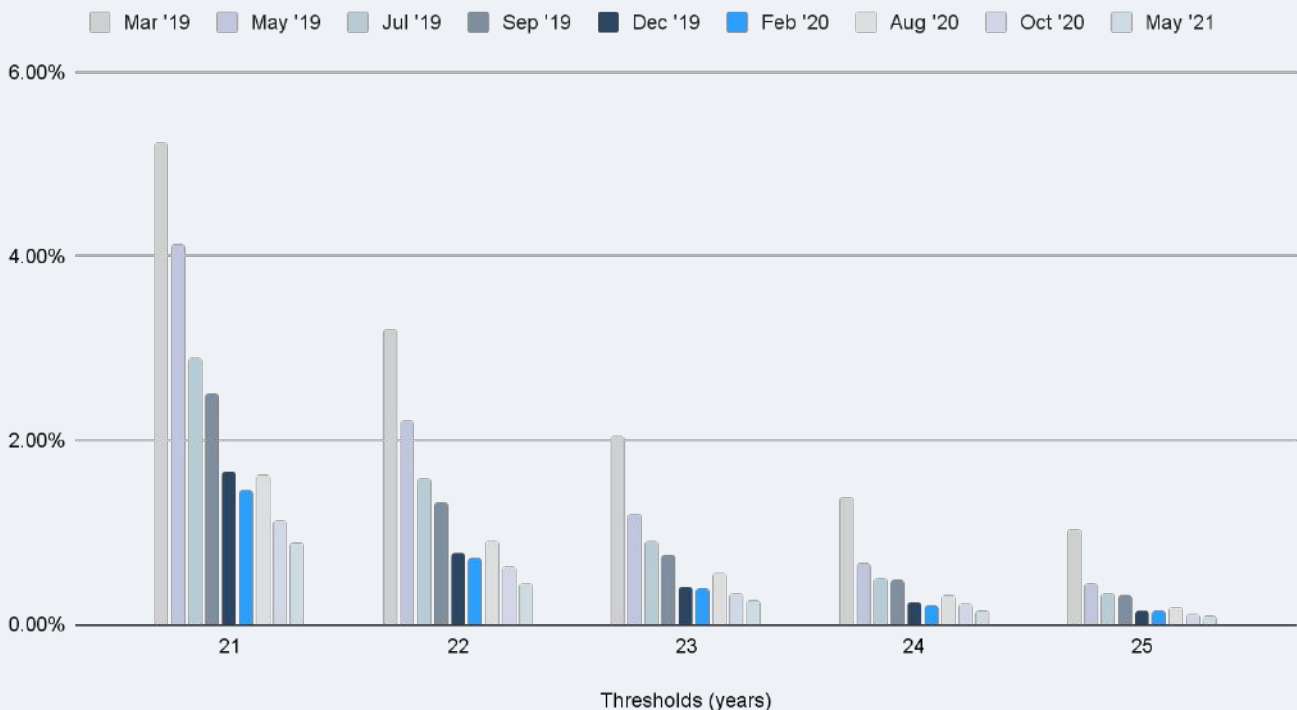
		Actual Age					Average False Positive Rate*
		16	17	18	19	20	
<i>Test Sample Size</i>		10,105	10,214	8,749	4,995	3,918	
Thresholds (years)	24	0.15%	0.46%	1.23%	3.70%	10.03%	3.12%
	25	0.06%	0.27%	0.75%	1.84%	6.20%	1.83%
	26	0.03%	0.17%	0.30%	0.94%	2.96%	0.88%
	27	0.01%	0.06%	0.17%	0.46%	1.43%	0.43%
	28	0.01%	0.06%	0.10%	0.24%	0.69%	0.22%
	29	0.01%	0.04%	0.06%	0.04%	0.23%	0.08%
	30	0.01%	0.03%	0.03%	0.02%	0.13%	0.04%
	31	0.01%	0.03%	0.01%	0.02%	0.05%	0.02%
	32	0.01%	0.03%	0.01%	0.02%	0.05%	0.02%
	33	0.01%	0.03%	0.01%	0.02%	0.05%	0.02%
	34	0.01%	0.03%	0.01%	0.02%	0.05%	0.02%
	35	0.01%	0.02%	0.01%	0.02%	0.05%	0.02%
	36	0.01%	0.01%	0.00%	0.00%	0.05%	0.01%
	37	0.01%	0.01%	0.00%	0.00%	0.05%	0.01%
	38	0.01%	0.01%	0.00%	0.00%	0.05%	0.01%
39	0.01%	0.01%	0.00%	0.00%	0.05%	0.01%	
40	0.01%	0.00%	0.00%	0.00%	0.05%	0.01%	

False positive improvements over time

Our false positive rates have shown steady improvement over the period between January 2019 and May 2021. We are confident this trend will continue as our training data set grows in volume and diversity. This is illustrated for a selection of thresholds in the table and chart below.

Average false positives for 14-17 year old (by threshold) - improvements over time

Thresholds (years)	Jan '19	Mar '19	May '19	Jul '19	Sep '19	Dec '19	Feb '20	Aug '20	Oct '20	May '21
21	9.34%	5.23%	4.12%	2.89%	2.50%	1.65%	1.46%	1.62%	1.13%	0.89%
22	4.11%	3.20%	2.21%	1.58%	1.32%	0.78%	0.72%	0.91%	0.63%	0.45%
23	3.31%	2.05%	1.19%	0.90%	0.75%	0.40%	0.38%	0.55%	0.34%	0.25%
24	2.65%	1.39%	0.66%	0.49%	0.47%	0.24%	0.20%	0.31%	0.22%	0.15%
25	2.14%	1.04%	0.44%	0.33%	0.31%	0.15%	0.14%	0.19%	0.11%	0.10%



We have not included our Sep 21 data because all of the other historic data was calculated on an average MAE for all testing data.

Trade-off between false negatives and false positives

False negatives are an annoyance to those trying to access an age-restricted service or purchase age-restricted goods. They can cause friction and conflict between customers and retail staff, with assaults and abuse being a growing problem,^{18, 19, 20} and customers having to carry physical ID documents. These documents (such as passports and driving licences) can be expensive to apply for and obtain, and a significant proportion of young people do not possess them. Large numbers of physical ID documents are also lost every year, increasing the risk of identity fraud as well as incurring a replacement cost.

Earlier in this paper, when discussing choice of a threshold age and safety buffer for use with facial age estimation, we have generally framed this in terms of trying to minimise false positives (effectively, where facial age estimation is too lenient), as these carry a greater risk of harm to young people. However, it is also sensible to consider false negative rates too (facial age estimation being too cautious). Choosing higher thresholds will tend to decrease false positives at the expense of causing more false negatives. It is important for regulators (or businesses in unregulated sectors) to consider their risk tolerance for any given deployment of facial age estimation, and choose a threshold which is likely to deliver an acceptable balance between false positive and false negative rates.

The table overleaf illustrates this for comparison against a typical 'Challenge 25' retail scenario, where the 'age of interest' (the legal age for buying age-restricted goods) is 18.

For each threshold, the 'false positives' column shows the small percentage of under-age teenagers that facial age estimation would be likely let through. The next column shows the percentage of young people from 18–25 that facial age estimation would be likely to reject, meaning they would have to present physical ID to prove their age instead. Note that this not only includes 'false negatives' (young people who were actually older than the threshold, but facial age estimation incorrectly estimated they were under it), but also 'genuine negatives' (where facial age estimation has correctly estimated that the young person is over the legal age, but they are still below the chosen threshold age).

18. *An analysis of abuse and violence towards retail staff when challenging customers for ID* (Allen & Rudkin, 2017)

<https://nfrnline.com/wp-content/uploads/Abuse-and-Violence-Report-2.pdf>

19. *'It's not part of the job': Violence and verbal abuse towards shop workers—A review of evidence and policy* (Taylor, 2019)

https://assets.ctfassets.net/5ywmq66472jr/22QfMejeWYbimJ9vkX9W9h/0e99f15c0ed24c16ab74d38b42d5129a/It_s_not_part_of_the_io_b_report.pdf

20. *Freedom from Fear: Survey of violence and abuse against shop staff in 2018* (Union of Shop, Distributive & Allied Workers, 2018)

<https://www.usdaw.org.uk/2018FFFReport>

We feel these rates compare favourably with the current 'Challenge 25' scheme, where shopkeepers have to estimate young people's ages, and require all those they think are under 25 to produce physical ID. Depending on risk tolerance, we believe facial age estimation offers clear potential to maintain robust protection for under-18s whilst substantially reducing the numbers of young people over 18 who have to bring physical ID with them when they go shopping.

Comparison of false positives for underage teenagers versus rejection rates for young people over the legal age of interest (18), for a selection of safety buffer thresholds

Choice of Threshold (years)	Average* False Positive Rate (for ages 14-17)	Combined average* rejection rate (false negatives & genuine negatives) (for ages 18-25)
21	1.04%	48.26% (genuine negatives for 18-20 year olds ÷ false negatives for 21-25 year olds)
22	0.63%	55.55% (genuine negatives for 18-21 year olds ÷ false negatives for 22-25 year olds)
23	0.38%	62.39% (genuine negatives for 18-22 year olds ÷ false negatives for 23-25 year olds)
24	0.24%	69.21% (genuine negatives for 18-23 year olds ÷ false negatives for 24-25 year olds)
25	0.17%	76.37% (genuine negatives for 18-24 year olds ÷ false negatives for 25 year olds)

*Note that the numbers of subjects of each age in the test data set was not equal. Therefore to avoid skewing the results, the false positive and negatives figures in this table are averages, weighted equally for the contribution of each age.

Memberships, associations and accreditations



Reviewed by





To find out more visit yoti.com

© 2022 Yoti Ltd