Real-World Adversarial Evaluation of Biometric Age Assurance Services

Madelyne Xiao
Princeton University

Lena Habtu
Swarthmore College

Shaanan Cohney
University of Melbourne

Sarah Scheffler
Carnegie Mellon University

Jonathan Mayer
Princeton University

Abstract

Several dozen U.S. states and national governments have mandated age restrictions on online content. One common enforcement mechanism for these age gates is biometric facial scanning. Despite the widespread deployment of biometric age assurance technologies to meet policy mandates, few public evaluations transparently assess age assurance systems as deployed. Moreover, existing audits of age assurance evaluate performance using controlled exemplar images rather than adversarial conditions where inputs are deliberately designed for age misclassification, as would occur in realistic settings. There is a critical gap in our understanding of the reliability of widely-deployed biometric facial scanning for age assurance.

To evaluate this technology in the context of age assurance policy goals, we conducted a preregistered in-person study of age assurance services with participants aged 13 years and older (n = 94) under adversarial testing conditions.

We asked participants to wear masks, wigs, adhesive mustaches and beards, and other head and facial coverings while sitting for age assurance checks in front of two mobile devices running age assurance apps deployed by three popular commercial vendors: Yoti, Facetec, and Privately.

We find substantial performance limitations for both baseline and adversarial conditions. Even under normal use, services return high null response rates. Under adversarial testing, a costume full-face mask induces age overestimates for users in two out of three services. High false negative rates (i.e., adults unable to pass the check) are also present both at baseline and for adversarial conditions including KN95 masks, costume masks, and eyepatches. Demographic factors, including race, gender, and skin tone, impact performance: dark-complected users experience higher false negative and rejection rates than other users on most evaluation conditions.

The magnitude and prevalence of errors we identify under both baseline and adversarial use, and especially considering demographic differences, raise significant concerns about the reliability of current state-of-the-art systems deployed to meet mandatory age assurance goals.

1 Introduction

Do age assurance technologies work as advertised? Platforms worldwide are turning to commercial facial analysis systems to comply with laws intended to limit access according to a user's age. Yet, public evidence as to their efficacy is severely limited. This paper aims to help bridge that gap.

In November 2024, Australia enacted a first-of-its-kind social media ban that aimed to prevent users under age sixteen from logging onto accounts on popular social media platforms such as Facebook, Instagram, Youtube, and Twitter/X [1]. The Australian law now requires ''that [a] provider of an age-restricted social media platform must take reasonable steps to prevent age-restricted users having accounts with the age-restricted social media platform.'' The eSafety Commissioner identifies commercial facial analysis algorithms as one of a few permissible technical approaches for complying with the law—and one that all the covered platforms have adopted [2–4]. Meanwhile, in the United States, at least 28 states have already enacted laws that require certain online services to implement similar age checking before users can access content [5, 6]. The Supreme Court recently allowed states to impose these requirements for adult online services, and there is ongoing First Amendment litigation challenging state laws that impose these requirements for social media platforms [7]. That case—and many like it challenging other state laws—involves factual questions about the effectiveness and burdens of age checking. Namely: do these systems do what they are supposed to do, effectively categorizing users based on age in real deployments?

This factual question does not, at present, have a clear answer. Despite the expanding scope of online age-checking requirements—affecting tens to hundreds of millions of people—there have been few independent and transparent evaluations of the commercial services used to implement these requirements. Further, many evaluations do not appear to include end-to-end system evaluation in realistic or adversarial settings—and where they do, they do not provide adequate information to assess the sufficiency (or details) of

their methodology, or the reliability or validity of their results. The National Institute of Standards and Technology's (NIST) independent assessments only tested facial analysis algorithms for age checking on non-representative high-quality images that had good lighting, image quality, and pose [8]. Further, NIST's trials of commercial *liveness prediction tests*, which determine if a living human being or a spoof artifact (e.g., a photo, video feed, or mask) is in front of the camera, were conducted with negative cases only: still images and video of people, but no live participants [9].

To address a substantial need for evaluation in realistic settings, we consider the following research questions in this work:

RQ1. How well do biometric age assurance services perform on age prediction, liveness prediction, and joint tasks when target demographics interact with the services *as directed*?

RQ2. How well do these services perform on age prediction, liveness prediction, and joint tasks when target demographics interact with the services *under adversarial conditions*?

To our knowledge, **this is the first paper in the academic literature to evaluate commercial age assurance systems under a realistic threat model.** In particular, our study adversarially tests these systems with participants in the target age bracket (ages 13-17).

This paper is organized as follows: In Section 2, we briefly outline relevant policy and academic work, with an emphasis on age assurance. In Section 3, we describe our participant recruitment strategy, noting specific challenges with recruiting minors, the physical study setup, our study protocol design, and our preregistered analysis plan. We present analyses of our collected data in Section 4, tracking the research questions outlined above. Finally, in Section 5, we consider the implications of our findings for the future of biometric age assurance implementation and age gating-related policy and conclude with a series of policy recommendations.

2 Related work

There are now enacted age assurance laws covering adult content in at least 28 U.S. states [6], alongside broader age assurance regimes in Australia and the U.K., as of August 2026. Requirements are pending or likely in many more jurisdictions. This has spurred significant research interest across a variety of domains. Of most salience are works that attempt to assess the viability of different approaches to age-assurance and their impact as policies are implemented.

Measurements of deployed age assurance systems. Rescorla, ET AL. [10] provide an overview of available age assurance technologies, barriers to implementation, and trade-offs inherent to each choice of technology. Minocha, ET AL. [11] measured the deployment of age-assurance technologies in Georgia and Texas (U.S. states early to enact requirements into law) and identified the presence of age assurance providers on only around 14% of adult sites in their measured set, and that YOTI dominated (60%+) observed deployments. However, this landscape has changed rapidly since their work was conducted in mid-2025.

As of writing, it has been eight months since the rollout of the social-media policy in Australia. The regulator designated to oversee the policy has made meaningful attempts to conduct regular evaluations of the policy-as-implemented, the results of which are made publicly accessible [12]. These span checking the extent to which and how platforms are meeting their compliance obligations, as well as the extent to which the policy is changing social media use among young people and its impact on families [13]. The behavioral findings across the Australian studies are largely consistent: predominantly, young people are still accessing platforms that laws were designed to restrict. Measures and methodologies vary across studies. Platforms reported removing, deactivating, or restricting some 4.7 million under-16 accounts shortly after implementation [13]. Yet independent work found that more than 85% of under-16 participants were still using restricted platforms three months later, with the other major study of this topic reaching a broadly consistent conclusion despite using different measures [14, 15]. This reinforces the fact that a policy that succeeds at reducing account ownership does not necessarily produce a comparable reduction in actual access.

Circumvention. There are also notable results on circumvention of age assurance systems, from both regulator-sponsored research and academic researchers. While news media reports of circumvention focused on weaknesses of the age-assurance technology itself (for example, using video game characters to bypass biometric age assurance, early work suggests that much circumvention comes from more mundane activities such as siblings, parents, or friends helping younger people bypass schemes [12, 16, 17]. This is consistent with qualitative interview work from Nansen, ET AL. [18], which found that, while young people do report attempting to bypass biometrics directly, it is frequently through ''borrowing'' other people to present to the camera.

Alongside this, a variety of studies on the U.K. experience suggest that VPN use meaningfully increased following the introduction of age-assurance requirements, potentially in an attempt to avoid privacy risks and frustrations with content gating. U.K. regulator Ofcom observed a sharp increase in mobile VPN use after the Online Safety Act's age-assurance requirements came into force [19]; independent work found an 89% increase in U.K. Google search interest for VPNs at the July 2025 age assurance deadline, alongside increases in VPN-related discussion around other regulatory milestones [20]. Notably, this work suggests that some of this interest was driven not simply by a desire to access restricted content, but by concerns about privacy and a desire to avoid age-assurance

systems. These concerns about the friction introduced by age assurance are also borne out in a lab setting by Lin, ET AL. [21], who found that biometric and ID-based systems substantially lowered user acceptance (measured through completion of the gating process) to 52% and 27% respectively.

Evaluations of age assurance and facial recognition models. Methodologically, this paper is most directly comparable to work by the National Institute for Standards and Technology (NIST) who have conducted the longest-running evaluation of facial recognition and facial age estimation technologies [8]. Our work is also conducted partially in response to Australia's Age Assurance Technology Trial (AATT), an official government trial conducted in collaboration with an industry body, the Age Check Certification Scheme (ACCS). As a part of the AATT's in-person assessments of the efficacy of existing commercial vendors, the Australian government worked with vendors to deploy and test age estimation algorithms in Australian middle and high schools [2]. Notably, the AATT, while involving an ostensibly independent academic as an advisor, did not consult with security and privacy experts. This resulted in standalone published conclusions, as opposed to the best practice of disclosing clear methodology and results. The closest academic works have generally evaluated synthetic systems or have performed evaluations with synthetic data sets [22–24].

Adversarial evaluation of facial recognition. Our evaluation also draws on a long history of adversarial research on facial recognition and biometrics, in particular [25–27]. We necessarily omit summaries of large volumes of research on general attacks against biometric models and red-teaming [27–41], and instead focus on those components most relevant to the policy issue at hand.

Policy considerations. A fulsome discussion of the merits (or lack of) of the policies—including the empirical work that speaks to them—is beyond scope, other than to acknowledge that proponents of the scheme reasonably note that it is still early days and urge continuing evaluations of the schemes across levels (behavioral change, well-being impacts, technical impacts, and an attempt to measure substitution effects and causality to the extent possible). A variety of such longitudinal studies are reported to be in progress, along with a number of policy-focused whitepapers [10, 42–45]. In all of these discussions, it is critical to recall that an underlying goal of these policies is to improve youth well-being.

3 Methods

In this section, we describe our approach to participant recruitment, our physical study setup, and our study protocol for in-person study sessions. We also present the goals and overarching structure of our preregistered analysis plan.

Table 1: Demographic distribution of study participants. Age bins are left-inclusive and right-exclusive. Participants could report more than one race.

Age		**Race**	
Age range	*N*	Race	*N*
13–18	14	Asian	46
18–20	24	White (Non-Hispanic)	38
20–24	22	Two or more races	15
24–28	14	Hispanic or Latino	10
28–32	8	Black or African-American	6
32–44	5	American Indian	3
44–64	7	Other	3

3.1 Participants and recruiting

As we describe in Section 1, our decision to complete testing with human subjects (and minors, in particular) was motivated by the absence of similarly rigorous in-person adversarial testing with minors in regimes where online age gating policies that specifically target adolescent users have already taken effect [2, 8]. We further discuss the risk mitigations we built into our study protocol in Section 7.

There are no preexisting datasets suitable for evaluating this form of adversarial robustness, and relying on generative AI to modify images would raise questions about external validity. Although several datasets contain face and age data (e.g., those described in [46]) and include images in both controlled and uncontrolled settings, none is appropriate for our use case: adversarial evaluation using images intentionally designed to mislead age-assurance systems about age. Further, many of these datasets pose thorny ethical issues for researchers, since many of them were assembled from images scraped from the public Internet. Accordingly, we designed a study that involved minor study participants, for whom we minimized risks and required both parental permission and minor assent.

Our recruitment strategy was motivated by a desire to engage a diverse cross-section of the Pittsburgh area. We posted calls for participation on bulletin boards on the Carnegie Mellon University and University of Pittsburgh campuses, as well as in local cafés, libraries, parks, restaurants, and gyms. We additionally disseminated information about the study through community organizations, bus ads, and the Center for Adolescent and Young Adult Health (CAYAH), an affiliate clinic of the University of Pittsburgh Medical Center (UPMC). Additional details about our recruiting methods, including flyers, emails, and other documentation, are available in our online supplement [47]. We recruited participants aged 13 years and older residing in the greater Pittsburgh area. Participant age was the only explicit eligibility criterion; the study site was the CMU campus. In all, we recruited $n = 94$ participants, ranging from 13 to 62 years of age. Full race and age breakdowns for the participant pool are available in Table 1. Addition-

Figure 1: Dual-phone rig for in-person study sessions.

ally, 54% of participants self-identified as female, 40% of participants self-identified as female, and 6% of participants self-identified as transgender or non-binary.

3.2 Experimental setup

Our primary desiderata in designing the software and hardware setups for this work were 1) mitigating any risk that minors might learn new strategies for circumventing facial biometric age assurance as a result of completing the study protocol, and 2) securing participants' privacy through deletion agreements and secure on-device data storage and processing for all commercial vendors under evaluation. The age assurance vendors and apps selected for this study and the physical hardware setup for study sessions were motivated by these desiderata.

Software. In order to define a set of prospective apps for evaluation, we consulted the membership list of the Age Verification Providers Association (AVPA) [48]. This list comprises thirty-four commercial age assurance vendors as of this writing. We reviewed the age assurance solutions provided by each vendor and attempted to identify 1) whether or not these solutions are on-device; 2) whether or not the provider explicitly guarantees deletion of data collected during the assurance process; and 3) whether or not the provider made a sandbox, developer account, or free API checks publicly available. If any of the above was true, we contacted the service provider to inform them that we were interested in accessing their age assurance service for a research study. We ultimately received positive responses from three age assurance providers: Yoti, Facetec, and Privately. Cross-checking these providers against reverse-engineered APKs and platform policy documentation confirmed that, of the social-media platforms restricted in Australia, 7 of 10 used one of our tested providers (see Appendix O). Not every service was available on both operating systems. We accessed Yoti's on-browser age assurance service on iPhone and Android. We installed Facetec's demo app on both phones. Privately's AgeAI for Retail app is only available for Android phones.

- **Yoti.** We were able to negotiate access to a specialized testbed with Yoti. Per their whitepaper, Yoti guarantees deletion of user data once assurance is complete [49]. Yoti is the age assurance provider for Claude, the AI assistant deployed by Anthropic, and Meta's family of apps (WhatsApp, Instagram, Facebook) [50, 51].
- **Facetec.** The Facetec Demo app is freely available on the Apple App Store and the Google Play Store; Facetec agreed to delete all data associated with our study data collection on written request. Facetec is a popular liveness detection provider for other age assurance services (e.g., Yoti's liveness detection functionality is provided by Facetec [52]). As of this writing, four state ID programs (Colorado, Virginia, New York, Utah) use Facetec for biometric liveness detection [53, 54].
- **Privately.** Privately's on-device age assurance app (called AgeAI for Retail) is freely available on the Google Play Store and can be used without a network connection; all image processing occurs on-device. Privately provides biometric age assurance services for Discord via a partnership with k-ID [55, 56].

Hardware. We used two mobile devices for testing: an Apple iPhone 15 (iOS version 26.3) and a Google Pixel 8a (Android version 16). We chose these devices because they are mainstream models, and are representative of iOS and Android ecosystems.

In order to maintain consistent image framing, angle, and lighting across devices, we designed a custom rig to simultaneously hold both mobile devices (see Figure 1). We designed and 3D-printed two custom arms, which we coupled with out-of-the-box phone holders. We mounted this rig on a camera tripod; we were able to position both arms such that two phones, when placed in their phone holders and mounted on the tripod, could take photos of a seated study participant from approximately the same height and angle.

To mitigate the risk that minor participants might learn strategies for circumventing age assurance systems, we prevented the study participants from seeing any content on the phone's screens by taping opaque covers over both phone screens. Each phone was connected to an external monitor and mouse that the researchers could use to conduct the study session; these monitors were visible only to research team. This design allowed us to complete testing with minor participants without revealing to them that the systems under evaluation specifically evaluated age (minor participants were informed that they were testing ''facial analysis AI''), and without requiring that they directly interact with the services

themselves. During normal use, both YOTI and FACETEC display onscreen instructions to users that require them to move their heads in certain ways in relation to the phone camera; to facilitate these adjustments for participants in the absence of an onscreen visual, research team members manipulated the phone rig or verbally instructed study participants to make adjustments to their head position.

3.3 Study protocol

We provide an overview of our study protocol from the point of participant consent or assent to compensation. The full study protocol, with all artifacts and documents made available to participants during in-person study sessions, is included in our online supplement [47].

Consent, assent, and permission. All adult participants received full information about the three services being tested, data collection and storage, and the policy and technology context for AI-driven biometric age assurance. Adult participants were then asked to provide their informed consent to participate prior to the start of their in-person study session.

We struck a balance of disclosing truthful information about the systems we were studying, but we only disclosed these details to parents/guardians who could make a decision about whether or not to share it with their children. This avoided any affirmative deception and gave parents the decision-making authority about the risks of evasion versus disclosure of additional information. Minor participants were informed that they were participating in an audit of facial analysis algorithms, and were not told that these algorithms specifically produced age estimates from facial images, to mitigate the risk of future age assurance circumvention and potential harm resulting from this study. Parents/guardians of minor participants received full information about the age assurance services being tested and possible risks of participation and our mitigation measures; they were informed that, while the research team would not disclose the names of age assurance vendors or age assurance outcomes to their child, parents/guardians could voluntarily disclose these details to their child at any point before, during, or after the study session. Additional detail about our recruitment process and permission, consent, and assent forms are available in our online supplement [47].

Evaluation conditions. In order to understand how well each service functions during standard use, we first recorded baseline age estimates for all three services. For baseline age estimates, participants posed with no props or facial coverings and maintained a neutral facial expression. We asked participants who arrived to the study session with glasses and head coverings to remove their glasses and head coverings if they felt comfortable doing so. Before the start of the study session, we informed participants that they could request to skip test conditions or stop testing at any time.

We then asked participants to wear surgical and costume masks, put on costume glasses, and apply adhesive costume mustaches and beards and pose in front of both mobile devices; a full list of study conditions is available in Table 5. As possible, we tried to match costume beard and mustache colors to each participant's hair color (if applicable) on the day of the study session. These test cases were generated from our reading of literature on physical attacks against (facial) biometric models and include the use of specially designed facial adhesives, glasses, and other accessories [27–41]. After preliminary testing with AI-modified facial photos suggested that facial hair might induce age overestimates in younger users, we added more realistic adhesive mustache and beard conditions to the list of possible test cases several months after in-person sessions began in February of 2026 [57].

Controlling the testing environment. Participants who arrived to the testing site with facial hair, eyewear, or headwear were requested to remove their eyewear or headwear if they felt comfortable doing so, and participants with their own facial hair were asked to don our mustache and beard disguises over their own facial hair. Testing was conducted in one of two rooms located in an academic research lab. While we attempted to maintain consistent testing settings *across* participants, we deliberately opted for a physical setting that would not guarantee absolutely ''clean'' image backgrounds: artifacts of the office setting, including furniture and wall decor, are visible in most images. We believe that this image ''noise'' is representative of the conditions under which many users engage with age assurance systems.

Observations. We designed our study protocol to maximize the efficiency of each study session we were able to undertake: participants in the full 90-minute study session were asked to complete baseline measurements and 11 test conditions, each comprising 17 independent observations of age assurance outputs (6 service thresholds tested on two devices for FACETEC; 1 age assurance output on two devices for YOTI; 3 service thresholds tested on one device for PRIVATELY), for a maximum of 204 observations per participant.[1] Some participants opted for an abbreviated, 60-minute study session, which excluded the facial expression conditions (i.e., open-mouth, smile, and frown), for a reduced maximum of 153 observations per participant.

Conclusion and compensation. At the conclusion of the study session, participants were asked to complete an optional exit survey that asked for their real age (treated as ground truth), self-reported race and gender identities, and prior experience using social media and biometric age assurance mechanisms. The full survey instrument is available in our online supplement [47]. Participants were then emailed an electronic gift card code a day or two after the conclusion of their study session: participants in 90-minute sessions received $45 Amazon gift cards, and participants in 60-minute sessions received

[1] We initially attempted to complete three ''trials'' per age check in the event of a null response but found, during pilot testing, that these repeated trials were too time-consuming, particularly during tests with FACETEC's demo app.

$30 Amazon gift cards, a pay rate well above minimum wage in the state of Pennsylvania.

3.4 Analysis plan

We preregistered our analysis plan on the Open Science Framework (OSF) and embargoed the plan pending publication of our full post-collection analysis. The analysis plan was drafted concurrently with pilot tests of the in-person study protocol. During those pilot tests, colleagues and research team members completed all test cases and offered feedback on the protocol; we adjusted our protocol and analysis plan accordingly, then preregistered the analysis plan on April 29, 2026, in advance of data analysis. We did not include the results of our pilot tests in our final data analysis. In light of limited sample size, we omit regression analyses and multi-comparison corrections from our preregistered analysis plan.

3.5 Limitations

The limitations we describe below pertain to the difficulties of recruiting minors and the external validity of our evaluation conditions.

Recruiting minor participants. We leveraged institutional resources from a participating university (e.g., recruiting pools, common areas accessible to students) to contact eligible college-age participants, but found that minor participants were difficult to reach through passive advertisements, such as posted flyers.

Sample size. We tried to obtain a sufficiently large sample for statistically significant analysis, but were constrained by the challenges of recruiting, the burden of carrying out the study, and the resources available: a sample size large enough to support statistically significant observations of medium- to small-sized effects in some of our demographic strata would have numbered in the multiple hundreds, and we did not have the research staff or financial means to undertake a study of that scale. As such, where possible, we report observational statistics over the set of participants we test. We note, however, that this sample size is still larger than those used by regulators and industry for adversarial testing [2].

External validity. Toward understanding how a resourceful preteen or adolescent might try to circumvent age assurance systems, we deliberately chose a slate of physical presentation attacks with low barriers to access: all disguises were available to purchase online, and no single disguise cost more than $20. During evaluation, we asked that participants maintain fixed poses or move their heads closer to or farther from both phone cameras. In practice, a user to whom the phone screen is fully visible would likely be able to engage in adversarial ''play'' with the age assurance app, manipulating their head position and facial expression in real time to induce a desirable outcome. We were unable to exactly emulate this process during our in-person tests. Furthermore, we did not exhaustively explore all combinations of physical presentations (e.g., combining a Groucho disguise with a wig, or combining facial manipulations with facial coverings). We leave such evaluations for future work.

3.6 Terminology and documentation

In this section, we distinguish between two different notions of *thresholds* for age assurance evaluation and describe our approach to reporting—and unifying—results from all three age assurance providers. Additionally, we discuss and motivate our choice of error reporting methods.

Evaluation and service thresholds. We distinguish between *service* and *evaluation* thresholds, defined respectively as t_s and t_e in Table 2. Service thresholds (t_s) are age thresholds pre-set by each age assurance service; services that return a pass/fail (e.g., Privately, Facetec) return these Boolean results with respect to a service threshold t_s. The service threshold usually corresponds to an age with policy import, such as 16, 18, or 21 years of age.

Evaluation thresholds (t_e) are set by the research team in order to understand the performance of age assurance services on age thresholds that might not be available in-app; or to convert non-Boolean results into Boolean results (e.g., in the case of Yoti) to facilitate cross-vendor comparisons.

Normalization of reported results. All three age assurance services produce age-related predictions, though the reporting mechanism for each service is unique: Facetec returns a Boolean pass-fail result in response to a threshold that the user chooses (e.g., user selects the 21 or older threshold; Facetec returns a Pass/Fail result); Yoti returns a four-year integer age interval estimate (e.g., Yoti might estimate that a user's age lies in the interval 21-25, inclusive); and Privately recommends that the user provide further ID proof (or not), in response to three pre-set age thresholds ($\geq$ 21, $\geq$ 25, and $\geq$ 30; for instance, a minor user might choose the $\geq$ 30 threshold. In response, the app might recommend that the user provide a government-issued ID). Furthermore, Facetec and Yoti also provide *null* responses, or age prediction results that are neither pass nor fail.

For ease of comparison with Facetec and Privately, we convert Yoti's four-year age interval into a Boolean response: we identify the midpoint p of the reported interval. (As an example, for a 21-25 interval, p would be 23 years of age.) We then apply a buffer value b—a fixed offset—to this midpoint estimate.[2] We then compute $d = [p \geq t_s + b]$, where t_s is the service threshold as defined in Table 2, and where d is the corresponding Boolean result. We record d and compare this Boolean outcome with the results returned by Privately and Facetec for that participant at service threshold t_s.

[2]We consider common buffer values reported in Yoti's own whitepapers and by regulatory agencies worldwide; for our analysis, we use buffer $b = \{2, 3, 5, 7\}$. These are values used by [49, 58, 59], respectively.

Appendix B describes each service's responses and our recording methodology. We note that false positive rates are likely deceptively low in our reported results; this is due to the fact that, for certain age bands, the number of true negative outcomes was proportionally lower (e.g., for age band 13-17; we had no participants under 13 years of age).

Error reporting. For services reporting continuous age estimates or intervals (i.e., Yoti), we report MAE and error quartiles. We define our MAE in Table 2 as the sum of the differences between users' real ages a and point estimates p divided by the size of the full participant pool. A primary motivation for this choice of metric (MAE) is comparison with existing evaluations, such as those completed and published by NIST [8]. We note, however, that MAE does not fully reflect the magnitude and direction of age assurance errors: that is, age overestimates and underestimates are treated identically under this error definition and might even cancel one another out. For this reason, we also report error quartiles in Appendix G.

Demographic reporting. We stratify our dataset along age group, race, and gender axes. These dimensions are self-reported by study participants in an optional exit survey administered at the end of each study session. For participants reporting multiple racial identities, we include that participant in each constituent race category they report and additionally include them in a ''Multi-racial/One or more races'' category. (e.g., a participant reporting themselves as ''White'' and ''Middle Eastern'' would be counted separately in aggregate analyses of White-identifying participants and Middle Eastern-identifying participants, and would additionally be counted in all aggregate analyses of the demographic subgroup of multi-racial-identifying participants.) We refrain from performing analyses over unique combinations of multi-racial subcategories (e.g., ''White/Middle Eastern'') because these combined categories are too numerous to exhaustively analyze and sample sizes are too small to yield stable FPR/FNR results. We additionally label participant images using the Monk skin tone scale (these images were deleted at the conclusion of this analysis) to understand how complexion might affect age assurance and liveness prediction performance [60]. The Monk skin tone scale applies 1-10 labels, where 1 indicates lightest skin tone and 10 indicates darkest skin tone. We elect to use Monk scale labels because this scaling system represents the current state-of-the-art in fairness-related AI evaluations.

Joint liveness and age prediction. We combine age assurance and liveness detection to simulate whether a person would be allowed to access an online service (d). In the typical configuration of online age assurance services, a user must both appear to meet an age threshold and pass liveness detection (when implemented[3]). We define $d = e \wedge l$, where e is a Boolean age prediction result (with $e = 1$ denoting a ''pass'' and $e = 1$ denoting a ''fail'') and l is a Boolean liveness prediction result. Then a false negative results when a user is of age and receives joint decision $d = 0$; a false positive results when a user is underage and receives a joint decision $d = 1$. When treating null outputs as equivalent to liveness failures, we define $d = e \wedge l \wedge \neg n$.

[3]We note that, in separate conversations with Privately engineers, we were informed that Privately does, in fact, have an in-house liveness prediction model. This model is likely only selectively implemented, and was not obviously available in the AgeAI for Retail app used for our testing.

Table 2: Terms and definitions.

Term	Definition
Age (a)	Participant's ground truth age in years, as self-reported in our exit survey.
Age point estimate (p)	Midpoint of the 4-year (inclusive) age interval returned by Yoti's age estimation service.
Buffer (b)	A buffer that we use to convert point estimates of age (from Yoti) into binary predictions.
True positive (TP)	The participant's age a is greater than or equal to the evaluation threshold t_e, and a service outputs a binary age estimate e with value 1 when using the service threshold t_s.
True negative (TN)	The participant's age a is less than the evaluation threshold t_e, and a service outputs a binary age estimate e with value 0 when using the service threshold t_s.
False positive (FP)	The participant's age a is less than the evaluation threshold t_e, and a service outputs a binary age estimate e with value 1 when using the service threshold t_s.
False negative (FN)	The participant's age a is greater than or equal to the evaluation threshold t_e, and a service outputs a binary prediction with value 0 when using the service threshold t_s.
Service threshold (t_s)	An age classification threshold supported by a service, which we configure the service to use in an evaluation.
Evaluation threshold (t_e)	An age classification threshold that we fix in an evaluation, for purposes of determining positive and negative classes. The evaluation threshold may be a service threshold, or it may differ when constructing performance curves.
Mean absolute error (MAE)	For Yoti only, because it provides an age interval estimate, we will report MAE defined as: MAE = $(\Sigma_{participants}\lvert p - a\rvert/(\text{count of participants}))$.

Interpretations of liveness ground truth. There is some ambiguity regarding the correct interpretation of liveness ground truth for some of our test conditions: while ground truth liveness for a participant wearing a full-face mask is arguably a failing (''not live'') score, it is less clear if the ground truth liveness prediction outcome for a participant wearing a KN95 mask and other partial facial occlusions (e.g., an eyepatch) should be *live* or *not live*. One might argue that

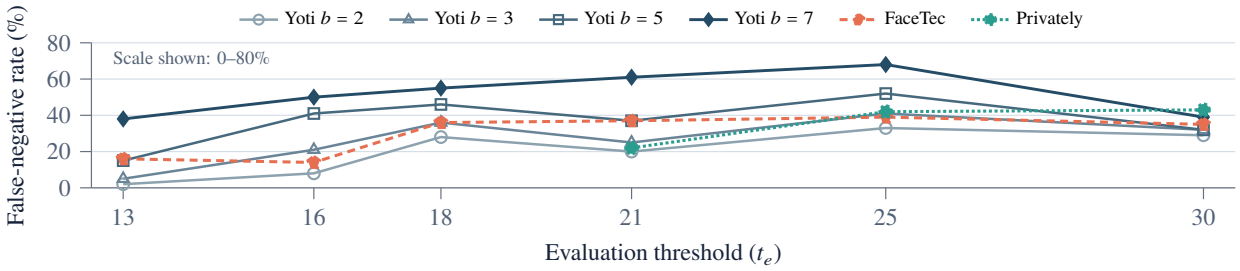

Figure 2: Baseline false-negative rates across matched thresholds ($t_e = t_s$), aggregated across devices. Yoti curves vary buffer b; Privately is available only at $t_e \in \{21, 25, 30\}$.

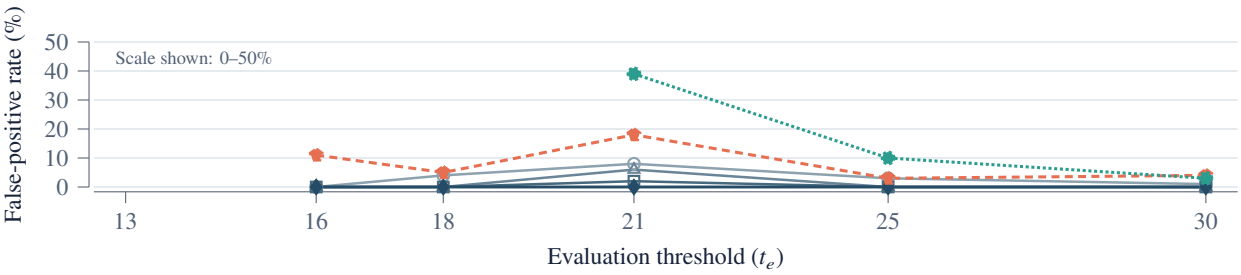

Figure 3: Baseline false-positive rates across matched thresholds ($t_e = t_s$), aggregated across devices. Series encodings match Figure 2; values are unavailable at $t_e = 13$, and Privately is available only at $t_e \in \{21, 25, 30\}$.

a surgical (or other medical) mask is a routine presentation for medical professionals, or in the midst of a pandemic, and that a user wearing such a mask should be correctly identified as *live*; similarly, one might make the argument that partial facial coverings might be understood as spoofing attempts and should be correctly characterized as *not live* presentations. Age assurance providers, in their own documentation, specify that liveness detection is intended to confirm the presence of a living person in-frame.[4]

4 Results

We organize our presentation of results by research question: we present performance results under baseline conditions (Section 4.1) before proceeding to a discussion of performance under adversarial testing conditions (Section 4.2). We preface our results presentation with definitions of terms (see Table 2) used throughout our results, analysis, and discussion sections. We summarize high-level findings below:

- **Age prediction (without liveness).** We find that a full-face costume mask effectively induces age overestimates for users. The KN95, Groucho, and glasses evaluation conditions induce high FNRs among older users.
- **Liveness prediction.** Darker skin tone induces outstandingly high null response rates and FNRs on liveness prediction tasks. Additionally, liveness prediction performance appears to track with the degree of facial coverage and the visibility of certain facial landmarks, such as the eyes and mouth.
- **Joint age and liveness prediction.** A selection of facial manipulation and partial facial occlusion conditions induce elevated FPRs in underage users, suggesting that subtle modifications to a user's facial expression can successfully bypass joint age and liveness checks. Additionally, a much larger number of evaluation conditions induce high FNRs for older users, indicating that partial and full facial coverings—including KN95 masks, costume masks, and eyepatches—might induce erroneous blocks of access.
- **Demographic attributes.** Race and gender appear to induce differential effects in age and liveness performance, with Black- and African-American-identifying users often receiving the highest FPR or FNR scores and white-identifying users often receiving the lowest. On stratifying participants by age bucket, we noticed the highest FPR and FNR near service and evaluation thresholds, with some services ''aging up'' younger users and ''de-

[4] In a whitepaper from January 2026, Yoti defines the objective of liveness detection as ''ensur[ing] that an identity check is being performed by a real, physically present human being, rather than a static image, recorded video, or synthetic representation'' [61]. FaceTec defines a liveness check as ''the ability for biometric systems to determine if the data it has collected was from a live human or an inanimate or non-living object, like an Artifact in a Spoof attempt'' [62].

aging'' older users: that is, younger users induced higher FPR scores at boundaries, and older users induced higher FNR scores at boundaries.

4.1 Performance under baseline conditions

We first present *baseline* performance results for all three services. As we describe in Section 3, we define the ''baseline'' test case as one in which study participants maintain a neutral facial expression and attempt to comply with each age assurance service's usage instructions.

4.1.1 Age predictive performance (baseline).

We discuss FPR and FNR on age prediction tasks for all three services and MAE for Yoti, which is the only service that returns a predicted age interval rather than a Boolean outcome. Facetec returns outstandingly high FNRs (i.e., instances in which users were incorrectly characterized as underage with respect to an service threshold t_s) for older users at threshold values: for instance, 0.68 for users aged 17-20 (n = 41) at the 18-year threshold, and 0.62 for users aged 21-24 (n = 26) at the 21-year threshold. We note a similar tendency in Privately, which returns FNR of 0.41 for users aged 21-24 (n= 17) at the 21-year threshold and FNR of 0.67 for users aged 25-28 (n = 15) at the 25-year threshold. FPRs are less pronounced but still notable for both services: FPR of 0.26 for users aged 17-20 (n = 43) at the 21-year threshold for Facetec and 0.45 for users aged 17-20 (n = 31) at the 21-year threshold for Privately. Yoti returns FPR of 0.36 for users aged 13-16 (n = 26) at the 16-year threshold, and FPR of 0.35 for users aged 13-16 (n = 26) at the 18-year threshold. We observe that, in general, Yoti returned lower false negative rates than both Privately and Facetec for buffer value $b = 2$, and lower FNR than Privately for most thresholds when $b \in \{2, 3\}$.

Even so, we observe MAE of more than two years of age for Yoti's real-valued age predictions, even at baseline (here, we take point estimate p as defined in Table 2)—a significant margin of error, in light of the precision required to gate out users at well-defined age boundaries (e.g., 16 years of age [58]). These results roughly align with NIST's findings on a similar task, for users aged 18-30. Table 3 compares NIST outcomes to MAEs for participants in the same age band in our study.[5] Full FNR/FPR tables for Privately and Facetec are available in Appendix N and Appendix M, respectively, and baseline age prediction performance results for all three services are available in Figure 2.

4.1.2 Liveness predictive performance (baseline).

Two of the three services we tested—Yoti and Facetec—provide explicit liveness prediction outcomes separately from age prediction outcomes. Yoti displays a liveness pass/fail concurrently with an age estimate, which is reported as a real-valued four-year interval. Facetec completes liveness prediction in advance of producing a pass/fail age prediction result, or displaying a null outcome.

We define a true positive result as one in which a study participant is correctly identified as ''live''; we note that there are no true negative outcomes among our baseline test cases (our test cases involved only living human participants). FNRs on liveness prediction tasks are notably high: FNRs at baseline for Yoti and Facetec are 0.32 and 0.49, respectively. Facetec returns liveness pass rates of 0.48 and 0.55 for Android and iPhone, respectively: that is, for at least half of all participants presenting themselves to the Facetec liveness check under standard usage conditions, the application returned a ''null'' or ''Did not pass'' result. (A null response implies a failed liveness detection test, per the demo app's UI. We further discuss usability issues that we observed during Facetec tests in Section 5.) Yoti reports liveness pass percentages of 0.45 and 0.90 for Android and iPhone, respectively. While we have not investigated the hardware requirements of each application, we hypothesize that the Yoti liveness pass percentage discrepancy that we observe might be attributable to differences in available camera hardware on the iPhone and Android devices.

We note that null responses for both services are distinct from age prediction failures: Yoti displays a ''Face not found'' or ''Multiple faces detected'' message in lieu of an age estimate and liveness pass/fail score, and Facetec displays a ''Let's try that again'' message.[6] At baseline, we find that Yoti produced no null responses, and Facetec produced null responses for 0.32 of participants.

4.1.3 Joint age and liveness prediction performance (baseline).

For Yoti, at evaluation threshold $t_e = 13$ at baseline, we observe FNR equal to 0.62 for users aged 13-16. That is, more than half of all study participants aged 13 to 16 (inclusive, with n = 26) were incorrectly determined to be under 13 years of age

Table 3: Comparison of Yoti mean absolute error (MAE) by gender for participants aged 18–30.

Source	Model	Age Band	Gender	MAE
NIST	yoti-004	18–30	Female	2.60
NIST	yoti-004	18–30	Male	1.00
Our work	Yoti age antispoofing	18–30	Female	2.42
Our work	Yoti age antispoofing	18–30	Male	1.25

[5] Notably, this NIST report does not report FNRs outright, and instead reports a composite ''inconvenience'' score [8].

[6] This message is occasionally accompanied by some explanatory text, e.g., ''Neutral expression, no smiling,'' or ''Please use ideal lighting.'' Screenshots are available in our online supplement [47].

when age verifying with Yoti. FNR for users aged 17-20 (n = 62) was 0.79 for $t_e = 16$, consistent with a tendency toward age *underestimates* for users above a given age threshold for this service provider. This pattern maintains among users in the 21-24 age group (n = 34) at $t_e = 21$: the FNR for those users is 0.79. We note that Facetec does not advance to the point of providing an age prediction if a user does pass its liveness test. As such, the FNRs we report on its joint prediction tasks are identical to its liveness FNRs at baseline: outstandingly high null response rates translate to FNR of 1.0 for most age bands on $t_s = \{13, 16\}$, and FNR above 0.60 for almost all age bands at $t_s = 18$.

4.1.4 Demographics and other factors impacting performance (baseline).

MAE for participants self-identifying as *white* was 1.95 years on Yoti's baseline age prediction task. This figure was 1.97 for participants self-identifying as *Asian* (all-inclusive of East, South, and Southeast Asian-identifying participants), 2.3 for participants self-identifying as *Hispanic or Latino*, and 2.27 for participants self-identifying as *multi-racial*. These results appear to be consistent with prior reports of Yoti's accuracy, which place its MAE on age assurance tasks between 2.0 and 2.7 years [8].

Additionally, we find that liveness pass percentages for participants with a Monk skin tone score of 8 (the highest scaled score we observed in our dataset) were 0.33 and 0.17, respectively, for Yoti and Facetec: these were the lowest pass percentages across all Monk score buckets for each service. 0.30 of participants with Monk score 5 passed Facetec's liveness detection test, and 0.22 of participants with Monk score 6 passed. These percentages are not monotonically decreasing with darker participant skin tone, however: we note that 0.39 of participants with Monk score 7 passed Facetec's liveness detection. Full liveness performance results with respect to participant skin tone are available in Appendix AA and Appendix AB.

Approximately 0.72 of participants self-identifying as *male* passed Yoti's liveness prediction task, while approximately 0.65 of participants self-identifying as *female* passed the liveness prediction task. By comparison, approximately 0.65 of participants self-identifying as *male* passed Facetec's liveness prediction task, while approximately 0.47 of participants self-identifying as *female* passed the same liveness prediction task. (We include a discussion of Facetec's usability issues—which might have contributed to these comparatively low pass percentages at baseline—in Section 5.) These are not statistically significant differences ($p > 0.05$ on a two-proportion z-test), in light of limited test population size. As a reminder, the ground truth liveness outcome at baseline is a ''pass.''

We additionally report age prediction FPR and FNR by racial subgroup in Appendices U and W for Facetec and Privately, which output Boolean pass/fail results over pre-specified thresholds. Results are mixed: Privately FPR was 0.10 for white-identifying users; FPR was 0.15 and 0 for Asian- and Black-identifying users, respectively. Yoti FPR was 0.11 for participants self-identifying as white; FPR was 0.17 and 0.11 for Black- and Asian-identifying participants, respectively. While these results suggest elevated FPR for Black-identifying participants on some conditions, these differences do not attain statistical significance ($p > 0.05$ on a χ^2 test).

Similarly, white-identifying participants had FNR of 0.052, 0.54, and 0.56 with Yoti, Privately, and Facetec, respectively; Black-identifying participants had FNR of 0, 0.56, and 0.55; and Asian-identifying participants had FNR of 0.047, 0.65, and 0.62. Though Asian-identifying participants had elevated FNRs on a number of conditions, these results are not statistically significant.

Taken together, these results suggest that race, skin tone, and gender have observable impacts on age and liveness prediction outcomes, and more extensive testing might prove out the statistical significance of these observations. Darker skin tone appears to induce higher rates of liveness failure, and male users consistently induce lower error rates (either FNR or FPR) than female users. The directionality of these changes (i.e., whether membership in a racial or gender subgroup induces false positive or false negative results) varies across services. This suggests that intensive evaluations of each service provider are crucial ahead of widespread implementation, since one service provider's performance outcomes and biases likely do not transfer to other service providers.

4.2 Performance under adversarial conditions

We present performance results for all three services on our adversarial test cases, which require participants to wear masks, wigs, and other head and face coverings while completing age and liveness checks. The full list of test conditions is available in Appendix A.

4.2.1 Age predictive performance (adversarial).

As in Section 4.1.1, we present FPR and FNR results for all services and additionally discuss Yoti MAE.

Certain test conditions, including the KN95 mask, costume mask, and Groucho disguise, appear to cause participants with ages close to the threshold to be estimated as younger or older than they really are. In particular, FPR is largest among participants below age thresholds, and FNR is largest in age groups at or just above age thresholds. We discuss these errors in greater detail in Section 4.2.4.

We note that the full face costume mask, which covers the participant's entire head (with cutouts for both eyes), induces age estimates in the 60s and 70s for *all* participants during testing with Yoti, irrespective of their ground truth age. In fact, the false positive rate (i.e., age overestimate) for the costume

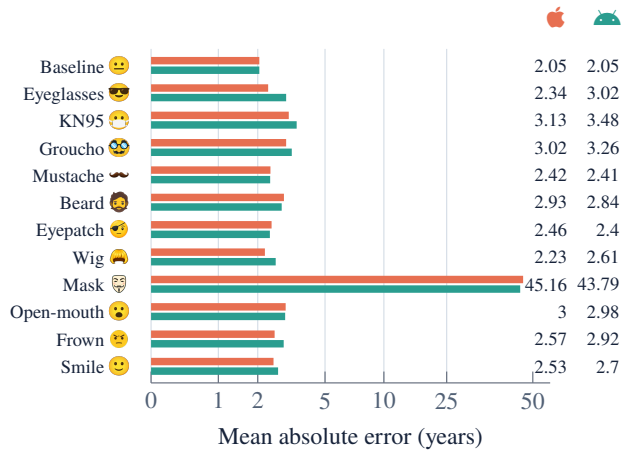

Figure 4: Mean average error (MAE) with respect to participant-reported ground truth ages (a) for Yoti only, per device type. Error is the absolute difference between a and midpoint estimate p as defined in Table 2.

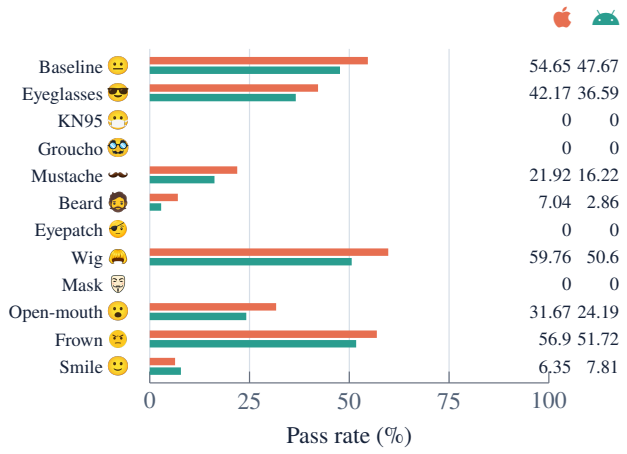

Figure 5: Liveness pass percentage relative to device-specific baseline/control for Facetec only. The baseline pass percentage for an iPhone was 54.65%; the baseline pass percentage for an Android was 47.67%.

mask condition was 1.00. Additionally, Privately returned ''pass'' age prediction outcomes for all users using the $t_s = 30$ threshold during testing with the costume mask condition, irrespective of their ground truth age. For both services, the performance differentials between baseline and the costume mask condition are statistically significant ($p < 0.05$ on a χ^2 test). We observe the smallest deviations from baseline performance for test conditions that only require the participant to manipulate their facial expression, or that minimally occlude the face (i.e., the wig condition). Yoti MAE for female participants with darker skin tone (Monk score of 8) exceeded MAE for female participants with lighter skin tone (Monk score of 5) on all adversarial conditions, sometimes by a margin of more than 3 years (see Appendix Z).

4.2.2 Liveness predictive performance (adversarial).

We consider the impact of adversarial conditions on liveness performance overall. (For a description of liveness outputs from all three services, see Table 2.) Absolute accuracy rates on Facetec's per-task performance for each test condition are available in Figure 5, and Yoti's performance graph is available in Appendix F. We observe that Facetec's liveness performance appears to correlate with degree of facial coverage and the obstruction of certain facial landmarks, such as the eyes and mouth: while facial manipulations resulted in lower pass percentages over baseline (with the ''smile'' condition performing particularly poorly—we note that Facetec's UI specifically requests that users ''not smile'') and hairpiece conditions (wig, mustache, beard) resulted in comparable or reduced pass percentages, disguises that covered the mouth or eyes in part or in full (KN95 masks, eyepatches, full face masks, Groucho Marx glasses-and-nose sets) resulted in 0.0 pass rates. Per Facetec's own documentation, the company's Liveness Detection functionality ''detect[s] concurrent human traits such as skin texture, reflections in the eyes, eye focus, pupil dilation, and many more'' [62]. Yoti's liveness detection (see Appendix F) appeared to identify most partial facial coverings as liveness failures. The pass rates for those test conditions—which include the KN95 mask, wig, eyepatch, beard and mustache—were 0.01 or less.

To address the ambiguity regarding interpretations of liveness ground truth (discussed in Section 3.6), we include both ground truth characterizations of our partial facial covering conditions (including the KN95 mask, eyepatch, and wig disguises) in our analysis of Yoti and Facetec's liveness detection performance. In this instance, this only requires that we take the complement of our FNR outcomes when comparing both definitions of ground truth side-by-side: whereas the FNR for the KN95 mask is 1.00 for Facetec when ground truth liveness is a passing score, FPR is 0.0 when ground truth is a failing score. In fact, both Yoti and Facetec returned FNRs of nearly 1.00 on the KN95 and eyepatch conditions with ground truth liveness set to a passing score; Facetec returned FNR of 0.41 on the wig condition for ground truth set to a passing score, implying a FPR of 0.59 in the case that ground truth is a failing score.

4.2.3 Joint age and liveness prediction performance (adversarial).

We find that, when age and liveness prediction tasks are jointly considered (with pass/fail responses determined as reported in Table 2), more null-response-prone liveness outcomes prevail: that is, services (and Facetec, in particular) might ''fallback'' to liveness detection failures to gate out users instead

of age prediction outcomes, which tend to be more sensitive to errors and less binary. Notably, at baseline, Facetec's pass rate across the full dataset is only 0.50 (see Figure 5).

For Yoti, partial facial occlusion conditions (e.g., the KN95 mask, eyepatch, Groucho disguise) induce outstandingly high FNR (1.00 across almost every age band and threshold), indicating that nearly all of-age participants are gated out by the joint age-liveness rule, even under minimal facial coverage (for a further discussion of liveness interpretations, see Section 4.2.2). High rejection rates are especially pronounced for older users: at $t_e = 13$ and $t_e = 16$, for instance, the open-mouth condition return FNRs at or above 0.50 for all age groups, with the highest error rates (0.91 at both thresholds) for participants aged 25-28 (n = 22). KN95, Groucho, mustache, and beard conditions also return FNR of 1.0 for nearly all age groups at every threshold.

As we note in Section 4.2.2, Facetec returns 0.0 pass rates on a number of partial occlusion conditions. These outcomes are reflected in outstandingly high FNRs across partial occlusion conditions for joint prediction tasks: Facetec rejects almost all users at every threshold for the smile, squint, mustache, KN95, eyepatch, and Groucho conditions. Both Facetec and Yoti reject almost all users on the mustache, beard, KN95, Groucho, costume mask, eyepatch, and wig conditions. These outcomes suggest that of-age users are likely to be subject to erroneous blocks, even with no or modest adversarial adjustments to their physical appearance. Full joint liveness and age prediction results for Yoti and Facetec are available in Appendix X and Appendix Y, respectively.

4.2.4 Demographics and other factors impacting performance (adversarial).

Across all three services, errors concentrate near decision boundaries, with younger participants more susceptible to false positives and older participants more susceptible to false negatives. Table 4 summarizes the strongest Yoti effects: at $t_e = 16$, FPR is elevated among 13-16-year-olds, reaching 0.50 under KN95 and eyepatch conditions. At higher thresholds, FNR becomes more prominent immediately above the threshold; for example, at $t_e = 21$, FNR exceeds 0.50 for KN95 and Groucho among participants aged 21-24. Facetec exhibits a particularly strong tendency toward false negatives near threshold ages, while Privately shows both high FPR below the threshold and high FNR above the threshold under several disguise conditions (Table 4).

Finally, for Yoti, the costume mask condition (a full-coverage face mask that mimics the appearance of an elderly man) induced age overestimates in *all* participants.

We additionally hypothesized that services with liveness detection functions might infer user gender from facial characteristics; users with female-presenting features might be more likely to receive failing scores on facial hair test conditions than users with male-presenting features. We did not notice outstandingly high performance differentials between male and female-identifying participants on the beard condition (FNR on this condition was 0.96 and 0.86 for female and male users, respectively, for Yoti, and 0.97 and 0.88 for Facetec)—as we note in the latter half of this section, we observe similar performance deltas between genders for *all* testing conditions, for both Yoti and Facetec. Interestingly, we observe much higher performance differentials between male- and female-identifying participants on the mustache condition: there, the FNR was 0.88 and 0.63 for female and male participants for Facetec, and 0.86 and 0.61 for Yoti.

Table 4: Selected age-estimation error rates near decision thresholds for Yoti, Facetec, and Privately. FPR and FNR are reported as proportions.

Service	Threshold	Age Group	Condition	FPR	FNR
Yoti	16	13–16	Baseline	0.36	–
		13–16	KN95	0.50	–
		13–16	Groucho	0.46	–
		13–16	Glasses	0.46	–
		13–16	Eyepatch	0.50	–
		13–16	Open-mouth	0.64	–
		13–16	Frown	0.60	–
	18	13–16	Glasses	0.24	–
		13–16	Groucho	0.24	–
		13–16	Eyepatch	0.25	–
		17–20	Groucho	–	0.30
		17–20	Open-mouth	–	0.32
		17–20	Mustache	–	0.31
		17–20	Glasses	–	0.08
		17–20	Eyepatch	–	0.15
	21	21–24	KN95	–	0.58
		21–24	Groucho	–	0.52
	25	25–28	KN95	–	0.53
		25–28	Groucho	–	0.60
Facetec	18	17–20	Baseline	–	0.68
		17–20	Open-mouth	–	0.50
		17–20	Glasses	–	0.51
		17–20	Frown	–	0.59
		17–20	Wig	–	0.73
	21	21–24	Open-mouth	–	0.77
		21–24	Frown	–	0.67
		21–24	Glasses	–	0.21
	25	25–28	Glasses	–	0.76
Privately	21	17–20	Baseline	0.45	–
		21–24	Baseline	–	0.41
		25–28	Baseline	–	0.20
		17–20	Frown	0.45	–
		13–16	Beard	0.85	–
		17–20	Beard	0.70	–
		17–20	Groucho	0.62	–
		21–24	KN95	–	0.71
		25–28	KN95	–	0.60
		21–24	Wig	–	0.65
		25–28	Wig	–	0.57

With the exception of the mustache, smile, and eyepatch conditions, MAE for Black participants was generally higher than MAE for white participants by at least 0.5 years during tests with YOTI. (For the costume mask condition, MAE for Black participants was actually approximately 4 years less.)

We also note that FNR for male participants completing liveness detection tests for FACETEC and YOTI was lower at baseline; for FACETEC, FNR was consistently lower for male users on all test conditions. For YOTI, however, FNR for male users was higher on a handful of test conditions with non-zero liveness pass percentages, namely the frown, open-mouth, glasses, and smile test conditions.

Taken together, these results indicate the presence of gender- and race-induced effects on age and liveness detection performance. Notably, skin tone appears to have an observable effect on the likelihood that a user might receive a null or negative response from a liveness prediction test. Additionally, our analysis of near-boundary error rates suggests that younger users disproportionately received age overestimates, and older users disproportionately received age underestimates; in effect, users whom age gating policies intend to protect might be able to bypass age assurance systems in significant numbers, and of-age users might be mistakenly blocked from access.

5 Discussion

We conclude this work with a discussion of usability issues observed during study protocol completion, obstacles encountered during the recruitment process, a possible remedy to these difficulties, and a series of policy recommendations.

Generalizability of outcomes. We remark that these results are highly individual to the test condition under analysis, the services being tested, and the personal attributes of individual study participants. In particular, each age assurance provider has different specifications regarding proper usage: For instance, FACETEC implicitly requires that participants maintain a neutral facial expression (during testing, a smiling expression in particular would trigger a null response and an admonition to maintain a neutral facial expression). As such, while there are some general trends regarding, for instance, the impact of skin tone and degree of facial coverage on liveness performance, we emphasize that the idiosyncrasies we observe suggest that further due diligence testing is needed ahead of policy implementation. In particular, each application should be assessed with a large and diverse user base, and compliant usage should not be assumed.

Notes on usability. Each age assurance app presented users with unique usability challenges: notably, FACETEC returned null responses in more than 50% of checks, and YOTI was occasionally unable to detect a user's face. In particular, the UI for FACETEC's age assurance and liveness prediction processes was difficult to navigate. Users were told to follow a series of on-screen instructions, including bringing their head toward and away from the camera and framing their face in an outline displayed on the phone screen; at various points during these manual adjustments, the application recorded video and took photographs of the user's face. On repeated attempts, however, research team members were unable to determine *when* these still images and video were being taken. (The result: lots of age check images were taken when the participant was not ready, contributing to the high rate of null outcomes we observe.)

Recruiting populations of interest. We remark that identifying and recruiting populations of interest—including non-majority racial groups and teens—posed a significant challenge to the completion of the study protocol. While we were able to recruit a significant number of participants between the ages of 18 and 21 for our study, minor participants (under 18 years of age) were harder to reach. This was due in large part to the added complexity of moving adult parent/guardians through the full study permissions process. Though our status as university-affiliated researchers helped gain parent/guardian trust in some instances, we found that many guardians were reflexively disinclined to allow their child to be photographed for a research study on AI systems. We included non-technical explanations of our study and our risk mitigations in our parent permission documents; in many cases, however, parent/guardians neglected to read these documents in full.

We ultimately found our greatest success in snowball recruiting: parents who completed the study advertised it to their own networks. This allowed us to leverage established trust relationships between families to advertise our study.

False negative rates. Policy on age assurance must be attentive to high false negative and rejection rates. We demonstrate in this work that these rates can be outstandingly high during regular use and under adversarial evaluation conditions. High FNRs carry a number of consequences: 1) they have annoyance, burden, and speech implications for affected users; 2) they might channel users to more burdensome proofs of identity, such as ID checks, which have significant privacy and security implications; 3) they create incentives for platforms to *not* trigger age assurance flows for users because platforms risk losing customers in valuable age groups; and 4) they may prompt users toward evasive strategies.

Government intervention in due diligence testing. The previous discussion about the difficulties of recruiting minor populations suggests that an official avenue for informing, recruiting, and testing populations of interest could helpfully aid in due diligence testing. Such intervention is not without precedence: as mentioned in Section 2, the Australian and U.S. governments have deployed tests of age assurance providers, though the scope of these tests was fairly limited (e.g., NIST's tests were conducted solely on datasets of still images, and AATT did not report on liveness or disclose methodology or results) [2, 8, 9]. We urge regulatory bodies, governments, and third party age assurance providers to work in concert to develop and execute research programs that en-

gage the public as participants *in advance of* sweeping policy implementation. These tests should 1) partner with independent researchers and publicly release transparent methods and results; 2) extensively assess adversarial robustness, performance on a demographically diverse user base, and realistic deployment scenarios; and 3) consider null responses and liveness prediction performance.

Future work. There are many avenues for further investigation. As mentioned previously, a version of this protocol, completed with a much larger study population, could provide additional confirmatory evidence of the observations we present here. We additionally recommend additional adversarial testing of media streams, such as video injection and tampering with mobile device camera hardware. These types of attacks are more technically sophisticated than the presentation attacks we investigate here (we deliberately choose low entry-to-access attacks, to emulate scenarios replicable by a resourceful preteen or adolescent), and present larger attack surfaces. Finally, we recognize that facial scans are only one of a variety of (biometric) verification methods available to vendors and users. We encourage exploration of other, privacy-preserving alternatives [63].

6 Open Science

Our analysis plan, supplemental materials, and analysis notebooks are available at https://github.com/citp/face_off_os.

7 Ethical Considerations

IRB Review. This study was approved by the Institutional Review Boards of Princeton University and Carnegie Mellon University.

Stakeholders in ethics analysis. Primary ethical stakeholders in this project include study participants, minor and adult Internet users, age verification service providers, and online services required to implement age verification mechanisms. We consider each stakeholder in this analysis, with a particular focus on participants in this study.

Study participants. We evaluate the ethical implications of undertaking this work alongside those criteria set forth by the Belmont and Menlo Reports: respect for persons, beneficence, justice, and respect for law and the public interest.

Respect for persons. We recognized that study participants might have questions how about how their facial image data might be used, stored, or transmitted during or after the study. In order to study commercial facial analysis algorithms, we collected images and videos of participants' faces. We maintained this information on devices that we own and only we have access to, using strong encryption, authentication, and other security measures. We did not—and will not—disclose images or videos of study participants, with the one exception that we describe here.

Commercial facial analysis algorithms are commonly deployed as a service, where analyzing an image or video requires submitting a copy for automated analysis. We were only able to study these algorithms by submitting images and videos rather than running the analysis on our own devices.

We only submitted images or videos to a commercial service if it committed to the following stringent data protection requirements: 1) The service immediately deletes images and videos after automated analysis, not storing them at all. 2) The service solely uses the images and videos for automated analysis, not repurposing them in any way or allowing its employees to access them. 3) The service protects submitted images and videos with strong encryption, authentication, and other safeguards, consistent with a rigorous industry standard. 4) An independent auditor has confirmed that the service has implemented these requirements.

Beneficence. We carefully designed this study to minimize the risk that minor participants might learn strategies for attempting to bypass age checks. We did not ask minor participants to interact with or attempt to bypass any age check. During in-person study sessions, we provided verbal instructions and took pictures and video of participants' faces in various poses and with various props. We did not tell participants which algorithms we were testing, what the algorithms do, how we are testing the algorithms, which apps or websites use the algorithms, or how they use the algorithms. We also did not show participants any of the photos or videos that we collected, live feeds from the cameras that we used, or any analysis or result displays.

Additionally, we disclosed the full study protocol (including names of age verification service providers, data handling practices, and circumvention risks and our mitigation approaches) to the parents/guardians of the minor participants; parents/guardians were able to voluntarily disclose these details to their children at any point before, during, or after study sessions. We compiled and made available to parents/guardians a document describing some alternatives to biometric age verification methods; this document also appears in our online supplement. These alternatives include app-level parental controls, which allow parents/guardians and children to negotiate best practices for Internet usage among themselves. Our goal was to provide parents/guardians with as many resources as possible in order that they might make informed decisions about their children's Internet browsing habits.

Justice. We believe that risks to participants were already minimal, but we ensured that no particular study group was disproportionately disadvantaged by our study protocol or outreach methods. We sought to reach a diverse cross-section of Pittsburgh residents, and compensated all study participants at a rate well above minimum wage for the state of Pennsylvania. We made debrief materials available to all adult completers of our study session and additionally informed parents/guardians about the full study protocol in advance of their child's par-

ticipation. All adult participants gave their informed consent prior to the start of their study session; any information about the study protocol that was withheld from minor participants was only withheld in the interest of preventing these participants from learning strategies for circumventing age verification systems (as mentioned in the previous section, their parents/guardians were fully briefed in advance of the study session and were free to discuss all details with their children as they deemed appropriate).

Respect for law and the public interest. This study did not involve any breaches of the law, or of terms of use for any of the age verification services we studied. We believe that the results of this work will inform both sides of the ongoing debate about the benefits and harms of biometric age verification.

Other stakeholders and decision to proceed with research. We believe that this study's outcomes will be of immediate interest to age verification service providers, website administrators and app developers, and policymakers. We acknowledge that age verification as a practice has been the subject of extensive policy and sociological debates: proponents of age verification argue that children should be shielded from harmful content online, and that age verification is a remedy for documented harms to children and teen's mental health [64]. Opponents of age verification (and biometric age verification in particular) point out that the security and privacy harms of sweeping age gates—including data leaks and mishandling of PII [65]—are not commensurate with the actual performance of these systems, many of which do not effectively bar underage users from accessing prohibited content to begin with [16].

Irrespective of the benefits or shortcomings of either position, this study provides much-needed empirical evidence of the actual functioning and performance of these systems when deployed in a diversity of use cases: this data might inform policymakers seeking to understand the robustness and per-vendor performance of these systems; it might also be used by age verification service providers to understand where their systems fall short. We acknowledged some potential future cost to age verification providers and online services that might be required to update or re-implement their age verification mechanisms in light of these outcomes. We believed, however, that the public benefit arising from an independent audit of these systems far outweighs any material cost to vendors and platforms. As such, we decided to move forward with this work.

8 Acknowledgements

We would like to thank our research assistants for their diligence and attention to detail. Without their support, this work would not have been possible: Simran Agrawal, Zane Aqra, Aditi Balaji, Rachel Cheng, Jiao Deng, Rene Xu, and Mandy Yin. We are grateful to all study participants for their engagement. We would additionally like to thank Michael Taylor for co-designing and constructing mechanical arms for our phone rig. MX was supported by NSF Grant DGE-2039656. SC was supported by an Australian Research Council Discovery Early Career Researcher Award DE260100249.

References

[1] Parliament of Australia *Online Safety Amendment (Social Media Minimum Age) Act 2024*. https://www.legislation.gov.au/C2024A00127/latest/text. 2024 (1).

[2] Age Check Certification Scheme *Age Assurance Technology Trial: Final Report*. https://ageassurance.com.au/publications/. 2025 (1, 3, 6, 13).

[3] Richard Guest, ET AL. *Facial Age Estimation: A Research Roadmap for Technological and Legal Development and Deployment. arXiv preprint arXiv:2505.22401* (2025). URL: https://arxiv.org/abs/2505.22401 (1).

[4] eSafety Commissioner *Social Media Minimum Age Regulatory Guidance*. Tech. rep. eSafety Commissioner, Sept. 2025. URL: https://www.esafety.gov.au/sites/default/files/2025-09/eSafety-SMMA-Regulatory-Guidance.pdf (1).

[5] Shuang Liu AND Sarah Scheffler *Adequately Tailoring Age Verification Regulations. arXiv preprint arXiv:2601.20241* (2026). URL: https://arxiv.org/abs/2601.20241 (1).

[6] Free Speech Coalition *All Age Verification Bills*. Action Center, filtered to enacted bills, and grouped by jurisdiction. URL: https://action.freespeechcoalition.com/age-verification-bills/all/ (visited on 08/13/2026) (1, 2).

[7] Supreme Court of the United States *Free Speech Coalition, Inc. v. Paxton*. https://www.supremecourt.gov/opinions/24pdf/23-1122_3e04.pdf. 2025 (1).

[8] Patrick J. Grother, ET AL. *Face Analysis Technology Evaluation: Age Estimation and Verification*. https://www.nist.gov/programs-projects/face-analysis-technology-evaluation-fate/age-estimation-verification. 2024 (2, 3, 7, 9, 10, 13).

[9] Mei Ngan, Patrick Grother, AND Austin Hom *"Face Analysis Technology Evaluation (FATE) Part 10: Performance of Passive, Software-Based Presentation Attack Detection (PAD) Algorithms"*. https://nvlpubs.nist.gov/nistpubs/ir/2023/NIST.IR.8491.pdf. Sept. 2023 (2, 13).

[10] Eric Rescorla, Zander Arnao, AND Alissa Cooper *Age Assurance Online: A Technical Assessment of Current Systems and Their Limitations*. Tech. rep. Knight-Georgetown Institute, Jan. 2026. URL: https://kgi.georgetown.edu/wp-content/uploads/2026/01/Age_Assurance_Online_Technical-Assessment_Report_KGI.pdf (2, 3).

[11] Shreyas Minocha, ET AL. *Papers, Please: A First Look at Age Verification on the Web*. In: *2026 IEEE Symposium on Security and Privacy (SP)*. IEEE, May 2026. URL: https://mikespecter.com/assets/pdf/AgeVerification.pdf (2).

[12] eSafety Commissioner *Early Days, Early Insights: Understanding Experiences of Social Media Age Restrictions Three Months After Implementation*. Tech. rep. eSafety Commissioner, Aug. 2026. URL: https://www.esafety.gov.au/sites/

default/files/2026-07/Early-days-early-insights-SMMA-evaluation-three-months-July-2026.pdf (2).

[13] eSafety Commissioner *Social Media Minimum Age: Compliance Update*. Tech. rep. eSafety Commissioner, Mar. 2026. URL: https://www.esafety.gov.au/sites/default/files/2026-03/SocialMediaMinimumAgeComplianceUpdateMarch2026.pdf (2).

[14] C. Barnes, ET AL. *Assessing Early Effects of Australia's Social Media Minimum Age Act on Adolescents' Social Media Use: Observational Study*. *BMJ* 393 (2026), e363695 (2).

[15] Leonardo Bursztyn, ET AL. *Why Bans Fail: Tipping Points and Australia's Social Media Ban*. Working Paper 2026-57. Becker Friedman Institute for Economics at the University of Chicago, Apr. 2026. URL: https://bfi.uchicago.edu/wp-content/uploads/2026/04/BFI_WP_2026-57.pdf (2).

[16] Jacob Ridley *"Brits can get around Discord's age verification thanks to Death Stranding's photo mode, bypassing the measure introduced with the UK's Online Safety Act. We tried it and it works—thanks, Kojima"*. PC Gamer. July 2025. URL: https://www.pcgamer.com/hardware/brits-can-get-around-discords-age-verification-thanks-to-death-strandings-photo-mode-bypassing-the-measure-introduced-with-the-uks-online-safety-act-we-tried-it-and-it-works-thanks-kojima/ (2, 15).

[17] Ofcom *Children's Online Experiences: Including a Comparative Analysis of the Periods Before and After the Children's Safety Duties Came into Force*. Tech. rep. Ofcom, May 2026. URL: https://www.ofcom.org.uk/siteassets/resources/documents/online-safety/research-statistics-and-data/protecting-children/childrens-online-experiences-research-report.pdf (2).

[18] Bjørn Nansen, ET AL. *From Phreaking to Sneaking: Children's Circumvention of Social Media Age Verification Systems*. *arXiv preprint arXiv:2605.00368* (2026). Preprint. Accepted to OzCHI 2026 (2).

[19] Ofcom *Online Nation 2025*. Tech. rep. Published 10 December 2025. Ofcom, Dec. 2025. URL: https://www.ofcom.org.uk/siteassets/resources/documents/research-and-data/online-research/online-nation/2025/online-nations-report-2025.pdf (2).

[20] Dhyey Mehta, ET AL. *Online Safety Regulation Increases Privacy Risk: Evidence from the UK Online Safety Act*. arXiv preprint arXiv:2606.05273. Submitted to PoPETs 2027. 2026 (2).

[21] Yanzi Lin, ET AL. *What Adults Will (and Won't) Do to Prove Their Age: Empirical Evidence from a Deceptive Web Experiment*. In: *35th USENIX Security Symposium (USENIX Security 26)*. USENIX Association, 2026. URL: https://www.usenix.org/system/files/usenixsecurity26-lin-yanzi.pdf (3).

[22] Ioannis Sarridis, Ioannis Kompatsiaris, AND Symeon Papadopoulos *Face Age Verification Vulnerabilities Under Simple Appearance Manipulations*. 2026. URL: https://arxiv.org/abs/2607.24194 (3).

[23] Xingyu Shen, ET AL. *Can a Teenager Fool an AI? Evaluating Low-Cost Cosmetic Attacks on Age Estimation Systems*. 2026. URL: https://arxiv.org/abs/2602.19539 (3).

[24] Simone Lavermicocca, Michele Carminati, AND Stefano Longari *X-rated Compliance Theater: An Empirical Evaluation of European Age Verification Systems in Adult Websites*. In: *35th USENIX Security Symposium (USENIX Security 26)*. USENIX Association, 2026. URL: https://www.usenix.org/conference/usenixsecurity26/presentation/lavermicocca (3).

[25] James Alexander AND Jonathan M. Smith *Engineering Privacy in Public: Confounding Face Recognition*. In: *Privacy Enhancing Technologies*. Ed. by Roger Dingledine. Vol. 2760. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, 2003, 88–106. URL: https://doi.org/10.1007/978-3-540-40956-4_7 (3).

[26] Katherine L. Gibson AND Jonathan M. Smith *The Emperor's New Masks: On Demographic Differences and Disguises*. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. June 2015, 57–64. URL: https://openaccess.thecvf.com/content_cvpr_workshops_2015/W02/html/Gibson_The_Emperors_New_2015_CVPR_paper.html (3).

[27] Mahmood Sharif, ET AL. *Accessorize to a Crime: Real and Stealthy Attacks on State-of-the-Art Face Recognition*. In: *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. CCS '16. Association for Computing Machinery, New York, NY, USA, 2016, 1528–1540. URL: https://doi.org/10.1145/2976749.2978392 (3, 5).

[28] Christian Rathgeb, ET AL., eds. *Handbook of Digital Face Manipulation and Detection: From DeepFakes to Morphing Attacks*. Advances in Computer Vision and Pattern Recognition. Springer, Cham, 2022. URL: https://doi.org/10.1007/978-3-030-87664-7 (3, 5).

[29] Takayuki Yamada, Seiichi Gohshi, AND Isao Echizen *Privacy Visor: Method Based on Light Absorbing and Reflecting Properties for Preventing Face Image Detection*. In: *2013 IEEE International Conference on Systems, Man, and Cybernetics*. IEEE, 2013, 1572–1577. URL: https://doi.org/10.1109/SMC.2013.271 (3, 5).

[30] Adam Harvey *CV Dazzle: Camouflage from Computer Vision*. Master's thesis project, Interactive Telecommunications Program, New York University. 2010. URL: https://adam.harvey.studio/cvdazzle (3, 5).

[31] Nesli Erdogmus AND Sébastien Marcel *Spoofing in 2D Face Recognition with 3D Masks and Anti-Spoofing with Kinect*. In: *2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*. IEEE, 2013, 1–6. URL: https://doi.org/10.1109/BTAS.2013.6712688 (3, 5).

[32] Yan Li, ET AL. *Understanding OSN-Based Facial Disclosure Against Face Authentication Systems*. In: *Proceedings of the 9th ACM Symposium on Information, Computer and Communications Security*. ACM, 2014, 413–424. URL: https://doi.org/10.1145/2590296.2590315 (3, 5).

[33] Matthew Turk AND Alex Pentland *Eigenfaces for Recognition*. *Journal of Cognitive Neuroscience* 3, 1 (1991), 71–86. URL: https://doi.org/10.1162/jocn.1991.3.1.71 (3, 5).

[34] Peter N. Belhumeur, João P. Hespanha, AND David J. Kriegman *Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection*. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19, 7 (1997), 711–720. URL: https://doi.org/10.1109/34.598228 (3, 5).

[35] Ranran Feng and Balakrishnan Prabhakaran *Facilitating Fashion Camouflage Art*. In: *Proceedings of the 21st ACM International Conference on Multimedia*. ACM, 2013, 793–802. URL: https://doi.org/10.1145/2502081.2502121 (3, 5).

[36] Xin Zheng, et al. *Robust Physical-World Attacks on Face Recognition*. *Pattern Recognition* 133 (2023), 109009. URL: https://doi.org/10.1016/j.patcog.2022.109009 (3, 5).

[37] Yinpeng Dong, et al. *Efficient Decision-Based Black-Box Adversarial Attacks on Face Recognition*. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2019, 7706–7714. URL: https://doi.org/10.1109/CVPR.2019.00790 (3, 5).

[38] Dinh-Luan Nguyen, et al. *Adversarial Light Projection Attacks on Face Recognition Systems: A Feasibility Study*. In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, 2020, 3548–3556. URL: https://doi.org/10.1109/CVPRW50498.2020.00415 (3, 5).

[39] Esha Sarkar, et al. *FaceHack: Attacking Facial Recognition Systems Using Malicious Facial Characteristics*. *IEEE Transactions on Biometrics, Behavior, and Identity Science* 4, 3 (2022), 361–372. URL: https://doi.org/10.1109/TBIOM.2021.3132132 (3, 5).

[40] Shuai Jia, et al. *Adv-Attribute: Inconspicuous and Transferable Adversarial Attack on Face Recognition*. In: *Advances in Neural Information Processing Systems*. Vol. 35. Curran Associates, Inc., 2022. URL: https://proceedings.neurips.cc/paper_files/paper/2022/hash/dccbeb7a8df3065c4646928985edf435-Abstract-Conference.html (3, 5).

[41] Meng Shen, et al. *Effective and Robust Physical-World Attacks on Deep Learning Face Recognition Systems*. *IEEE Transactions on Information Forensics and Security* 16 (2021), 4063–4077. URL: https://doi.org/10.1109/TIFS.2021.3102492 (3, 5).

[42] Amy Winecoff and Alissa Cooper *Age Assurance Rules: An Implementation Guide*. Tech. rep. Knight-Georgetown Institute, July 2026. URL: https://kgi.georgetown.edu/research-and-commentary/age-assurance-rules-implementation/ (3).

[43] Amy Winecoff and Alissa Cooper *Assigning Responsibility for Age Assurance: Recommendations for Youth Online Safety*. Tech. rep. Knight-Georgetown Institute, July 2026. URL: https://kgi.georgetown.edu/research-and-commentary/assigning-responsibility-for-age-assurance/ (3).

[44] European Commission *A Digital Decade for Children and Youth: The New European Strategy for a Better Internet for Kids (BIK+)*. COM(2022) 212 final. May 2022. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52022DC0212 (3).

[45] European Commission *EU Strategy on the Rights of the Child*. COM(2021) 142 final. Mar. 2021. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021DC0142 (3).

[46] Khaled ELKarazle, Valliappan Raman, and Patrick Then *Facial Age Estimation Using Machine Learning Techniques: An Overview*. *Big Data and Cognitive Computing* 6, 4 (2022). URL: https://www.mdpi.com/2504-2289/6/4/128 (3).

[47] Anonymous *Online Supplement*. Open Science Framework. 2026 (3, 5, 9).

[48] AVPA *Age Verification Providers Association*. https://avpassociation.com/. 2026 (4).

[49] Yoti *Facial Age Estimation White Paper*. https://www.yoti.com/blog/facial-age-estimation-white-paper/. 2024 (4, 6).

[50] Claude Support *Age assurance on Claude* (May 2026). URL: https://support.claude.com/en/articles/15171100-age-assurance-on-claude (4).

[51] Meta *''Introducing New Ways to Verify Age on Instagram''*. https://about.fb.com/news/2022/06/new-ways-to-verify-age-on-instagram/. 2022 (4).

[52] Yoti *''What does Yoti do to ensure its services are accessible by all?''* https://support.yoti.com/yotisupport/s/article/What-does-Yoti-do-to-ensure-its-services-are-accessible-by-all. 2025 (4).

[53] FaceTec *NIST FRVT-PAD Commentary* (Apr. 2022). URL: https://facetec.com/FaceTec_NIST_FRVT-PAD_Comments.pdf (4).

[54] ID Tech Editorial Team *FaceTec Reports UR Code Deployments Across Law Enforcement and Public Safety* (Aug. 2026). URL: https://idtechwire.com/facetec-reports-ur-code-deployments-across-law-enforcement-and-public-safety/ (4).

[55] Stanislav Vishnevskiy *''Getting Global Age Assurance Right: What We Got Wrong and What's Changing''*. https://discord.com/blog/getting-global-age-assurance-right-what-we-got-wrong-and-whats-changing. 2026 (4).

[56] Ashley Belanger *''Users hate it, but age-check tech is coming. Here's how it works.''* https://arstechnica.com/tech-policy/2026/03/after-discord-fiasco-age-check-tech-promises-privacy-by-running-locally-does-it-work/. 2026 (4).

[57] *''Age Against the Machine: Offline Model Extraction, Evaluation, and Evasion for Production Biometric Age Assurance Systems''* (2026) (5).

[58] Ofcom *Guidance on Highly Effective Age Assurance*. https://www.ofcom.org.uk/online-safety/protecting-children/highly-effective-age-assurance/. 2025 (6, 9).

[59] Matt Prendergast *"German regulator KJM cuts buffer for Yoti's Facial Age Estimation for 18+ content checks from 5 to 3 years"*. 2025. URL: https://www.yoti.com/blog/age-estimation-approved-kjm-highest-age-assurance-level-18-adult-content/ (6).

[60] Ellis Monk *The Monk Skin Tone Scale*. *SocArXiv* (May 4, 2023). URL: https://doi.org/10.31235/osf.io/pdf4c (7).

[61] Yoti *Yoti MyFace liveness white paper*. https://cdn.aws.yoti.com/wp-content/uploads/2026/01/Yoti-MyFace-Liveness-Paper-Jan-2026-public.pdf. 2026 (8).

[62] Facetec *''Facetec Glossary of Terms''*. https://www.facetec.com/glossary (8, 11).

[63] Steven M Bellovin *Privacy-Preserving Age Verification—and Its Limitations*. *Preprint, submitted October* (2025) (14).

[64] Nancy Costello, et al. *Algorithms, addiction, and adolescent mental health: An interdisciplinary study to inform state-level policy action to protect youth from the dangers of social media*. *American Journal of Law & Medicine* 49, 2-3 (2023), 135–172 (15).

[65] Rindala Alajaji AND Samantha Baldwin *"Discord Voluntarily Pushes Mandatory Age Verification Despite Recent Data Breach"*. Electronic Frontier Foundation. Feb. 2026. URL: https://www.eff.org/deeplinks/2026/02/discord-voluntarily-pushes-mandatory-age-verification-despite-recent-data-breach (15).

A Test cases

Table 5: Test cases and descriptions.

Test condition	Participant instruction
Baseline (ordinary pose)	We ask the participant to pose with a neutral expression.
Eyeglasses	We ask the participant to put on a pair of glasses with plastic lenses.
KN95 mask	We ask the participant to put on a white surgical KN95 mask.
Groucho Marx glasses	We ask the participant to put on a costume disguise with a mustache, nose, eyebrows, and glasses.
Adhesive mustache	We ask the participant to apply a stick-on mustache to their upper lip.
Costume beard	We ask the participant to put on a costume beard.
Eyepatch	We ask the participant to put on an eyepatch over their right eye.
Salt & pepper wig	We ask the participant to put on a curly gray wig.
Full face mask	We ask the participant to put on a full-face rubber costume mask.
Open-mouth expression	We ask the participant to mime an open-mouthed, shocked expression.
Frowning expression	We ask the participant to frown with knitted brows.
Smiling expression	We ask the participant to smile with an open mouth.

B Age outputs

Table 6: Age estimation outputs for all three providers and how we documented them.

Service	Output & Documentation
Facetec	**Binary age estimate**; binary age prediction is with respect to four or more age thresholds, selected from 13, 16, 18, 21, 25, and 30. Facetec may output a message instead of an age estimate, such as ‘‘Let’s try that again.’’ The service may also fail to output a value, instead continuing to provide instructions such as moving closer or further away. We treat these outputs as null responses (*n*).
Privately	**Binary age estimate**; binary age estimation is with respect to three age thresholds, 21, 25, and 30. We have not observed any instances in study data collection of Privately outputting a message instead of an age estimate or not outputting a value.
Yoti	**Integer age estimate**; we convert integer age estimates to binary age estimates, as described in Section 3.4. Yoti may also output a message instead of an age estimate, such as “Face not found” or “Multiple faces detected.” We record these as null responses (*n*).

C FNR by condition and gender identity

Table 7: FNR by condition and self-reported gender identity for YOTI.

Condition	Gender	*N*	FNR
Baseline	Female	96	0.34
	Male	70	0.29
	Non-binary	4	0.50
Beard	Female	78	1.00
	Male	60	1.00
	Non-binary	4	1.00
Costume mask	Female	64	1.00
	Male	50	0.98
	Non-binary	4	1.00
Eyepatch	Female	90	1.00
	Male	68	1.00
	Non-binary	4	1.00
Frown	Female	72	0.40
	Male	39	0.54
	Non-binary	4	0.75
Open-mouth	Female	74	0.78
	Male	39	0.82
	Non-binary	3	1.00
Glasses	Female	96	0.25
	Male	70	0.30
	Non-binary	4	0.75
Groucho	Female	89	1.00
	Male	58	1.00
	Non-binary	4	1.00
KN95	Female	89	0.99
	Male	65	1.00
	Non-binary	4	1.00
Smile	Female	77	0.38
	Male	44	0.57
Mustache	Female	81	0.99
	Male	60	1.00
	Non-binary	4	1.00
Wig	Female	92	1.00
	Male	69	1.00
	Non-binary	4	1.00

Table 8: FNR by condition and self-reported gender identity for FACETEC.

Condition	Gender	*N*	FNR
Baseline	Female	87	0.51
	Male	54	0.39
	Non-binary	4	0.50
Beard	Female	56	0.96
	Male	36	0.86
Costume mask	Female	48	1.00
	Male	28	1.00
Eyepatch	Female	60	1.00
	Male	32	1.00
Frown	Female	63	0.48
	Male	30	0.37
Open-mouth	Female	55	0.71
	Male	28	0.50
Glasses	Female	80	0.63
	Male	49	0.45
	Non-binary	4	0.50
Groucho	Female	56	1.00
	Male	22	1.00
KN95	Female	61	1.00
	Male	32	1.00
Smile	Female	54	0.91
	Male	22	0.82
Mustache	Female	66	0.86
	Male	44	0.61
	Non-binary	3	0.67
Wig	Female	84	0.46
	Male	48	0.33
	Non-binary	4	0.00

D FNR and FPR by condition and racial group

Table 9: FNR and FPR by condition and self-reported racial group for FACETEC. The table is continued in the right-hand panel.

Condition	Race	*N*	FNR	FPR
Baseline	AI/AN	6	0.5455	0
	AS	228	0.6244	0.1038
	BL/AA	23	0.5455	0
	HI/LA	59	0.7222	0.0571
	ME/NA	5	0.5	0
	MR	61	0.5893	0.125
	NH/PI	5	0.6667	0
	OT	17	0.5357	0
	WH	187	0.564	0.0722
Glasses	AS	176	0.6023	0.1798
	BL/AA	16	0.5714	0
	HI/LA	52	0.6522	0.1852
	ME/NA	6	0.5	0
	MR	52	0.625	0.2667
	NH/PI	3	1	0
	OT	8	0.5455	0
	WH	122	0.5571	0.1852
Open-mouth	AI/AN	3	0.5	0
	AS	92	0.6753	0.0698
	BL/AA	16	0.5217	0.3333
	HI/LA	16	0.6667	0
	MR	26	0.5417	0.1333
	NH/PI	8	0.5714	0.3333
	OT	12	0.5333	0
	WH	41	0.5484	0.04
Frown	AI/AN	7	0.6	0
	AS	162	0.6352	0.1408
	BL/AA	32	0.561	0
	HI/LA	38	0.76	0
	MR	58	0.6071	0.0769
	NH/PI	8	0.5714	0
	OT	16	0.5417	0
	WH	136	0.5847	0.1299

Condition	Race	*N*	FNR	FPR
Smile	AS	22	0.6111	0
	BL/AA	18	0.5455	0.1429
	HI/LA	13	0.6667	0
	MR	17	0.5714	0.1
	NH/PI	6	0.5	0.25
	OT	7	0.6	0
	WH	26	0.6842	0.0714
Mustache	AS	62	0.623	0.2
	BL/AA	15	0.5	0.25
	HI/LA	25	0.6087	0.0833
	MR	13	0.5385	0.1429
	NH/PI	6	0.5	0.25
	WH	35	0.5278	0.1579
Beard	AS	35	0.5957	0.125
	HI/LA	7	0.7143	0
	MR	15	0.7333	0
	WH	15	0.7333	0
Wig	AI/AN	12	0.6667	0
	AS	212	0.6862	0.046
	BL/AA	39	0.6087	0
	HI/LA	59	0.8485	0
	MR	64	0.7143	0
	NH/PI	8	0.6667	0
	OT	15	0.5217	0
	WH	157	0.6303	0.0465

E Pooled MAE by skin tone

Table 10: Pooled MAE by self-reported gender identity and Monk skin tone score across all test conditions.

Gender	Skin tone	*N*	MAE
Female	5.0	178	0.051
	6.0	296	0.058
	7.0	151	0.056
	8.0	44	0.068
Male	5.0	43	0.044
	6.0	138	0.047
	7.0	163	0.048
Non-binary	8.0	24	0.056
Transgender	7.0	23	0.046

Table 11: Pooled MAE by age band and Monk skin tone score across all test conditions.

Age band	Skin tone	*N*	MAE
13–16	5.0	44	0.065
	6.0	82	0.040
	7.0	63	0.068
17–20	5.0	86	0.041
	6.0	88	0.066
	7.0	139	0.050
	8.0	48	0.071
21–24	5.0	45	0.044
	6.0	104	0.046
	7.0	22	0.079
25–28	6.0	71	0.069
29–32	5.0	46	0.054
	6.0	41	0.035
	7.0	20	0.027

F Liveness pass percentages, Yoti

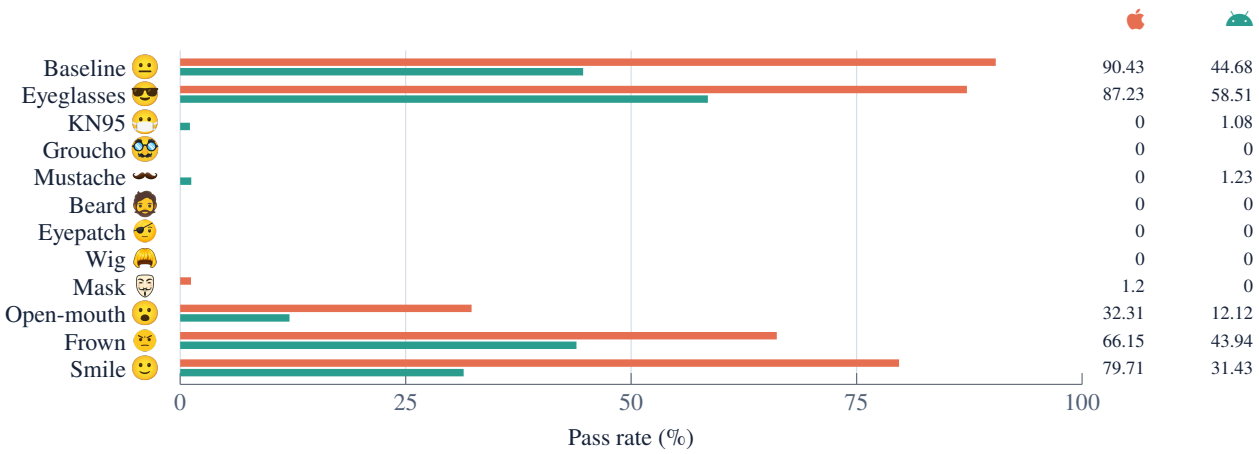

Figure 6: Liveness pass percentage relative to device-specific baseline/control for Yoti only. The baseline pass percentage for an iPhone was 90.43%; the baseline pass percentage for an Android was 44.68%.

G Error quartiles by device, Yoti

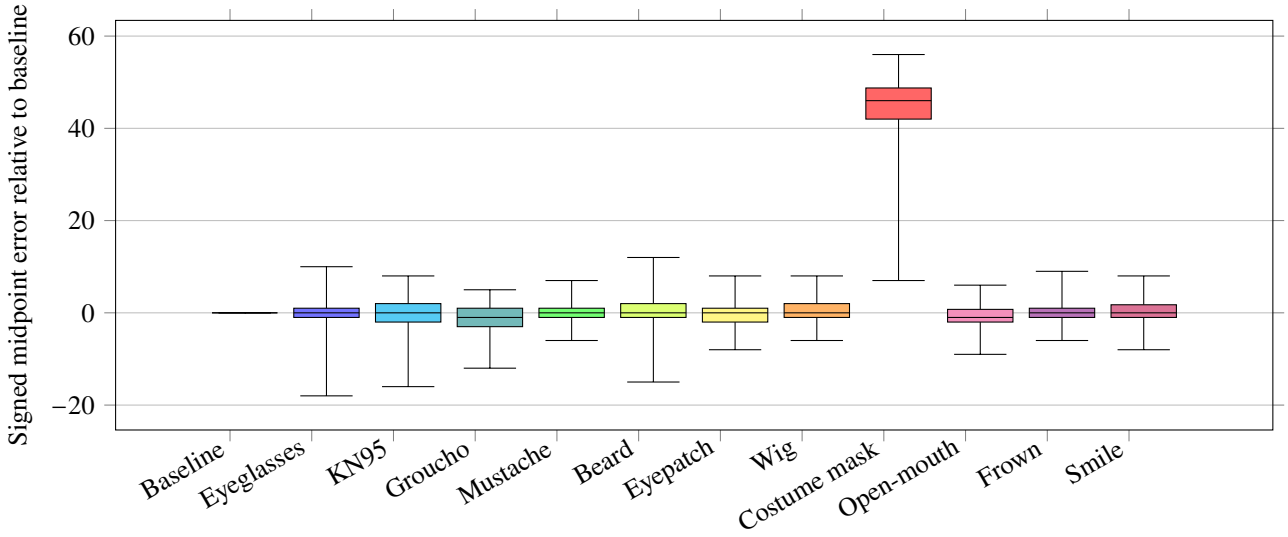

Figure 7: Distribution of signed midpoint error relative to Android baseline.

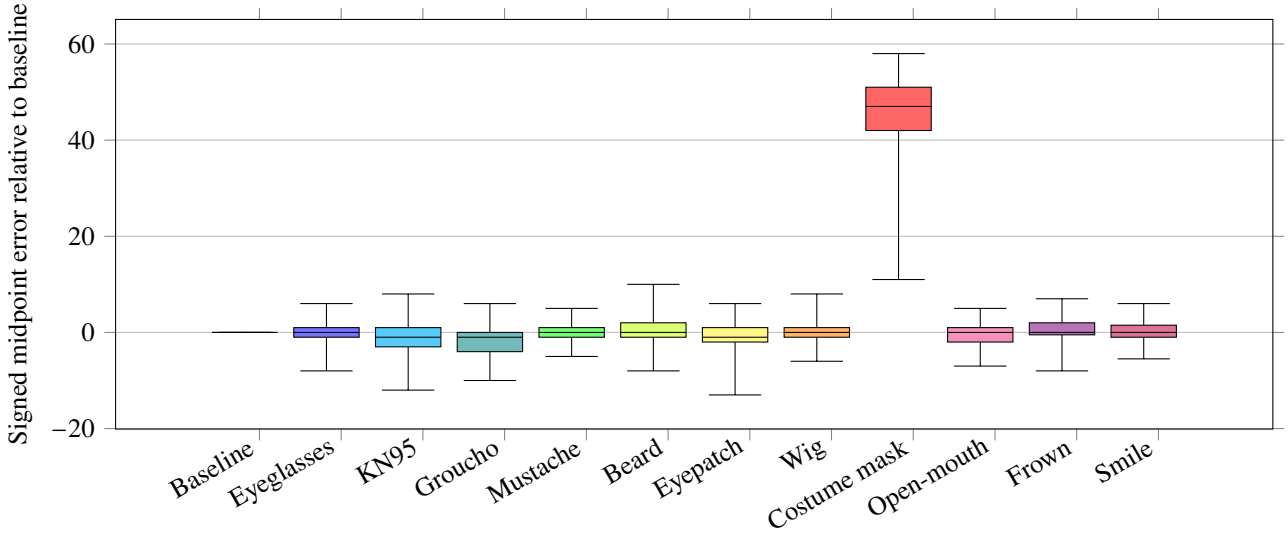

Figure 8: Distribution of signed midpoint error relative to iPhone baseline.

H FNR and FPR by age band, Yoti

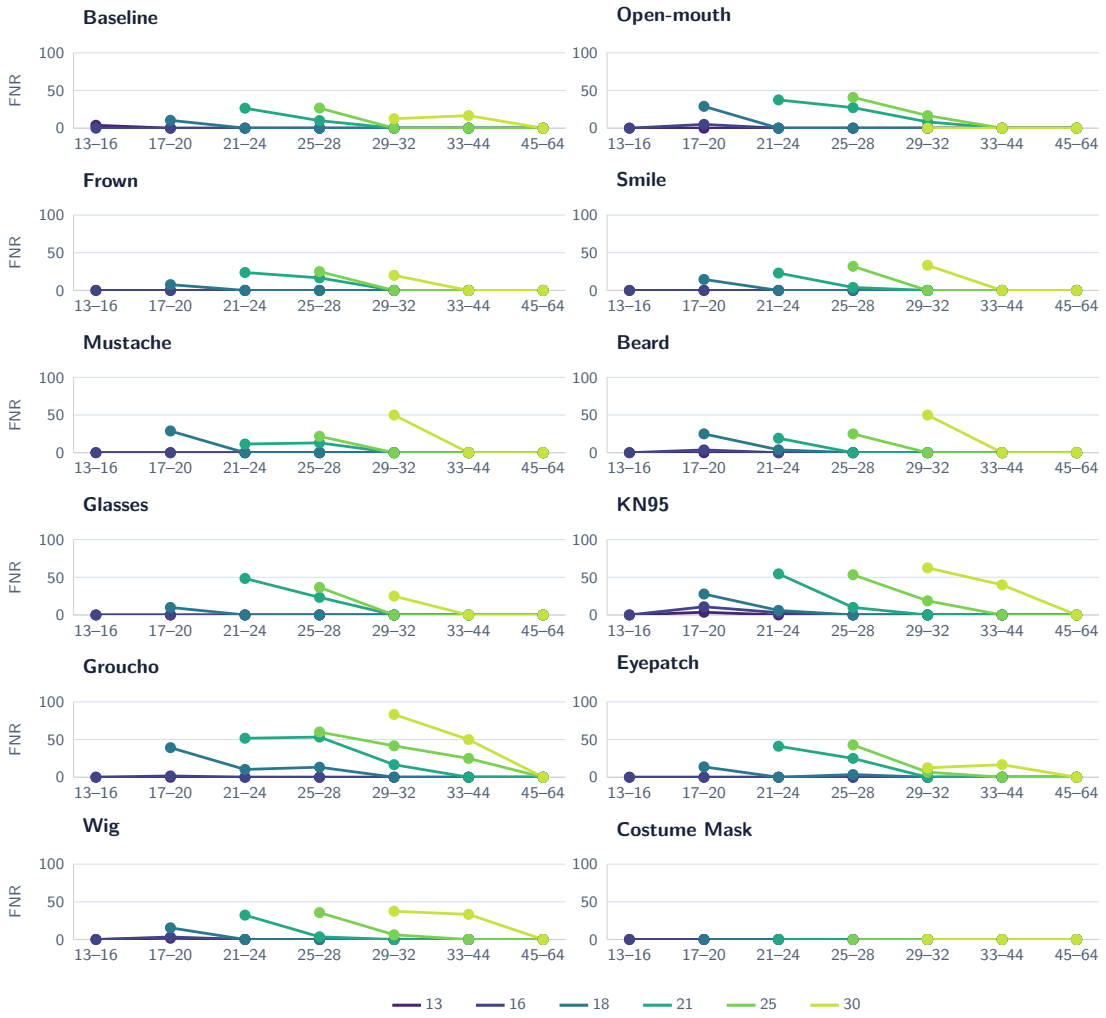

Figure 9: Yoti's FNR performance on all conditions and baseline, stratified by age group.

Figure 10: Yoti's FPR performance on all conditions and baseline, stratified by age group.

I Buffer sensitivity, Yoti

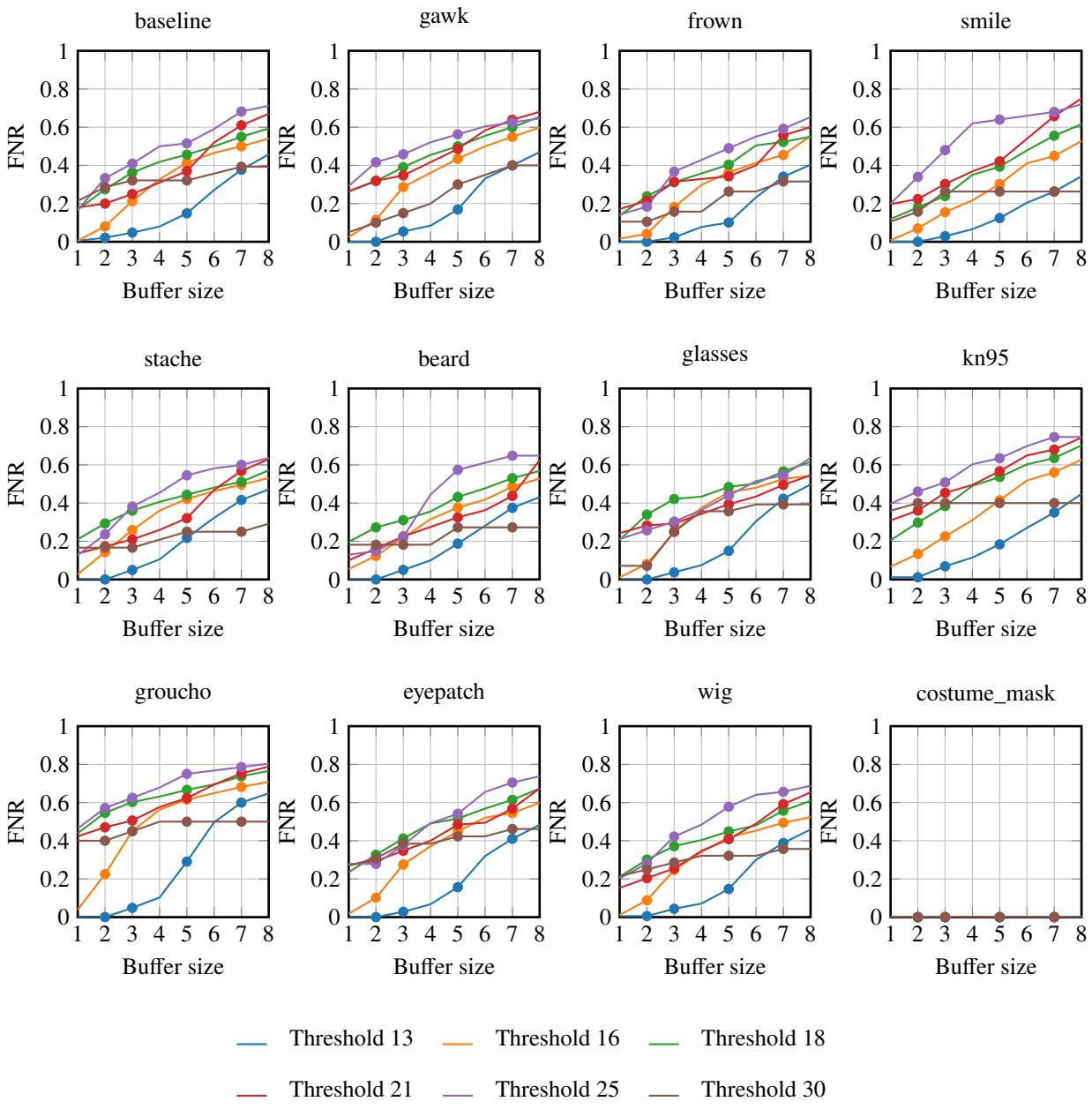

Figure 11: False negative rate as a function of Yoti buffer size across age thresholds and test conditions. Each colored curve represents an age threshold and traces buffer sizes 1–8. Points identify buffer sizes 2, 3, 5, and 7. For test conditions, interval-valued Yoti age estimates are converted to their interval midpoints before applying the buffer. iPhone and Android observations are pooled within each condition.

J FNR and FPR by age band, PRIVATELY

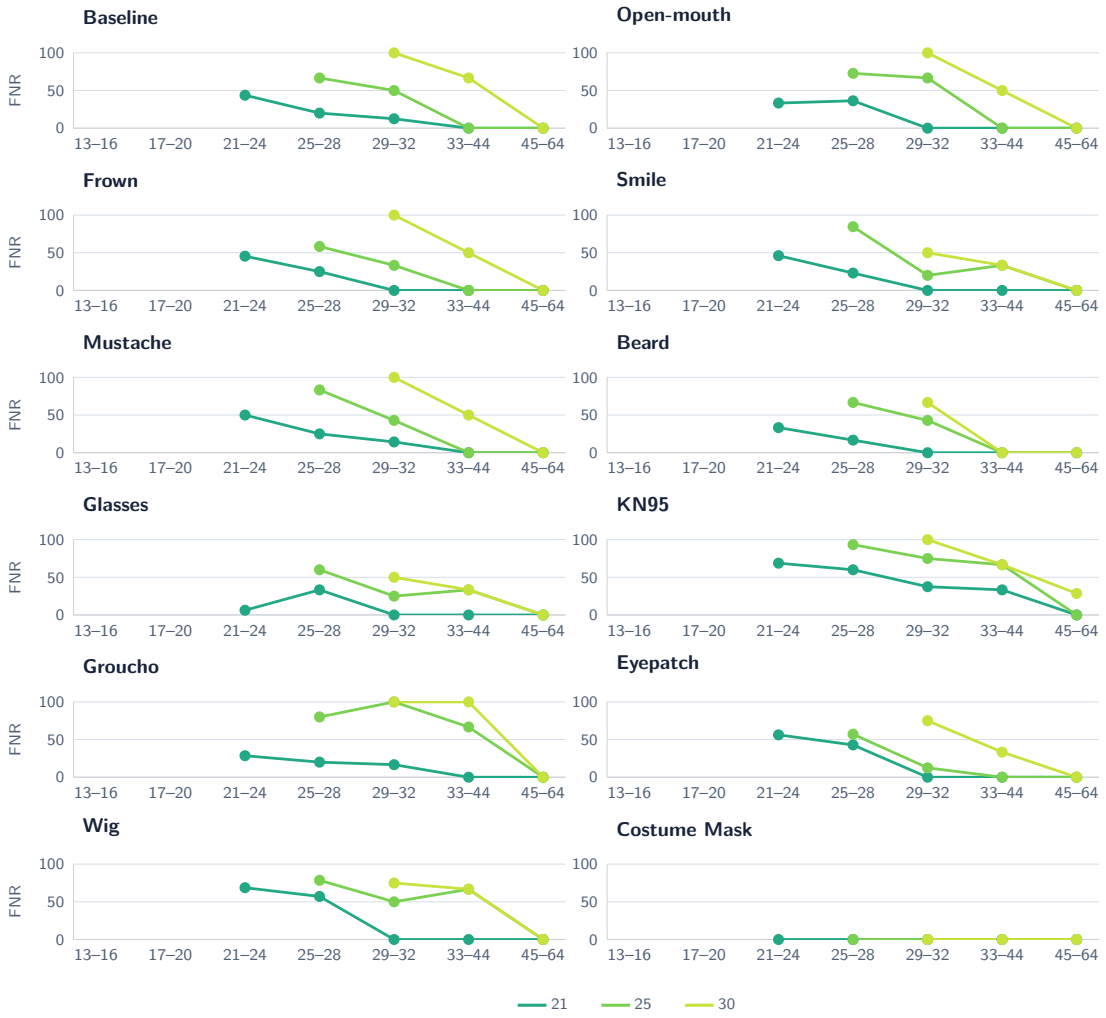

Figure 12: PRIVATELY's FNR performance on all conditions and baseline, stratified by age group. Each line represents a different service threshold t_s. Null responses are recorded as blank or NaN values (not zeroes).

Figure 13: PRIVATELY's FPR performance on all conditions and baseline, stratified by age group. Each line represents a different service threshold t_s. Null responses are recorded as blank or NaN values (not zeroes).

K FNR and FPR by age band, FACETEC

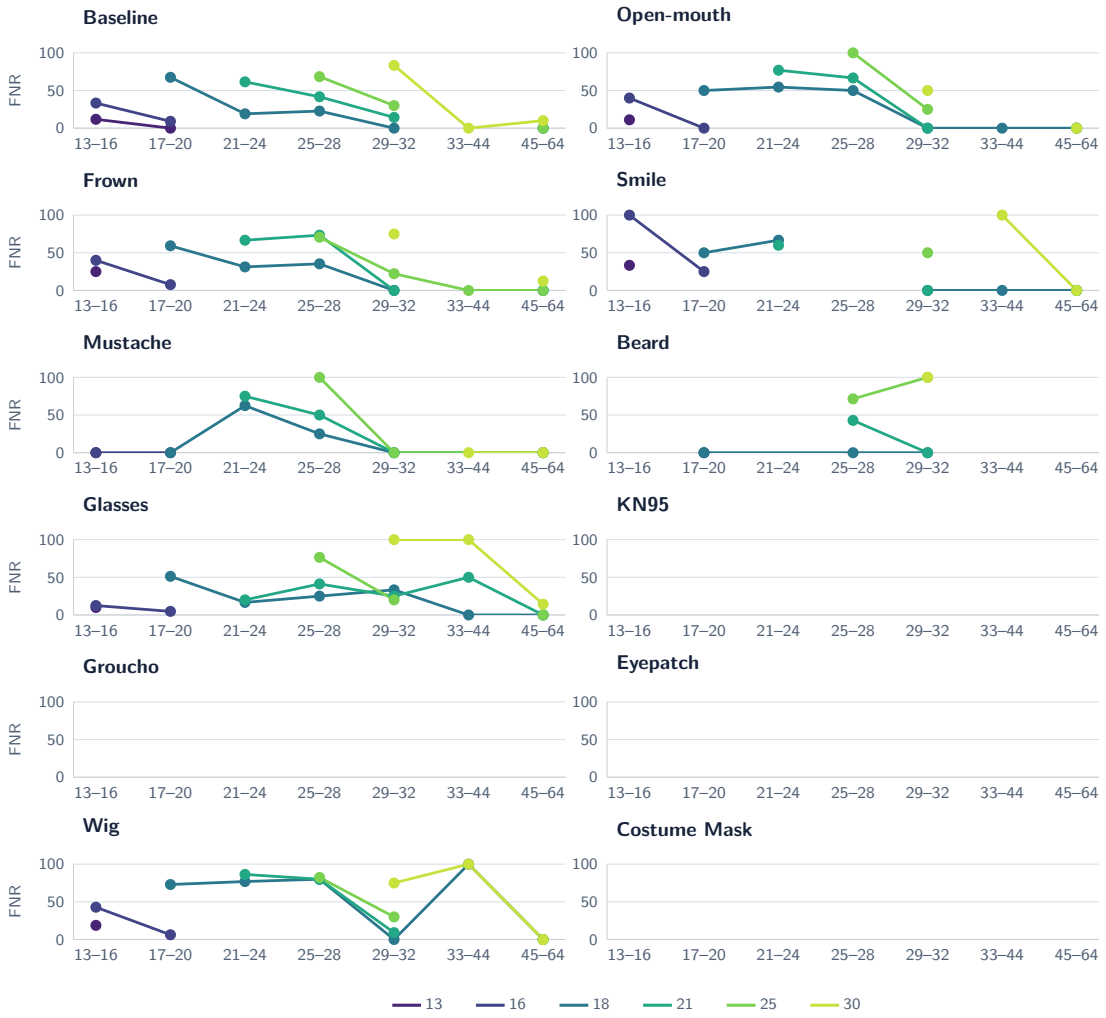

Figure 14: Facetec's FNR performance on all conditions and baseline, stratified by age group. Each line represents a different service threshold t_s. Null responses are recorded as blank or NaN values (not zeroes).

Figure 15: Facetec's FPR performance on all conditions and baseline, stratified by age group. Each line represents a different service threshold t_s. Null responses are recorded as blank or NaN values (not zeroes).

L FNR and FPR by age band, Yoti

Table 12: Performance by age band and test condition. FPR and FNR are proportions rounded to two decimal places.

Threshold	Age Band	*N*	FPR	FNR	Threshold	Age Band	*N*	FPR	FNR	Threshold	Age Band	*N*	FPR	FNR
	Baseline					**Open-mouth**					**Frown**			
13	13-16	26	–	0.04	13	13-16	26	–	0.00	13	13-16	26	–	0.00
	17-20	62	–	0.00		17-20	62	–	0.00		17-20	62	–	0.00
	21-24	34	–	0.00		21-24	34	–	0.00		21-24	34	–	0.00
	25-28	30	–	0.00		25-28	30	–	0.00		25-28	30	–	0.00
	29-32	16	–	0.00		29-32	16	–	0.00		29-32	16	–	0.00
	33-44	6	–	0.00		33-44	6	–	0.00		33-44	6	–	0.00
	45-64	14	–	0.00		45-64	14	–	0.00		45-64	14	–	0.00
16	13-16	26	0.36	0.00	16	13-16	26	0.25	0.00	16	13-16	26	0.60	0.00
	17-20	62	–	0.00		17-20	62	–	0.05		17-20	62	–	0.00
	21-24	34	–	0.00		21-24	34	–	0.00		21-24	34	–	0.00
	25-28	30	–	0.00		25-28	30	–	0.00		25-28	30	–	0.00
	29-32	16	–	0.00		29-32	16	–	0.00		29-32	16	–	0.00
	33-44	6	–	0.00		33-44	6	–	0.00		33-44	6	–	0.00
	45-64	14	–	0.00		45-64	14	–	0.00		45-64	14	–	0.00
18	13-16	26	0.35	–	18	13-16	26	0.64	–	18	13-16	26	0.60	–
	17-20	62	0.00	0.15		17-20	62	0.00	0.29		17-20	62	0.50	0.08
	21-24	34	–	0.00		21-24	34	–	0.00		21-24	34	–	0.00
	25-28	30	–	0.00		25-28	30	–	0.00		25-28	30	–	0.00
	29-32	16	–	0.00		29-32	16	–	0.00		29-32	16	–	0.00
	33-44	6	–	0.00		33-44	6	–	0.00		33-44	6	–	0.00
	45-64	14	–	0.00		45-64	14	–	0.00		45-64	14	–	0.00
21	13-16	26	0.00	–	21	13-16	26	0.14	–	21	13-16	26	0.13	–
	17-20	62	0.23	–		17-20	62	0.28	–		17-20	62	0.34	–
	21-24	34	–	0.26		21-24	34	–	0.38		21-24	34	–	0.24
	25-28	30	–	0.10		25-28	30	–	0.27		25-28	30	–	0.17
	29-32	16	–	0.00		29-32	16	–	0.08		29-32	16	–	0.00
	33-44	6	–	0.00		33-44	6	–	0.00		33-44	6	–	0.00
	45-64	14	–	0.00		45-64	14	–	0.00		45-64	14	–	0.00
25	13-16	26	0.00	–	25	13-16	26	0.00	–	25	13-16	26	0.00	–
	17-20	62	0.05	–		17-20	62	0.05	–		17-20	62	0.12	–
	21-24	34	0.32	–		21-24	34	0.21	–		21-24	34	0.19	–
	25-28	30	–	0.27		25-28	30	–	0.41		25-28	30	–	0.25
	29-32	16	–	0.00		29-32	16	–	0.17		29-32	16	–	0.00
	33-44	6	–	0.00		33-44	6	–	0.00		33-44	6	–	0.00
	45-64	14	–	0.00		45-64	14	–	0.00		45-64	14	–	0.00
30	13-16	26	0.00	–	30	13-16	26	0.00	–	30	13-16	26	0.00	–
	17-20	62	0.00	–		17-20	62	0.00	–		17-20	62	0.00	–
	21-24	34	0.00	–		21-24	34	0.00	–		21-24	34	0.00	–
	25-28	30	0.07	–		25-28	30	0.05	–		25-28	30	0.25	–
	29-32	16	0.50	0.12		29-32	16	0.00	0.00		29-32	16	0.17	0.20
	33-44	6	–	0.17		33-44	6	–	0.00		33-44	6	–	0.00
	45-64	14	–	0.00		45-64	14	–	0.00		45-64	14	–	0.00
	Smile					**Mustache**					**Beard**			
13	13-16	26	–	0.00	13	13-16	26	–	0.00	13	13-16	26	–	0.00
	17-20	62	–	0.00		17-20	62	–	0.00		17-20	62	–	0.00
	21-24	34	–	0.00		21-24	34	–	0.00		21-24	34	–	0.00
	25-28	30	–	0.00		25-28	30	–	0.00		25-28	30	–	0.00
	29-32	16	–	0.00		29-32	16	–	0.00		29-32	16	–	0.00
	33-44	6	–	0.00		33-44	6	–	0.00		33-44	6	–	0.00
	45-64	14	–	0.00		45-64	14	–	0.00		45-64	14	–	0.00
16	13-16	26	0.40	0.00	16	13-16	26	0.30	0.00	16	13-16	26	0.58	0.00
	17-20	62	–	0.00		17-20	62	–	0.00		17-20	62	–	0.04
	21-24	34	–	0.00		21-24	34	–	0.00		21-24	34	–	0.00
	25-28	30	–	0.00		25-28	30	–	0.00		25-28	30	–	0.00
	29-32	16	–	0.00		29-32	16	–	0.00		29-32	16	–	0.00
	33-44	6	–	0.00		33-44	6	–	0.00		33-44	6	–	0.00
	45-64	14	–	0.00		45-64	14	–	0.00		45-64	14	–	0.00
18	13-16	26	0.53	–	18	13-16	26	0.36	–	18	13-16	26	0.50	–
	17-20	62	0.50	0.15		17-20	62	0.00	0.29		17-20	62	0.00	0.25
	21-24	34	–	0.00		21-24	34	–	0.00		21-24	34	–	0.04
	25-28	30	–	0.00		25-28	30	–	0.00		25-28	30	–	0.00
	29-32	16	–	0.00		29-32	16	–	0.00		29-32	16	–	0.00
	33-44	6	–	0.00		33-44	6	–	0.00		33-44	6	–	0.00
	45-64	14	–	0.00		45-64	14	–	0.00		45-64	14	–	0.00

Continued on next page

Table 12 – continued from previous page

Threshold	Age Band	N	FPR	FNR	Threshold	Age Band	N	FPR	FNR	Threshold	Age Band	N	FPR	FNR
21	13-16	26	0.07	–	21	13-16	26	0.00	–	21	13-16	26	0.00	–
	17-20	62	0.47	–		17-20	62	0.19	–		17-20	62	0.30	–
	21-24	34	–	0.23		21-24	34	–	0.12		21-24	34	–	0.19
	25-28	30	–	0.04		25-28	30	–	0.13		25-28	30	–	0.00
	29-32	16	–	0.00		29-32	16	–	0.00		29-32	16	–	0.00
	33-44	6	–	0.00		33-44	6	–	0.00		33-44	6	–	0.00
	45-64	14	–	0.00		45-64	14	–	0.00		45-64	14	–	0.00
25	13-16	26	0.00	–	25	13-16	26	0.00	–	25	13-16	26	0.00	–
	17-20	62	0.09	–		17-20	62	0.09	–		17-20	62	0.07	–
	21-24	34	0.23	–		21-24	34	0.38	–		21-24	34	0.38	–
	25-28	30	–	0.32		25-28	30	–	0.22		25-28	30	–	0.25
	29-32	16	–	0.00		29-32	16	–	0.00		29-32	16	–	0.00
	33-44	6	–	0.00		33-44	6	–	0.00		33-44	6	–	0.00
	45-64	14	–	0.00		45-64	14	–	0.00		45-64	14	–	0.00
30	13-16	26	0.00	–	30	13-16	26	0.00	–	30	13-16	26	0.00	–
	17-20	62	0.00	–		17-20	62	0.00	–		17-20	62	0.02	–
	21-24	34	0.00	–		21-24	34	0.00	–		21-24	34	0.00	–
	25-28	30	0.00	–		25-28	30	0.04	–		25-28	30	0.17	–
	29-32	16	0.00	0.33		29-32	16	0.38	0.50		29-32	16	0.00	0.50
	33-44	6	–	0.00		33-44	6	–	0.00		33-44	6	–	0.00
	45-64	14	–	0.00		45-64	14	–	0.00		45-64	14	–	0.00
	Glasses					**KN95**					**Groucho**			
13	13-16	26	–	0.00	13	13-16	26	–	0.00	13	13-16	26	–	0.00
	17-20	62	–	0.00		17-20	62	–	0.04		17-20	62	–	0.00
	21-24	34	–	0.00		21-24	34	–	0.00		21-24	34	–	0.00
	25-28	30	–	0.00		25-28	30	–	0.00		25-28	30	–	0.00
	29-32	16	–	0.00		29-32	16	–	0.00		29-32	16	–	0.00
	33-44	6	–	0.00		33-44	6	–	0.00		33-44	6	–	0.00
	45-64	14	–	0.00		45-64	14	–	0.00		45-64	14	–	0.00
16	13-16	26	0.46	0.00	16	13-16	26	0.50	0.00	16	13-16	26	0.46	0.00
	17-20	62	–	0.00		17-20	62	–	0.11		17-20	62	–	0.02
	21-24	34	–	0.00		21-24	34	–	0.03		21-24	34	–	0.00
	25-28	30	–	0.00		25-28	30	–	0.00		25-28	30	–	0.00
	29-32	16	–	0.00		29-32	16	–	0.00		29-32	16	–	0.00
	33-44	6	–	0.00		33-44	6	–	0.00		33-44	6	–	0.00
	45-64	14	–	0.00		45-64	14	–	0.00		45-64	14	–	0.00
18	13-16	26	0.24	–	18	13-16	26	0.41	–	18	13-16	26	0.24	–
	17-20	62	0.00	0.10		17-20	62	0.00	0.28		17-20	62	0.00	0.39
	21-24	34	–	0.00		21-24	34	–	0.06		21-24	34	–	0.10
	25-28	30	–	0.00		25-28	30	–	0.00		25-28	30	–	0.13
	29-32	16	–	0.00		29-32	16	–	0.00		29-32	16	–	0.00
	33-44	6	–	0.00		33-44	6	–	0.00		33-44	6	–	0.00
	45-64	14	–	0.00		45-64	14	–	0.00		45-64	14	–	0.00
21	13-16	26	0.08	–	21	13-16	26	0.14	–	21	13-16	26	0.10	–
	17-20	62	0.26	–		17-20	62	0.33	–		17-20	62	0.07	–
	21-24	34	–	0.48		21-24	34	–	0.55		21-24	34	–	0.52
	25-28	30	–	0.23		25-28	30	–	0.10		25-28	30	–	0.53
	29-32	16	–	0.00		29-32	16	–	0.00		29-32	16	–	0.17
	33-44	6	–	0.00		33-44	6	–	0.00		33-44	6	–	0.00
	45-64	14	–	0.00		45-64	14	–	0.00		45-64	14	–	0.00
25	13-16	26	0.00	–	25	13-16	26	0.05	–	25	13-16	26	0.05	–
	17-20	62	0.06	–		17-20	62	0.11	–		17-20	62	0.02	–
	21-24	34	0.30	–		21-24	34	0.15	–		21-24	34	0.14	–
	25-28	30	–	0.37		25-28	30	–	0.53		25-28	30	–	0.60
	29-32	16	–	0.00		29-32	16	–	0.19		29-32	16	–	0.42
	33-44	6	–	0.00		33-44	6	–	0.00		33-44	6	–	0.25
	45-64	14	–	0.00		45-64	14	–	0.00		45-64	14	–	0.00
30	13-16	26	0.00	–	30	13-16	26	0.00	–	30	13-16	26	0.00	–
	17-20	62	0.00	–		17-20	62	0.00	–		17-20	62	0.00	–
	21-24	34	0.00	–		21-24	34	0.00	–		21-24	34	0.00	–
	25-28	30	0.23	–		25-28	30	0.13	–		25-28	30	0.03	–
	29-32	16	0.50	0.25		29-32	16	0.12	0.62		29-32	16	0.00	0.83
	33-44	6	–	0.00		33-44	6	–	0.40		33-44	6	–	0.50
	45-64	14	–	0.00		45-64	14	–	0.00		45-64	14	–	0.00
	Eyepatch					**Wig**					**Costume mask**			
13	13-16	26	–	0.00	13	13-16	26	–	0.00	13	13-16	26	–	0.00
	17-20	62	–	0.00		17-20	62	–	0.02		17-20	62	–	0.00
	21-24	34	–	0.00		21-24	34	–	0.00		21-24	34	–	0.00
	25-28	30	–	0.00		25-28	30	–	0.00		25-28	30	–	0.00
	29-32	16	–	0.00		29-32	16	–	0.00		29-32	16	–	0.00
	33-44	6	–	0.00		33-44	6	–	0.00		33-44	6	–	0.00

Continued on next page

Table 12 – continued from previous page

Threshold	Age Band	N	FPR	FNR	Threshold	Age Band	N	FPR	FNR	Threshold	Age Band	N	FPR	FNR
	45-64	14	–	0.00		45-64	14	–	0.00		45-64	14	–	0.00
16	13-16	26	0.50	0.00	16	13-16	26	0.45	0.00	16	13-16	26	1.00	0.00
	17-20	62	–	0.00		17-20	62	–	0.03		17-20	62	–	0.00
	21-24	34	–	0.00		21-24	34	–	0.00		21-24	34	–	0.00
	25-28	30	–	0.00		25-28	30	–	0.00		25-28	30	–	0.00
	29-32	16	–	0.00		29-32	16	–	0.00		29-32	16	–	0.00
	33-44	6	–	0.00		33-44	6	–	0.00		33-44	6	–	0.00
	45-64	14	–	0.00		45-64	14	–	0.00		45-64	14	–	0.00
18	13-16	26	0.25	–	18	13-16	26	0.35	–	18	13-16	26	1.00	–
	17-20	62	0.50	0.14		17-20	62	0.50	0.16		17-20	62	1.00	0.00
	21-24	34	–	0.00		21-24	34	–	0.00		21-24	34	–	0.00
	25-28	30	–	0.04		25-28	30	–	0.00		25-28	30	–	0.00
	29-32	16	–	0.00		29-32	16	–	0.00		29-32	16	–	0.00
	33-44	6	–	0.00		33-44	6	–	0.00		33-44	6	–	0.00
	45-64	14	–	0.00		45-64	14	–	0.00		45-64	14	–	0.00
21	13-16	26	0.10	–	21	13-16	26	0.04	–	21	13-16	26	1.00	–
	17-20	62	0.27	–		17-20	62	0.20	–		17-20	62	1.00	–
	21-24	34	–	0.41		21-24	34	–	0.32		21-24	34	–	0.00
	25-28	30	–	0.25		25-28	30	–	0.04		25-28	30	–	0.00
	29-32	16	–	0.00		29-32	16	–	0.00		29-32	16	–	0.00
	33-44	6	–	0.00		33-44	6	–	0.00		33-44	6	–	0.00
	45-64	14	–	0.00		45-64	14	–	0.00		45-64	14	–	0.00
25	13-16	26	0.05	–	25	13-16	26	0.00	–	25	13-16	26	1.00	–
	17-20	62	0.03	–		17-20	62	0.08	–		17-20	62	1.00	–
	21-24	34	0.26	–		21-24	34	0.32	–		21-24	34	1.00	–
	25-28	30	–	0.43		25-28	30	–	0.36		25-28	30	–	0.00
	29-32	16	–	0.07		29-32	16	–	0.06		29-32	16	–	0.00
	33-44	6	–	0.00		33-44	6	–	0.00		33-44	6	–	0.00
	45-64	14	–	0.00		45-64	14	–	0.00		45-64	14	–	0.00
30	13-16	26	0.00	–	30	13-16	26	0.00	–	30	13-16	26	1.00	–
	17-20	62	0.00	–		17-20	62	0.00	–		17-20	62	1.00	–
	21-24	34	0.00	–		21-24	34	0.03	–		21-24	34	1.00	–
	25-28	30	0.14	–		25-28	30	0.04	–		25-28	30	1.00	–
	29-32	16	0.00	0.12		29-32	16	0.38	0.38		29-32	16	1.00	0.00
	33-44	6	–	0.17		33-44	6	–	0.33		33-44	6	–	0.00
	45-64	14	–	0.00		45-64	14	–	0.00		45-64	14	–	0.00

M FNR and FPR by age band, FACETEC

Table 13: Age-prediction performance by age band and test condition for FACETEC. FPR and FNR are reported as proportions.

Threshold	Age Band	*N*	FPR	FNR	Threshold	Age Band	*N*	FPR	FNR	Threshold	Age Band	*N*	FPR	FNR
	Baseline					**Open-mouth**					**Frown**			
13	13-16	17	–	0.12	13	13-16	9	–	0.11	13	13-16	16	–	0.25
	17-20	1	–	0.00		17-20	0	–	–		17-20	0	–	–
	21-24	0	–	–		21-24	0	–	–		21-24	0	–	–
	25-28	0	–	–		25-28	0	–	–		25-28	0	–	–
	29-32	0	–	–		29-32	0	–	–		29-32	0	–	–
	33-44	0	–	–		33-44	0	–	–		33-44	0	–	–
	45-64	1	–	1.00		45-64	0	–	–		45-64	0	–	–
16	13-16	15	0.11	0.33	16	13-16	9	0.25	0.40	16	13-16	17	0.14	0.40
	17-20	22	–	0.09		17-20	11	–	0.00		17-20	26	–	0.08
	21-24	0	–	–		21-24	0	–	–		21-24	0	–	–
	25-28	0	–	–		25-28	0	–	–		25-28	0	–	–
	29-32	0	–	–		29-32	0	–	–		29-32	0	–	–
	33-44	0	–	–		33-44	0	–	–		33-44	0	–	–
	45-64	0	–	–		45-64	0	–	–		45-64	0	–	–
18	13-16	19	0.05	–	18	13-16	10	0.10	–	18	13-16	18	0.11	–
	17-20	41	0.00	0.68		17-20	14	–	0.50		17-20	28	0.00	0.59
	21-24	21	–	0.19		21-24	11	–	0.55		21-24	16	–	0.31
	25-28	22	–	0.23		25-28	2	–	0.50		25-28	17	–	0.35
	29-32	8	–	0.00		29-32	5	–	0.00		29-32	7	–	0.00
	33-44	0	–	–		33-44	1	–	0.00		33-44	0	–	–
	45-64	10	–	0.00		45-64	2	–	0.00		45-64	6	–	0.00
21	13-16	18	0.00	–	21	13-16	9	0.00	–	21	13-16	16	0.06	–
	17-20	43	0.26	–		17-20	14	0.14	–		17-20	30	0.20	–
	21-24	26	–	0.62		21-24	13	–	0.77		21-24	15	–	0.67
	25-28	24	–	0.42		25-28	3	–	0.67		25-28	15	–	0.73
	29-32	14	–	0.14		29-32	6	–	0.00		29-32	8	–	0.00
	33-44	0	–	–		33-44	0	–	–		33-44	0	–	–
	45-64	12	–	0.00		45-64	1	–	0.00		45-64	6	–	0.00
25	13-16	2	0.00	–	25	13-16	0	–	–	25	13-16	1	0.00	–
	17-20	40	0.03	–		17-20	13	0.08	–		17-20	34	0.15	–
	21-24	25	0.04	–		21-24	13	0.15	–		21-24	13	0.23	–
	25-28	19	–	0.68		25-28	4	–	1.00		25-28	17	–	0.71
	29-32	10	–	0.30		29-32	4	–	0.25		29-32	9	–	0.22
	33-44	0	–	–		33-44	0	–	–		33-44	3	–	0.00
	45-64	12	–	0.00		45-64	2	–	0.00		45-64	8	–	0.00
30	13-16	2	0.00	–	30	13-16	0	–	–	30	13-16	0	–	–
	17-20	12	0.00	–		17-20	1	0.00	–		17-20	4	0.00	–
	21-24	25	0.00	–		21-24	13	0.00	–		21-24	14	0.00	–
	25-28	23	0.09	–		25-28	2	0.00	–		25-28	17	0.12	–
	29-32	13	0.14	0.83		29-32	6	0.00	0.50		29-32	9	0.20	0.75
	33-44	1	–	0.00		33-44	0	–	–		33-44	0	–	–
	45-64	10	–	0.10		45-64	2	–	0.00		45-64	8	–	0.12
	Smile					**Mustache**					**Beard**			
13	13-16	3	–	0.33	13	13-16	4	–	0.00	13	13-16	0	–	–
	17-20	0	–	–		17-20	0	–	–		17-20	0	–	–
	21-24	0	–	–		21-24	0	–	–		21-24	0	–	–
	25-28	0	–	–		25-28	0	–	–		25-28	0	–	–
	29-32	0	–	–		29-32	0	–	–		29-32	0	–	–
	33-44	0	–	–		33-44	0	–	–		33-44	0	–	–
	45-64	0	–	–		45-64	0	–	–		45-64	0	–	–
16	13-16	1	–	1.00	16	13-16	9	0.50	0.00	16	13-16	0	–	–
	17-20	8	–	0.25		17-20	8	–	0.00		17-20	2	–	0.00
	21-24	0	–	–		21-24	0	–	–		21-24	0	–	–
	25-28	0	–	–		25-28	0	–	–		25-28	0	–	–
	29-32	0	–	–		29-32	0	–	–		29-32	0	–	–
	33-44	0	–	–		33-44	0	–	–		33-44	0	–	–
	45-64	0	–	–		45-64	0	–	–		45-64	0	–	–
18	13-16	3	0.00	–	18	13-16	5	0.60	–	18	13-16	0	–	–
	17-20	10	0.00	0.50		17-20	10	–	0.00		17-20	2	–	0.00
	21-24	3	–	0.67		21-24	8	–	0.62		21-24	0	–	–
	25-28	0	–	–		25-28	8	–	0.25		25-28	8	–	0.00
	29-32	2	–	0.00		29-32	1	–	0.00		29-32	2	–	0.00
	33-44	1	–	0.00		33-44	0	–	–		33-44	0	–	–
	45-64	2	–	0.00		45-64	6	–	0.00		45-64	0	–	–

Continued on next page

Table 13 – continued from previous page

Threshold	Age Band	N	FPR	FNR	Threshold	Age Band	N	FPR	FNR	Threshold	Age Band	N	FPR	FNR
21	13-16	3	0.00	–	21	13-16	4	0.25	–	21	13-16	0	–	–
	17-20	10	0.20	–		17-20	10	0.40	–		17-20	1	1.00	–
	21-24	5	–	0.60		21-24	8	–	0.75		21-24	0	–	–
	25-28	0	–	–		25-28	6	–	0.50		25-28	7	–	0.43
	29-32	2	–	0.00		29-32	1	–	0.00		29-32	1	–	0.00
	33-44	0	–	–		33-44	0	–	–		33-44	0	–	–
	45-64	2	–	0.00		45-64	2	–	0.00		45-64	0	–	–
25	13-16	0	–	–	25	13-16	0	–	–	25	13-16	0	–	–
	17-20	10	0.10	–		17-20	9	0.11	–		17-20	1	0.00	–
	21-24	3	0.00	–		21-24	7	0.14	–		21-24	0	–	–
	25-28	0	–	–		25-28	9	–	1.00		25-28	7	–	0.71
	29-32	2	–	0.50		29-32	2	–	0.00		29-32	2	–	1.00
	33-44	0	–	–		33-44	0	–	–		33-44	0	–	–
	45-64	1	–	0.00		45-64	3	–	0.00		45-64	0	–	–
30	13-16	0	–	–	30	13-16	0	–	–	30	13-16	0	–	–
	17-20	0	–	–		17-20	1	0.00	–		17-20	0	–	–
	21-24	4	0.00	–		21-24	7	0.00	–		21-24	0	–	–
	25-28	1	0.00	–		25-28	9	0.00	–		25-28	7	0.00	–
	29-32	1	0.00	–		29-32	0	–	–		29-32	2	–	1.00
	33-44	1	–	1.00		33-44	1	–	0.00		33-44	0	–	–
	45-64	1	–	0.00		45-64	4	–	0.00		45-64	0	–	–
Glasses					**KN95**					**Groucho**				
13	13-16	10	–	0.10	13	13-16	0	–	–	13	13-16	0	–	–
	17-20	0	–	–		17-20	0	–	–		17-20	0	–	–
	21-24	0	–	–		21-24	0	–	–		21-24	0	–	–
	25-28	0	–	–		25-28	0	–	–		25-28	0	–	–
	29-32	0	–	–		29-32	0	–	–		29-32	0	–	–
	33-44	0	–	–		33-44	0	–	–		33-44	0	–	–
	45-64	0	–	–		45-64	0	–	–		45-64	0	–	–
16	13-16	10	1.00	0.12	16	13-16	0	–	–	16	13-16	0	–	–
	17-20	21	–	0.05		17-20	0	–	–		17-20	0	–	–
	21-24	0	–	–		21-24	0	–	–		21-24	0	–	–
	25-28	0	–	–		25-28	0	–	–		25-28	0	–	–
	29-32	0	–	–		29-32	0	–	–		29-32	0	–	–
	33-44	0	–	–		33-44	0	–	–		33-44	0	–	–
	45-64	0	–	–		45-64	0	–	–		45-64	0	–	–
18	13-16	7	0.00	–	18	13-16	0	–	–	18	13-16	0	–	–
	17-20	39	–	0.51		17-20	0	–	–		17-20	0	–	–
	21-24	18	–	0.17		21-24	0	–	–		21-24	0	–	–
	25-28	16	–	0.25		25-28	0	–	–		25-28	0	–	–
	29-32	6	–	0.33		29-32	0	–	–		29-32	0	–	–
	33-44	1	–	0.00		33-44	0	–	–		33-44	0	–	–
	45-64	8	–	0.00		45-64	0	–	–		45-64	0	–	–
21	13-16	9	0.00	–	21	13-16	0	–	–	21	13-16	0	–	–
	17-20	32	0.28	–		17-20	0	–	–		17-20	0	–	–
	21-24	15	–	0.20		21-24	0	–	–		21-24	0	–	–
	25-28	17	–	0.41		25-28	0	–	–		25-28	0	–	–
	29-32	8	–	0.25		29-32	0	–	–		29-32	0	–	–
	33-44	2	–	0.50		33-44	0	–	–		33-44	0	–	–
	45-64	9	–	0.00		45-64	0	–	–		45-64	0	–	–
25	13-16	2	0.00	–	25	13-16	0	–	–	25	13-16	0	–	–
	17-20	32	0.22	–		17-20	0	–	–		17-20	0	–	–
	21-24	17	0.18	–		21-24	0	–	–		21-24	0	–	–
	25-28	17	–	0.76		25-28	0	–	–		25-28	0	–	–
	29-32	5	–	0.20		29-32	0	–	–		29-32	0	–	–
	33-44	0	–	–		33-44	0	–	–		33-44	0	–	–
	45-64	9	–	0.00		45-64	0	–	–		45-64	0	–	–
30	13-16	0	–	–	30	13-16	0	–	–	30	13-16	0	–	–
	17-20	8	0.12	–		17-20	0	–	–		17-20	0	–	–
	21-24	16	0.06	–		21-24	0	–	–		21-24	0	–	–
	25-28	14	0.21	–		25-28	0	–	–		25-28	0	–	–
	29-32	4	0.33	1.00		29-32	0	–	–		29-32	0	–	–
	33-44	1	–	1.00		33-44	0	–	–		33-44	0	–	–
	45-64	7	–	0.14		45-64	0	–	–		45-64	0	–	–
Eyepatch					**Wig**					**Costume mask**				
13	13-16	0	–	–	13	13-16	16	–	0.19	13	13-16	0	–	–
	17-20	0	–	–		17-20	0	–	–		17-20	0	–	–
	21-24	0	–	–		21-24	0	–	–		21-24	0	–	–
	25-28	0	–	–		25-28	0	–	–		25-28	0	–	–
	29-32	0	–	–		29-32	0	–	–		29-32	0	–	–
	33-44	0	–	–		33-44	0	–	–		33-44	0	–	–

Continued on next page

Table 13 – continued from previous page

Threshold	Age Band	N	FPR	FNR	Threshold	Age Band	N	FPR	FNR	Threshold	Age Band	N	FPR	FNR
	45-64	0	–	–		45-64	0	–	–		45-64	0	–	–
16	13-16	0	–	–	16	13-16	17	0.10	0.43	16	13-16	0	–	–
	17-20	0	–	–		17-20	32	–	0.06		17-20	0	–	–
	21-24	0	–	–		21-24	0	–	–		21-24	0	–	–
	25-28	0	–	–		25-28	0	–	–		25-28	0	–	–
	29-32	0	–	–		29-32	0	–	–		29-32	0	–	–
	33-44	0	–	–		33-44	0	–	–		33-44	0	–	–
	45-64	0	–	–		45-64	0	–	–		45-64	0	–	–
18	13-16	0	–	–	18	13-16	20	0.00	–	18	13-16	0	–	–
	17-20	0	–	–		17-20	37	–	0.73		17-20	0	–	–
	21-24	0	–	–		21-24	26	–	0.77		21-24	0	–	–
	25-28	0	–	–		25-28	20	–	0.80		25-28	0	–	–
	29-32	0	–	–		29-32	7	–	0.00		29-32	0	–	–
	33-44	0	–	–		33-44	1	–	1.00		33-44	0	–	–
	45-64	0	–	–		45-64	10	–	0.00		45-64	0	–	–
21	13-16	0	–	–	21	13-16	16	0.00	–	21	13-16	0	–	–
	17-20	0	–	–		17-20	37	0.08	–		17-20	0	–	–
	21-24	0	–	–		21-24	22	–	0.86		21-24	0	–	–
	25-28	0	–	–		25-28	20	–	0.80		25-28	0	–	–
	29-32	0	–	–		29-32	11	–	0.09		29-32	0	–	–
	33-44	0	–	–		33-44	0	–	–		33-44	0	–	–
	45-64	0	–	–		45-64	10	–	0.00		45-64	0	–	–
25	13-16	0	–	–	25	13-16	2	0.00	–	25	13-16	0	–	–
	17-20	0	–	–		17-20	38	0.00	–		17-20	0	–	–
	21-24	0	–	–		21-24	26	0.12	–		21-24	0	–	–
	25-28	0	–	–		25-28	17	–	0.82		25-28	0	–	–
	29-32	0	–	–		29-32	10	–	0.30		29-32	0	–	–
	33-44	0	–	–		33-44	0	–	–		33-44	0	–	–
	45-64	0	–	–		45-64	7	–	0.00		45-64	0	–	–
30	13-16	0	–	–	30	13-16	0	–	–	30	13-16	0	–	–
	17-20	0	–	–		17-20	7	0.00	–		17-20	0	–	–
	21-24	0	–	–		21-24	27	0.00	–		21-24	0	–	–
	25-28	0	–	–		25-28	19	0.05	–		25-28	0	–	–
	29-32	0	–	–		29-32	9	0.00	0.75		29-32	0	–	–
	33-44	0	–	–		33-44	1	–	1.00		33-44	0	–	–
	45-64	0	–	–		45-64	10	–	0.00		45-64	0	–	–

N FNR and FPR by age band, PRIVATELY

Table 14: Age-prediction performance by age band and test condition for PRIVATELY. FPR and FNR are reported as proportions.

Threshold	Age Band	*N*	FPR	FNR	Threshold	Age Band	*N*	FPR	FNR	Threshold	Age Band	*N*	FPR	FNR
	Baseline					**Open-mouth**					**Frown**			
21	13-16	13	0.23	–	21	13-16	9	0.44	–	21	13-16	9	0.22	–
	17-20	31	0.45	–		17-20	21	0.62	–		17-20	21	0.62	–
	21-24	17	–	0.41		21-24	12	–	0.33		21-24	11	–	0.45
	25-28	15	–	0.20		25-28	11	–	0.36		25-28	12	–	0.25
	29-32	8	–	0.12		29-32	6	–	0.00		29-32	6	–	0.00
	33-44	3	–	0.00		33-44	2	–	0.00		33-44	2	–	0.00
	45-64	7	–	0.00		45-64	5	–	0.00		45-64	5	–	0.00
25	13-16	13	0.08	–	25	13-16	9	0.00	–	25	13-16	9	0.22	–
	17-20	31	0.06	–		17-20	21	0.10	–		17-20	21	0.24	–
	21-24	17	0.18	–		21-24	12	0.00	–		21-24	11	0.27	–
	25-28	15	–	0.67		25-28	11	–	0.73		25-28	12	–	0.58
	29-32	8	–	0.50		29-32	6	–	0.67		29-32	6	–	0.33
	33-44	3	–	0.00		33-44	2	–	0.00		33-44	2	–	0.00
	45-64	7	–	0.00		45-64	5	–	0.00		45-64	5	–	0.00
30	13-16	13	0.00	–	30	13-16	9	0.00	–	30	13-16	9	0.00	–
	17-20	31	0.00	–		17-20	21	0.00	–		17-20	21	0.00	–
	21-24	17	0.00	–		21-24	12	0.00	–		21-24	11	0.09	–
	25-28	15	0.13	–		25-28	11	0.09	–		25-28	12	0.17	–
	29-32	8	0.00	1.00		29-32	6	0.00	1.00		29-32	6	0.33	1.00
	33-44	3	–	0.67		33-44	2	–	0.50		33-44	2	–	0.50
	45-64	7	–	0.00		45-64	5	–	0.00		45-64	5	–	0.00
	Smile					**Mustache**					**Beard**			
21	13-16	9	0.11	–	21	13-16	13	0.31	–	21	13-16	13	0.85	–
	17-20	22	0.36	–		17-20	27	0.59	–		17-20	27	0.70	–
	21-24	13	–	0.46		21-24	13	–	0.54		21-24	13	–	0.31
	25-28	13	–	0.23		25-28	12	–	0.25		25-28	12	–	0.17
	29-32	5	–	0.00		29-32	7	–	0.14		29-32	7	–	0.00
	33-44	3	–	0.00		33-44	2	–	0.00		33-44	1	–	0.00
	45-64	5	–	0.00		45-64	7	–	0.00		45-64	7	–	0.00
25	13-16	9	0.00	–	25	13-16	13	0.08	–	25	13-16	13	0.00	–
	17-20	22	0.05	–		17-20	27	0.19	–		17-20	27	0.11	–
	21-24	13	0.08	–		21-24	13	0.08	–		21-24	13	0.23	–
	25-28	13	–	0.85		25-28	12	–	0.83		25-28	12	–	0.67
	29-32	5	–	0.20		29-32	7	–	0.43		29-32	7	–	0.43
	33-44	3	–	0.33		33-44	2	–	0.00		33-44	1	–	0.00
	45-64	5	–	0.00		45-64	7	–	0.00		45-64	7	–	0.00
30	13-16	9	0.00	–	30	13-16	13	0.00	–	30	13-16	13	0.00	–
	17-20	22	0.00	–		17-20	27	0.04	–		17-20	27	0.04	–
	21-24	13	0.08	–		21-24	13	0.00	–		21-24	13	0.00	–
	25-28	13	0.00	–		25-28	12	0.00	–		25-28	12	0.00	–
	29-32	5	0.00	0.50		29-32	7	0.00	1.00		29-32	7	0.00	0.67
	33-44	3	–	0.33		33-44	2	–	0.50		33-44	1	–	0.00
	45-64	5	–	0.00		45-64	7	–	0.00		45-64	7	–	0.00
	Glasses					**KN95**					**Groucho**			
21	13-16	13	0.31	–	21	13-16	13	0.38	–	21	13-16	12	0.42	–
	17-20	31	0.55	–		17-20	30	0.40	–		17-20	29	0.62	–
	21-24	17	–	0.06		21-24	17	–	0.71		21-24	15	–	0.33
	25-28	15	–	0.33		25-28	15	–	0.60		25-28	15	–	0.20
	29-32	8	–	0.00		29-32	8	–	0.38		29-32	6	–	0.17
	33-44	3	–	0.00		33-44	3	–	0.33		33-44	3	–	0.00
	45-64	7	–	0.00		45-64	7	–	0.00		45-64	6	–	0.00
25	13-16	13	0.00	–	25	13-16	13	0.00	–	25	13-16	12	0.00	–
	17-20	31	0.23	–		17-20	30	0.00	–		17-20	29	0.03	–
	21-24	17	0.47	–		21-24	17	0.06	–		21-24	15	0.13	–
	25-28	15	–	0.60		25-28	15	–	0.93		25-28	15	–	0.80
	29-32	8	–	0.25		29-32	8	–	0.75		29-32	6	–	1.00
	33-44	3	–	0.33		33-44	3	–	0.67		33-44	3	–	0.67
	45-64	7	–	0.00		45-64	7	–	0.00		45-64	6	–	0.00
30	13-16	13	0.00	–	30	13-16	13	0.00	–	30	13-16	12	0.00	–
	17-20	31	0.06	–		17-20	30	0.00	–		17-20	29	0.00	–
	21-24	17	0.00	–		21-24	17	0.00	–		21-24	15	0.00	–
	25-28	15	0.13	–		25-28	15	0.00	–		25-28	15	0.00	–
	29-32	8	0.00	0.50		29-32	8	0.00	1.00		29-32	6	0.00	1.00

Continued on next page

Table 14 – continued from previous page

Threshold	Age Band	*N*	FPR	FNR	Threshold	Age Band	*N*	FPR	FNR	Threshold	Age Band	*N*	FPR	FNR
	33-44	3	–	0.33		33-44	3	–	0.67		33-44	3	–	1.00
	45-64	7	–	0.00		45-64	7	–	0.29		45-64	6	–	0.00
	Eyepatch					**Wig**					**Costume mask**			
21	13-16	12	0.00	–	21	13-16	13	0.08	–	21	13-16	12	1.00	–
	17-20	31	0.26	–		17-20	30	0.20	–		17-20	29	1.00	–
	21-24	17	–	0.53		21-24	17	–	0.65		21-24	14	–	0.00
	25-28	14	–	0.43		25-28	14	–	0.57		25-28	14	–	0.00
	29-32	8	–	0.00		29-32	8	–	0.00		29-32	6	–	0.00
	33-44	3	–	0.00		33-44	3	–	0.00		33-44	2	–	0.00
	45-64	7	–	0.00		45-64	7	–	0.00		45-64	7	–	0.00
25	13-16	12	0.00	–	25	13-16	13	0.00	–	25	13-16	12	1.00	–
	17-20	31	0.06	–		17-20	30	0.07	–		17-20	29	1.00	–
	21-24	17	0.12	–		21-24	17	0.12	–		21-24	14	1.00	–
	25-28	14	–	0.57		25-28	14	–	0.79		25-28	14	–	0.00
	29-32	8	–	0.12		29-32	8	–	0.50		29-32	6	–	0.00
	33-44	3	–	0.00		33-44	3	–	0.67		33-44	2	–	0.00
	45-64	7	–	0.00		45-64	7	–	0.00		45-64	7	–	0.00
30	13-16	12	0.00	–	30	13-16	13	0.00	–	30	13-16	12	1.00	–
	17-20	31	0.00	–		17-20	30	0.00	–		17-20	29	1.00	–
	21-24	17	0.00	–		21-24	17	0.00	–		21-24	14	1.00	–
	25-28	14	0.00	–		25-28	14	0.07	–		25-28	14	1.00	–
	29-32	8	0.00	0.75		29-32	8	0.25	0.75		29-32	6	1.00	0.00
	33-44	3	–	0.33		33-44	3	–	0.67		33-44	2	–	0.00
	45-64	7	–	0.00		45-64	7	–	0.00		45-64	7	–	0.00

O Platform age verification provider attributions

Table 15: Provider attribution for platforms subject to Australia's 2026 social-media age restrictions.

Platform	Provider	Policies	APK
Snapchat	k-ID	✓	×
Instagram	Yoti	✓	×
TikTok	Yoti	✓	×
Facebook	Yoti	✓	×
Reddit	Persona	✓	✓
Threads	Yoti	✓	×
Twitch	k-ID	✓	×
X	Persona	✓	✓
YouTube	Private ID	✓	×
Kick	k-ID	✓	×

Note: A check indicates direct attribution in policy documents or APK analysis. Static analysis identified every age-verification path, but not providers selected behind hosted flows. k-ID is a gateway to multiple models and lists Privately as its sole on-device provider.

P Liveness FNR by test condition and gender identity, Yoti

Table 16: We stratify participants by self-reported gender identity, reporting FNR on all test conditions for Yoti.

Condition	Gender	*N*	FNR
Baseline	Female	87	0.51
	Male	54	0.39
	Non-binary	4	0.5
Beard	Female	56	0.96
	Male	36	0.86
Costume mask	Female	48	1
	Male	28	1
Eyepatch	Female	60	1
	Male	32	1
Frown	Female	63	0.48
	Male	30	0.37
Open-mouth	Female	55	0.71
	Male	28	0.5
Glasses	Female	80	0.63
	Male	49	0.45
	Non-binary	4	0.5
Groucho	Female	56	1
	Male	22	1
KN95	Female	61	1
	Male	32	1
Smile	Female	54	0.91
	Male	22	0.82
Mustache	Female	66	0.86
	Male	44	0.61
	Non-binary	3	0.67
Wig	Female	84	0.46
	Male	48	0.33
	Non-binary	4	0

Q Liveness FNR by test condition and gender identity, FACETEC

Table 17: We stratify participants by self-reported gender identity, reporting FNR on all test conditions for FACETEC.

Condition	Gender	*N*	FNR
Baseline	Female	93	0.53
	Male	60	0.35
	Non-binary	6	0.5
Beard	Female	62	0.97
	Male	42	0.88
	Non-binary	4	1
Costume mask	Female	54	1
	Male	34	1
	Non-binary	4	1
Eyepatch	Female	66	1
	Male	38	1
	Non-binary	4	1
Frown	Female	67	0.45
	Male	34	0.32
	Non-binary	4	0.75
Open-mouth	Female	59	0.71
	Male	32	0.53
	Non-binary	4	1
Glasses	Female	86	0.65
	Male	55	0.44
	Non-binary	6	0.67
Groucho	Female	62	1
	Male	26	1
	Non-binary	4	1
KN95	Female	67	1
	Male	38	1
	Non-binary	4	1
Smile	Female	58	0.91
	Male	26	0.85
	Non-binary	4	1
Mustache	Female	72	0.88
	Male	48	0.63
	Non-binary	5	0.8
Wig	Female	90	0.46
	Male	54	0.33
	Non-binary	6	0.33

R Age prediction FPR and FNR by condition and gender identity, Yoti

Table 18: Yoti FPR and FNR by condition, gender identity, and age threshold. Rates are proportions.

Baseline					Open-mouth					Frown				
Gender	Thr.	*N*	FPR	FNR	Gender	Thr.	*N*	FPR	FNR	Gender	Thr.	*N*	FPR	FNR
Female	13	102	–	0.01	Female	13	78	–	0.00	Female	13	76	–	0.00
	16	102	0.50	0.00		16	78	0.38	0.03		16	76	0.62	0.00
	18	102	0.39	0.04		18	78	0.44	0.07		18	76	0.44	0.02
	21	102	0.22	0.17		21	78	0.26	0.33		21	76	0.32	0.16
	25	102	0.13	0.25		25	78	0.07	0.46		25	76	0.12	0.25
	30	102	0.02	0.08		30	78	0.00	0.00		30	76	0.03	0.10
Male	13	76	–	0.00	Male	13	43	–	0.00	Male	13	43	–	0.00
	16	76	0.00	0.00		16	43	–	0.00		16	43	–	0.00
	18	76	0.25	0.06		18	43	1.00	0.12		18	43	1.00	0.05
	21	76	0.10	0.07		21	43	0.23	0.10		21	43	0.31	0.10
	25	76	0.11	0.00		25	43	0.14	0.00		25	43	0.15	0.00
	30	76	0.07	0.07		30	43	0.03	0.00		30	43	0.14	0.00
Non-binary	13	6	–	0.00	Non-binary	13	5	–	0.00	Non-binary	13	6	–	0.00
	16	6	–	0.00		16	5	–	0.00		16	6	–	0.00
	18	6	–	0.33		18	5	–	0.40		18	6	–	0.00
	21	6	0.00	–		21	5	0.00	–		21	6	0.00	–
	25	6	0.00	–		25	5	0.00	–		25	6	0.00	–
	30	6	0.00	–		30	5	0.00	–		30	6	0.00	–
Other	13	2	–	0.00	Other	13	2	–	0.00	Other	13	2	–	0.00
	16	2	–	0.00		16	2	–	0.00		16	2	–	0.00
	18	2	–	0.00		18	2	–	0.00		18	2	–	0.00
	21	2	–	0.00		21	2	–	0.00		21	2	–	0.00
	25	2	–	0.00		25	2	–	0.00		25	2	–	0.00
	30	2	–	0.00		30	2	–	0.00		30	2	–	0.00
Transgender	13	2	–	0.00	Transgender	13	2	–	0.00	Transgender	13	2	–	0.00
	16	2	–	0.00		16	2	–	0.00		16	2	–	0.00
	18	2	0.00	–		18	2	0.00	–		18	2	0.50	–
	21	2	0.00	–		21	2	0.00	–		21	2	0.00	–
	25	2	0.00	–		25	2	0.00	–		25	2	0.00	–
	30	2	0.00	–		30	2	0.00	–		30	2	0.00	–

Smile					Mustache					Beard				
Gender	Thr.	*N*	FPR	FNR	Gender	Thr.	*N*	FPR	FNR	Gender	Thr.	*N*	FPR	FNR
Female	13	81	–	0.00	Female	13	87	–	0.00	Female	13	84	–	0.00
	16	81	0.50	0.00		16	87	0.50	0.00		16	84	0.60	0.03
	18	81	0.38	0.01		18	87	0.39	0.13		18	84	0.44	0.09
	21	81	0.40	0.15		21	87	0.12	0.08		21	84	0.17	0.08
	25	81	0.14	0.30		25	87	0.10	0.20		25	84	0.07	0.25
	30	81	0.00	0.00		30	87	0.00	0.17		30	84	0.01	0.10
Male	13	48	–	0.00	Male	13	64	–	0.00	Male	13	66	–	0.00
	16	48	–	0.00		16	64	0.25	0.00		16	66	0.50	0.00
	18	48	1.00	0.09		18	64	0.12	0.07		18	66	0.50	0.10
	21	48	0.27	0.03		21	64	0.17	0.07		21	66	0.31	0.05
	25	48	0.09	0.04		25	64	0.25	0.00		25	66	0.26	0.00
	30	48	0.00	0.11		30	64	0.07	0.10		30	66	0.07	0.20
Non-binary	13	4	–	0.00	Non-binary	13	6	–	0.00	Non-binary	13	6	–	0.00
	16	4	–	0.00		16	6	–	0.00		16	6	–	0.00
	18	4	–	0.25		18	6	–	0.33		18	6	–	0.33
	21	4	0.25	–		21	6	0.00	–		21	6	0.00	–
	25	4	0.00	–		25	6	0.00	–		25	6	0.00	–
	30	4	0.00	–		30	6	0.00	–		30	6	0.00	–
Other	13	2	–	0.00	Other	13	2	–	0.00	Other	13	2	–	0.00
	16	2	–	0.00		16	2	–	0.00		16	2	–	0.00
	18	2	–	0.00		18	2	–	0.00		18	2	–	0.00
	21	2	–	0.00		21	2	–	0.00		21	2	–	0.00
	25	2	–	0.00		25	2	–	0.00		25	2	–	0.00
	30	2	–	0.00		30	2	–	0.00		30	2	–	0.00
Transgender	13	2	–	0.00	Transgender	13	2	–	0.00	Transgender	13	2	–	0.00
	16	2	–	0.00		16	2	–	0.00		16	2	–	0.00
	18	2	0.50	–		18	2	0.00	–		18	2	0.00	–
	21	2	0.00	–		21	2	0.00	–		21	2	0.00	–
	25	2	0.00	–		25	2	0.00	–		25	2	0.00	–
	30	2	0.00	–		30	2	0.00	–		30	2	0.00	–

Continued on next page

Yoti FPR/FNR by condition, gender identity, and threshold (continued)

Gender	Thr.	*N*	FPR	FNR	Gender	Thr.	*N*	FPR	FNR	Gender	Thr.	*N*	FPR	FNR
Glasses					**KN95**					**Groucho**				
Gender	Thr.	*N*	FPR	FNR	Gender	Thr.	*N*	FPR	FNR	Gender	Thr.	*N*	FPR	FNR
Female	13	101	–	0.00	Female	13	95	–	0.00	Female	13	94	–	0.00
	16	101	0.60	0.00		16	95	0.38	0.04		16	94	0.70	0.01
	18	101	0.33	0.00		18	95	0.44	0.09		18	94	0.28	0.21
	21	101	0.24	0.35		21	95	0.20	0.31		21	94	0.12	0.59
	25	101	0.12	0.31		25	95	0.08	0.42		25	94	0.08	0.72
	30	101	0.04	0.17		30	95	0.01	0.09		30	94	0.00	0.33
Male	13	76	–	0.00	Male	13	70	–	0.00	Male	13	61	–	0.00
	16	76	0.25	0.00		16	70	1.00	0.01		16	61	0.00	0.00
	18	76	0.00	0.06		18	70	0.33	0.12		18	61	0.00	0.17
	21	76	0.17	0.11		21	70	0.35	0.14		21	61	0.00	0.16
	25	76	0.14	0.03		25	70	0.17	0.20		25	61	0.03	0.12
	30	76	0.11	0.00		30	70	0.07	0.50		30	61	0.02	0.44
Non-binary	13	6	–	0.00	Non-binary	13	4	–	0.00	Non-binary	13	6	–	0.00
	16	6	–	0.00		16	4	–	0.00		16	6	–	0.00
	18	6	–	0.33		18	4	–	0.25		18	6	–	0.50
	21	6	0.17	–		21	4	0.75	–		21	6	0.00	–
	25	6	0.00	–		25	4	0.00	–		25	6	0.00	–
	30	6	0.00	–		30	4	0.00	–		30	6	0.00	–
Other	13	2	–	0.00	Other	13	2	–	0.00	Other	13	2	–	0.00
	16	2	–	0.00		16	2	–	0.00		16	2	–	0.00
	18	2	–	0.00		18	2	–	0.00		18	2	–	0.00
	21	2	–	0.00		21	2	–	0.00		21	2	–	0.00
	25	2	–	0.00		25	2	–	0.00		25	2	–	0.00
	30	2	–	0.00		30	2	–	0.00		30	2	–	0.00
Transgender	13	2	–	0.00	Transgender	13	1	–	0.00	Transgender	13	2	–	0.00
	16	2	–	0.00		16	1	–	1.00		16	2	–	0.00
	18	2	0.00	–		18	1	0.00	–		18	2	0.00	–
	21	2	0.00	–		21	1	0.00	–		21	2	0.00	–
	25	2	0.00	–		25	1	0.00	–		25	2	0.00	–
	30	2	0.00	–		30	1	0.00	–		30	2	0.00	–
Eyepatch					**Wig**					**Costume Mask**				
Gender	Thr.	*N*	FPR	FNR	Gender	Thr.	*N*	FPR	FNR	Gender	Thr.	*N*	FPR	FNR
Female	13	96	–	0.00	Female	13	98	–	0.00	Female	13	70	–	0.00
	16	96	0.71	0.00		16	98	0.60	0.01		16	70	1.00	0.00
	18	96	0.33	0.04		18	98	0.44	0.06		18	70	1.00	0.00
	21	96	0.32	0.35		21	98	0.21	0.18		21	70	1.00	0.00
	25	96	0.10	0.31		25	98	0.16	0.20		25	70	1.00	0.00
	30	96	0.01	0.08		30	98	0.02	0.00		30	70	1.00	0.00
Male	13	74	–	0.00	Male	13	75	–	0.00	Male	13	55	–	0.00
	16	74	0.25	0.00		16	75	0.33	0.00		16	55	1.00	0.00
	18	74	0.00	0.06		18	75	0.00	0.03		18	55	1.00	0.00
	21	74	0.07	0.09		21	75	0.07	0.07		21	55	1.00	0.00
	25	74	0.12	0.13		25	75	0.12	0.16		25	55	1.00	0.00
	30	74	0.05	0.08		30	75	0.05	0.36		30	55	1.00	0.00
Non-binary	13	5	–	0.00	Non-binary	13	5	–	0.00	Non-binary	13	6	–	0.00
	16	5	–	0.00		16	5	–	0.00		16	6	–	0.00
	18	5	–	0.40		18	5	–	0.20		18	6	–	0.00
	21	5	0.20	–		21	5	0.20	–		21	6	1.00	–
	25	5	0.00	–		25	5	0.00	–		25	6	1.00	–
	30	5	0.00	–		30	5	0.00	–		30	6	1.00	–
Other	13	1	–	0.00	Other	13	2	–	0.00	Other	13	2	–	0.00
	16	1	–	0.00		16	2	–	0.00		16	2	–	0.00
	18	1	–	0.00		18	2	–	0.00		18	2	–	0.00
	21	1	–	0.00		21	2	–	0.00		21	2	–	0.00
	25	1	–	0.00		25	2	–	0.00		25	2	–	0.00
	30	1	–	0.00		30	2	–	0.00		30	2	–	0.00
Transgender	13	2	–	0.00	Transgender	13	2	–	0.00	Transgender	13	2	–	0.00
	16	2	–	0.00		16	2	–	0.00		16	2	–	0.00
	18	2	0.50	–		18	2	0.50	–		18	2	1.00	–
	21	2	0.00	–		21	2	0.00	–		21	2	1.00	–
	25	2	0.00	–		25	2	0.00	–		25	2	1.00	–
	30	2	0.00	–		30	2	0.00	–		30	2	1.00	–

S Age prediction FPR and FNR by condition and gender identity, PRIVATELY

Table 19: PRIVATELY FPR and FNR by condition, gender identity, and age threshold. Rates are proportions.

Privately FPR/FNR by condition, gender identity, and threshold

Gender	Thr.	*N*	FPR	FNR
Baseline				
Female	21	51	0.32	0.58
	25	51	0.15	0.64
	30	51	0.02	0.60
Male	21	38	0.21	0.55
	25	38	0.00	0.64
	30	38	0.03	0.70
Non-binary	21	3	0.25	–
	25	3	0.00	–
	30	3	0.00	–
Other	21	1	–	0.50
	25	1	–	0.50
	30	1	–	0.50
Transgender	21	1	0.00	–
	25	1	0.00	–
	30	1	0.00	–
Gawk				
Female	21	39	0.39	0.59
	25	39	0.07	0.71
	30	39	0.00	0.62
Male	21	22	0.30	0.54
	25	22	0.00	0.65
	30	22	0.05	0.67
Non-binary	21	3	0.40	–
	25	3	0.00	–
	30	3	0.00	–
Other	21	1	–	0.50
	25	1	–	0.50
	30	1	–	0.50
Transgender	21	1	0.00	–
	25	1	0.00	–
	30	1	0.00	–
Frown				
Female	21	38	0.32	0.59
	25	38	0.24	0.63
	30	38	0.08	0.62
Male	21	23	0.30	0.53
	25	23	0.08	0.60
	30	23	0.05	0.67
Non-binary	21	3	0.50	–
	25	3	0.25	–
	30	3	0.00	–
Other	21	1	–	0.50
	25	1	–	0.50
	30	1	–	0.50
Transgender	21	1	0.00	–
	25	1	0.00	–
	30	1	0.00	–
Smile				
Female	21	41	0.26	0.58
	25	41	0.06	0.75
	30	41	0.03	0.57
Male	21	25	0.20	0.55
	25	25	0.00	0.62
	30	25	0.00	0.56
Non-binary	21	2	0.00	–
	25	2	0.00	–
	30	2	0.00	–
Other	21	1	–	0.50
	25	1	–	0.50
	30	1	–	0.50
Transgender	21	1	0.00	–
	25	1	0.00	–
	30	1	0.00	–
Stache				
Female	21	6	–	0.67
	25	6	0.00	0.67
	30	6	0.00	–
Male	21	1	–	0.50
	25	1	–	0.50
	30	1	–	1.00
Mustache				
Female	21	44	0.35	0.57
	25	44	0.14	0.68
	30	44	0.03	0.60
Male	21	32	0.29	0.59
	25	32	0.10	0.64
	30	32	0.00	0.62
Non-binary	21	3	0.40	–
	25	3	0.00	–
	30	3	0.00	–
Other	21	1	–	0.50
	25	1	–	0.50
	30	1	–	0.50
Transgender	21	1	0.00	–
	25	1	0.00	–
	30	1	0.00	–
Beard				
Female	21	42	0.44	0.56
	25	42	0.14	0.71
	30	42	0.03	0.56
Male	21	33	0.41	0.53
	25	33	0.00	0.58
	30	33	0.00	0.56
Non-binary	21	3	0.50	–
	25	3	0.25	–
	30	3	0.00	–
Other	21	1	–	0.50
	25	1	–	0.50
	30	1	–	0.50
Transgender	21	1	0.00	–
	25	1	0.00	–
	30	1	0.00	–
Glasses				
Female	21	51	0.32	0.55
	25	51	0.19	0.67
	30	51	0.02	0.55
Male	21	38	0.32	0.51
	25	38	0.21	0.57
	30	38	0.09	0.58
Non-binary	21	3	0.40	–
	25	3	0.25	–
	30	3	0.00	–
Other	21	1	–	0.50
	25	1	–	0.50
	30	1	–	0.50
Transgender	21	1	0.00	–
	25	1	0.00	–
	30	1	0.00	–
KN95				
Female	21	50	0.33	0.63
	25	50	0.03	0.76
	30	50	0.00	0.67
Male	21	38	0.17	0.72
	25	38	0.00	0.76
	30	38	0.00	0.78
Non-binary	21	3	0.40	–
	25	3	0.00	–
	30	3	0.00	–
Other	21	1	–	0.50
	25	1	–	0.50
	30	1	–	0.50
Transgender	21	1	0.00	–
	25	1	0.00	–
	30	1	0.00	–
Groucho				
Female	21	49	0.33	0.58
	25	49	0.08	0.76
	30	49	0.00	0.67
Male	21	32	0.38	0.53
	25	32	0.00	0.76
	30	32	0.00	0.71
Eyepatch				
Female	21	49	0.17	0.64
	25	49	0.00	0.65
	30	49	0.00	0.60
Male	21	38	0.12	0.55
	25	38	0.12	0.53
	30	38	0.00	0.58
Wig				
Female	21	49	0.20	0.66
	25	49	0.03	0.71
	30	49	0.02	0.60
Male	21	38	0.00	0.59
	25	38	0.08	0.67
	30	38	0.03	0.64

Continued on next page

Privately FPR/FNR by condition, gender identity, and threshold (continued)

Gender	Thr.	*N*	FPR	FNR	Gender	Thr.	*N*	FPR	FNR	Gender	Thr.	*N*	FPR	FNR
Non-binary	21	3	0.50	–	Non-binary	21	3	0.25	–	Non-binary	21	3	0.25	–
	25	3	0.00	–		25	3	0.25	–		25	3	0.25	–
	30	3	0.00	–		30	3	0.00	–		30	3	0.00	–
Other	21	1	–	0.50	Other	21	1	–	0.50	Other	21	1	–	0.50
	25	1	–	0.50		25	1	–	0.50		25	1	–	0.50
	30	1	–	0.50		30	1	–	0.50		30	1	–	0.50
Transgender	21	1	0.00	–	Transgender	21	1	0.00	–	Transgender	21	1	0.00	–
	25	1	0.00	–		25	1	0.00	–		25	1	0.00	–
	30	1	0.00	–		30	1	0.00	–		30	1	0.00	–

Costume Mask

Gender	Thr.	*N*	FPR	FNR	Gender	Thr.	*N*	FPR	FNR	Gender	Thr.	*N*	FPR	FNR
Female	21	43	0.50	0.50										
	25	43	0.50	0.50										
	30	43	0.50	0.50										
Male	21	36	0.50	0.50										
	25	36	0.50	0.50										
	30	36	0.50	0.50										
Non-binary	21	3	0.50	–										
	25	3	0.50	–										
	30	3	0.50	–										
Other	21	1	–	0.50										
	25	1	–	0.50										
	30	1	–	0.50										
Transgender	21	1	0.50	–										
	25	1	0.50	–										
	30	1	0.50	–										

T Age prediction FPR and FNR by condition and gender identity, FACETEC

Table 20: We stratify participants by self-reported gender identity, reporting FNR and FPR on all test conditions and thresholds for FACETEC.

Gender	Thr.	*N*	FPR	FNR	Gender	Thr.	*N*	FPR	FNR	Gender	Thr.	*N*	FPR	FNR
	Baseline					**Gawk**					**Frown**			
Female	13	12	–	0.52	Female	13	7	–	0.54	Female	13	14	–	0.58
	16	21	0.14	0.54		16	13	0.20	0.53		16	28	0.12	0.55
	18	65	0.07	0.63		18	26	0.11	0.64		18	56	0.06	0.62
	21	76	0.16	0.63		21	26	0.07	0.86		21	53	0.18	0.76
	25	54	0.03	0.67		25	20	0.15	0.75		25	49	0.17	0.65
	30	46	0.00	0.62		30	13	0.00	1.00		30	23	0.05	0.67
Male	13	7	–	0.58	Male	13	2	–	0.50	Male	13	2	–	0.50
	16	14	0.00	0.52		16	7	–	0.54		16	11	–	0.50
	18	49	0.00	0.57		18	17	0.00	0.62		18	32	0.33	0.60
	21	55	0.12	0.60		21	19	0.11	0.55		21	32	0.00	0.59
	25	48	0.00	0.59		25	15	0.00	0.67		25	31	0.11	0.60
	30	37	0.09	0.62		30	11	0.00	0.50		30	27	0.09	0.60
Non-binary	16	1	–	0.50	Non-binary	21	1	0.00	–	Non-binary	16	3	–	0.60
	18	4	–	1.00	Other	18	2	–	0.50		18	2	–	0.67
	21	3	0.25	–		25	1	–	0.50		21	2	0.00	–
	25	4	0.20	–							25	3	0.00	–
	30	2	0.00	–						Other	18	1	–	0.50
Other	18	2	–	0.50							21	2	–	0.50
	21	2	–	0.50							25	2	–	0.50
	25	2	–	0.50							30	2	–	0.50
	30	1	–	0.50						Transgender	16	1	–	1.00
Transgender	16	1	–	1.00							18	1	0.00	–
	18	1	0.00	–							21	1	0.00	–
	21	1	0.00	–										
	Smile					**Mustache**					**Beard**			
Female	13	3	–	0.60	Female	13	4	–	0.50	Female	18	7	–	0.50
	16	4	–	0.57		16	9	0.40	0.50		21	7	–	0.64
	18	13	0.00	0.67		18	19	0.33	0.54		25	7	–	0.88
	21	14	0.15	1.00		21	16	0.25	0.67		30	7	0.00	1.00
	25	9	0.10	–		25	12	0.20	0.80	Male	16	2	–	0.50
	30	4	0.00	1.00		30	10	0.00	0.50		18	5	–	0.50
Male	16	3	–	0.50	Male	16	7	0.00	0.50		21	2	0.50	0.50
	18	6	–	0.55		18	18	0.50	0.59		25	3	0.00	0.67
	21	8	0.00	0.50		21	13	0.33	0.70		30	2	0.00	–
	25	5	0.00	0.60		25	18	0.00	0.71					
	30	4	0.00	0.50		30	12	0.00	0.50					
Transgender	16	2	–	1.00	Non-binary	18	1	–	0.50					
	18	2	0.00	–	Transgender	16	1	–	0.50					
	25	2	0.00	–		21	2	0.00	–					
	Glasses					**Wig**								
Female	13	8	–	0.53	Female	13	13	–	0.54					
	16	18	0.50	0.52		16	29	0.11	0.54					
	18	53	0.00	0.63		18	66	0.00	0.79					
	21	49	0.16	0.61		21	64	0.08	0.76					
	25	45	0.18	0.74		25	50	0.06	0.67					
	30	21	0.14	0.67		30	36	0.00	0.62					
Male	13	2	–	0.50	Male	13	3	–	0.60					
	16	13	–	0.52		16	15	0.00	0.52					
	18	40	–	0.57		18	50	0.00	0.69					
	21	40	0.15	0.53		21	49	0.00	0.66					
	25	35	0.12	0.57		25	44	0.03	0.70					
	30	25	0.13	0.62		30	34	0.04	0.57					
Non-binary	18	1	–	0.50	Non-binary	16	4	–	0.57					
	21	3	0.40	–		18	3	–	0.75					
	25	1	0.50	–		21	2	0.00	–					
	30	2	0.00	–		25	3	0.00	–					
Other	18	1	–	0.50		30	1	0.00	–					
	25	1	–	0.50	Other	18	2	–	0.50					
	30	2	–	0.50		21	1	–	0.50					
						25	2	–	0.50					
						30	2	–	0.50					
					Transgender	16	1	–	0.50					
						25	1	0.00	–					

U FNR and FPR by test condition and race, FACETEC

Table 21: Age-prediction performance by self-reported racial group, threshold, and test condition for FACETEC. FPR and FNR are reported as proportions.

Race	Threshold	*N*	FPR	FNR	Race	Threshold	*N*	FPR	FNR	Race	Threshold	*N*	FPR	FNR
	Baseline					**Open-mouth**					**Frown**			
AI/AN	13	1	–	1.00	AI/AN	18	1	–	0.50	AI/AN	16	1	–	1.00
	18	1	–	0.50		21	1	–	0.50		18	1	–	1.00
	21	1	–	0.50		25	1	–	0.50		25	4	0.00	0.50
	25	1	–	0.50							30	1	–	0.50
	30	2	–	0.50	AS	13	2	–	0.50					
						16	8	–	0.50	AS	13	4	–	0.50
AS	13	3	–	0.50		18	25	0.33	0.66		16	15	–	0.54
	16	13	–	0.50		21	29	0.08	0.72		18	47	0.20	0.59
	18	63	0.20	0.58		25	26	0.10	0.70		21	45	0.23	0.68
	21	70	0.29	0.64		30	18	0.00	1.00		25	44	0.19	0.69
	25	58	0.03	0.75							30	31	0.06	1.00
	30	50	0.04	1.00	BL/AA	16	2	–	0.50					
						18	5	–	0.56	BL/AA	16	4	–	0.57
BL/AA	16	1	–	0.50		21	4	0.33	0.50		18	8	–	0.67
	18	7	–	0.64		25	2	0.33	–		21	7	0.00	0.50
	21	6	0.00	0.50		30	3	–	0.50		25	10	0.00	0.50
	25	6	0.00	0.50							30	3	–	0.50
	30	3	–	0.50	HI/LA	16	2	–	0.50					
						18	4	–	0.67	HI/LA	16	3	–	0.50
HI/LA	16	2	–	0.50		21	2	0.00	1.00		18	10	–	0.77
	18	12	–	0.75		25	5	0.00	1.00		21	9	0.00	1.00
	21	19	0.17	0.75		30	3	0.00	–		25	10	0.00	1.00
	25	15	0.00	0.75							30	6	0.00	–
	30	11	0.00	–	MR	16	3	–	0.50					
						18	6	–	0.60	MR	16	6	–	0.55
ME/NA	16	1	–	0.50		21	7	0.20	0.50		18	15	–	0.71
	18	1	–	0.50		25	7	0.14	0.50		21	12	0.00	0.56
	21	2	0.00	–		30	3	0.00	–		25	17	0.09	0.54
	25	1	0.00	–							30	8	0.12	0.50
					NH/PI	16	2	–	0.50					
MR	13	1	–	1.00		18	2	–	0.67	NH/PI	16	2	–	0.50
	16	2	–	0.50		21	2	0.33	–		18	2	–	0.67
	18	14	–	0.64		25	2	0.33	–		21	2	0.00	–
	21	17	0.27	0.56							25	2	0.00	–
	25	17	0.00	0.56	OT	13	2	–	0.50					
	30	10	0.11	0.50		16	1	–	1.00	OT	13	2	–	0.50
						18	4	0.00	0.50		16	2	–	1.00
NH/PI	16	1	–	0.50		21	3	0.00	0.50		18	3	0.00	0.50
	18	1	–	1.00		25	2	–	0.50		21	3	0.00	0.50
	21	1	0.00	–							25	3	–	0.50
	25	2	0.00	–	WH	13	5	–	0.56		30	3	–	0.50
						16	8	0.20	0.57					
OT	13	3	–	0.60		18	10	0.00	0.57	WH	13	10	–	0.62
	16	1	–	1.00		21	13	0.00	0.50		16	22	0.12	0.52
	18	4	0.00	0.50		25	6	0.00	0.50		18	36	0.07	0.64
	21	4	0.00	0.50		30	3	0.00	–		21	36	0.08	0.57
	25	3	–	0.50							25	30	0.17	0.53
	30	2	–	0.50							30	17	0.15	0.67
WH	13	14	–	0.56										
	16	20	0.10	0.58										
	18	47	0.00	0.63										
	21	52	0.09	0.54										
	25	40	0.04	0.52										
	30	29	0.10	0.59										
	Smile					**Mustache**					**Beard**			
AI/AN	18	1	–	0.50	AI/AN	30	1	–	0.50	AS	16	2	–	0.50
											18	11	–	0.50
AS	18	4	–	0.57	AS	13	2	–	0.50		21	7	0.50	0.60
	21	10	0.20	0.60		16	6	–	0.50		25	8	0.00	0.78
	25	5	0.00	0.67		18	19	0.50	0.61		30	7	0.00	1.00
	30	6	0.00	–		21	18	0.30	0.73					
						25	17	0.18	0.80	HI/LA	18	1	–	0.50

Continued on next page

Table 21 – continued from previous page

Race	Threshold	*N*	FPR	FNR	Race	Threshold	*N*	FPR	FNR	Race	Threshold	*N*	FPR	FNR
BL/AA	16	2	–	0.50		30	11	0.00	–		21	2	–	0.67
	18	6	–	0.60							25	2	–	1.00
	21	4	0.33	0.50	BL/AA	16	1	–	0.50		30	2	0.00	–
	25	5	0.00	0.50		18	5	–	0.50					
	30	1	–	0.50		21	2	0.50	0.50	MR	18	3	–	0.50
						25	4	0.00	0.50		21	4	–	0.80
HI/LA	16	2	–	0.50		30	3	–	0.50		25	4	–	1.00
	18	4	–	0.80							30	4	0.00	–
	21	3	0.00	–	HI/LA	16	2	–	0.50					
	25	4	0.00	–		18	6	–	0.55	WH	18	3	–	0.50
						21	6	0.33	0.67		21	4	–	0.80
MR	16	2	–	0.50		25	6	0.00	1.00		25	4	–	1.00
	18	4	–	0.67		30	5	0.00	–		30	4	0.00	–
	21	4	0.33	0.50										
	25	4	0.00	–	MR	16	1	–	0.50					
	30	3	0.00	–		18	4	–	0.50					
						21	2	0.33	–					
NH/PI	16	2	–	0.50		25	4	0.00	1.00					
	18	1	–	0.50		30	2	0.00	0.50					
	21	1	0.50	–										
	25	2	0.00	–	NH/PI	16	1	–	0.50					
						18	2	–	0.50					
OT	13	2	–	0.50		21	1	0.50	–					
	16	1	–	1.00		25	2	0.00	–					
	18	2	0.00	–										
	21	2	0.00	–	WH	13	2	–	0.50					
						16	8	0.33	0.50					
WH	13	1	–	1.00		18	10	0.25	0.50					
	16	4	–	0.67		21	6	0.00	0.50					
	18	7	0.00	0.67		25	5	0.00	0.75					
	21	6	0.00	0.60		30	4	0.00	0.50					
	25	4	0.20	–										
	30	4	0.00	1.00										

Glasses

AI/AN	21	1	0.00	–
AS	13	3	–	0.50
	16	16	–	0.52
	18	50	0.00	0.59
	21	51	0.25	0.61
	25	46	0.24	0.71
	30	30	0.12	–
BL/AA	16	1	–	0.50
	18	5	–	0.71
	21	3	0.00	0.50
	25	4	0.00	0.50
	30	3	0.00	0.50
HI/LA	16	5	–	0.50
	18	14	–	0.70
	21	13	0.25	0.58
	25	12	0.11	1.00
	30	8	0.20	–
ME/NA	16	1	–	0.50
	18	1	–	0.50
	21	2	0.00	–
	25	2	0.00	–
MR	16	4	–	0.57
	18	14	–	0.64
	21	12	0.25	0.55
	25	14	0.27	0.75
	30	8	0.27	–
NH/PI	18	2	–	1.00
	25	1	0.00	–
OT	13	1	–	0.50
	16	1	–	1.00
	18	1	–	0.50
	21	2	0.00	–

Wig

AI/AN	16	2	–	0.67
	18	4	–	0.80
	21	3	0.00	0.50
	25	2	0.00	–
	30	1	–	0.50
AS	13	3	–	0.50
	16	19	–	0.50
	18	61	0.00	0.70
	21	59	0.10	0.75
	25	57	0.06	0.75
	30	43	0.00	0.67
BL/AA	16	5	–	0.56
	18	11	–	0.73
	21	10	0.00	0.50
	25	9	0.00	0.50
	30	4	–	0.67
HI/LA	16	5	–	0.50
	18	16	–	1.00
	21	15	0.00	1.00
	25	15	0.00	1.00
	30	8	0.00	–
MR	16	6	–	0.55
	18	19	–	0.83
	21	14	0.00	0.83
	25	15	0.00	0.57
	30	10	0.00	0.50
NH/PI	16	2	–	0.50
	18	2	–	1.00
	21	2	0.00	–
	25	2	0.00	–
OT	13	1	–	0.50
	16	1	–	1.00
	18	5	0.00	0.50

Continued on next page

Table 21 – continued from previous page

Race	Threshold	*N*	FPR	FNR	Race	Threshold	*N*	FPR	FNR	Race	Threshold	*N*	FPR	FNR
	25	1	–	0.50		21	3	0.00	0.50					
	30	2	–	0.50		25	2	–	0.50					
						30	3	–	0.50					
WH	13	6	–	0.55										
	16	11	0.50	0.53	WH	13	12	–	0.57					
	18	36	0.00	0.56		16	21	0.09	0.58					
	21	32	0.12	0.51		18	42	0.00	0.74					
	25	30	0.15	0.59		21	39	0.04	0.64					
	30	15	0.23	0.71		25	28	0.05	0.54					
						30	25	0.05	0.54					

V FNR and FPR by test condition and race, Yoti

Table 22: Yoti FPR and FNR by condition, race, and age threshold. Rates are proportions.

Yoti FPR/FNR by condition, race, and age threshold														
Baseline					**Gawk**					**Frown**				
Race	Thr.	*N*	FPR	FNR	Race	Thr.	*N*	FPR	FNR	Race	Thr.	*N*	FPR	FNR
AS	13	92	–	0.00	AS	13	68	–	0.00	AS	13	64	–	0.00
	16	92	–	0.00		16	68	–	0.00		16	64	–	0.00
	18	92	1.00	0.02		18	68	1.00	0.05		18	64	1.00	0.02
	21	92	0.19	0.15		21	68	0.32	0.26		21	64	0.33	0.07
	25	92	0.14	0.14		25	68	0.10	0.32		25	64	0.14	0.14
	30	92	0.05	0.25		30	68	0.02	0.00		30	64	0.08	0.50
WH	13	76	–	0.01	WH	13	47	–	0.00	WH	13	50	–	0.00
	16	76	0.36	0.00		16	47	0.38	0.05		16	50	0.62	0.00
	18	76	0.18	0.15		18	47	0.29	0.27		18	50	0.36	0.06
	21	76	0.14	0.12		21	47	0.10	0.11		21	50	0.17	0.24
	25	76	0.04	0.08		25	47	0.00	0.00		25	50	0.03	0.00
	30	76	0.03	0.06		30	47	0.00	0.00		30	50	0.10	0.00
BL/AA	13	12	–	0.00	BL/AA	13	10	–	0.00	BL/AA	13	12	–	0.00
	16	12	–	0.00		16	10	–	0.00		16	12	–	0.00
	18	12	–	0.00		18	10	–	0.00		18	12	–	0.00
	21	12	0.33	0.00		21	10	0.50	0.00		21	12	0.33	0.00
	25	12	0.17	0.00		25	10	0.00	0.00		25	12	0.33	0.00
	30	12	0.00	0.00		30	10	0.00	0.00		30	12	0.00	0.00
HI/LA	13	20	–	0.00	HI/LA	13	8	–	0.00	HI/LA	13	10	–	0.00
	16	20	–	0.00		16	8	–	0.00		16	10	–	0.00
	18	20	–	0.00		18	8	–	0.00		18	10	–	0.00
	21	20	0.20	0.20		21	8	0.25	0.50		21	10	0.67	0.50
	25	20	0.19	0.75		25	8	0.50	1.00		25	10	0.12	1.00
	30	20	0.00	–		30	8	0.00	–		30	10	0.00	–
ME/NA	13	2	–	0.00	AI/AN	13	6	–	0.00	AI/AN	13	6	–	0.00
	16	2	–	0.00		16	6	–	0.00		16	6	–	0.00
	18	2	–	0.00		18	6	–	0.00		18	6	–	0.00
	21	2	0.00	–		21	6	0.00	0.00		21	6	0.00	0.00
	25	2	0.00	–		25	6	0.00	0.00		25	6	0.00	0.00
	30	2	0.00	–		30	6	0.00	0.00		30	6	0.00	0.00
AI/AN	13	6	–	0.00	NH/PI	13	2	–	0.00	NH/PI	13	2	–	0.00
	16	6	–	0.00		16	2	–	0.00		16	2	–	0.00
	18	6	–	0.00		18	2	–	0.00		18	2	–	0.00
	21	6	0.00	0.00		21	2	1.00	–		21	2	1.00	–
	25	6	0.00	0.00		25	2	0.00	–		25	2	1.00	–
	30	6	0.00	0.00		30	2	0.00	–		30	2	0.00	–
NH/PI	13	2	–	0.00	OT	13	6	–	0.00	OT	13	6	–	0.00
	16	2	–	0.00		16	6	–	0.00		16	6	–	0.00
	18	2	–	0.00		18	6	0.50	0.00		18	6	0.50	0.00
	21	2	1.00	–		21	6	0.00	0.00		21	6	0.00	0.00
	25	2	0.50	–		25	6	0.00	0.00		25	6	0.00	0.00
	30	2	0.00	–		30	6	0.00	0.00		30	6	0.00	0.00
OT	13	6	–	0.00	MR	13	15	–	0.00	MR	13	19	–	0.00
	16	6	–	0.00		16	15	–	0.00		16	19	–	0.00
	18	6	0.50	0.00		18	15	–	0.07		18	19	–	0.00
	21	6	0.00	0.00		21	15	0.29	0.00		21	19	0.44	0.10
	25	6	0.00	0.00		25	15	0.00	0.00		25	19	0.18	0.00
	30	6	0.00	0.00		30	15	0.00	0.00		30	19	0.13	0.00
MR	13	26	–	0.00										
	16	26	–	0.00										
	18	26	–	0.04										
	21	26	0.33	0.21										
	25	26	0.06	0.20										
	30	26	0.00	0.00										

Smile					**Mustache**					**Beard**				
Race	Thr.	*N*	FPR	FNR	Race	Thr.	*N*	FPR	FNR	Race	Thr.	*N*	FPR	FNR
AS	13	68	–	0.00	AS	13	75	–	0.00	AS	13	74	–	0.00
	16	68	–	0.00		16	75	–	0.00		16	74	–	0.00
	18	68	1.00	0.00		18	75	1.00	0.08		18	74	0.75	0.04
	21	68	0.62	0.13		21	75	0.27	0.09		21	74	0.29	0.07
	25	68	0.15	0.30		25	75	0.25	0.15		25	74	0.20	0.11
	30	68	0.00	–		30	75	0.04	1.00		30	74	0.07	0.50

Continued on next page

Yoti FPR/FNR by condition, race, and threshold (continued)

Race	Thr.	*N*	FPR	FNR	Race	Thr.	*N*	FPR	FNR	Race	Thr.	*N*	FPR	FNR
WH	13	52	–	0.00	WH	13	67	–	0.00	WH	13	68	–	0.00
	16	52	0.50	0.00		16	67	0.43	0.00		16	68	0.57	0.04
	18	52	0.29	0.11		18	67	0.14	0.24		18	68	0.36	0.26
	21	52	0.17	0.04		21	67	0.02	0.04		21	68	0.05	0.07
	25	52	0.09	0.06		25	67	0.00	0.05		25	68	0.04	0.09
	30	52	0.00	0.09		30	67	0.02	0.07		30	68	0.02	0.14
BL/AA	13	12	–	0.00	BL/AA	13	12	–	0.00	BL/AA	13	10	–	0.00
	16	12	–	0.00		16	12	–	0.00		16	10	–	0.00
	18	12	–	0.00		18	12	–	0.00		18	10	–	0.00
	21	12	0.50	0.00		21	12	0.33	0.00		21	10	0.33	0.00
	25	12	0.00	0.00		25	12	0.00	0.00		25	10	0.17	0.00
	30	12	0.00	0.00		30	12	0.00	0.00		30	10	0.00	0.00
HI/LA	13	14	–	0.00	HI/LA	13	18	–	0.00	HI/LA	13	18	–	0.00
	16	14	–	0.00		16	18	–	0.00		16	18	–	0.00
	18	14	–	0.14		18	18	–	0.00		18	18	–	0.00
	21	14	0.38	0.17		21	18	0.00	0.12		21	18	0.40	0.00
	25	14	0.08	0.00		25	18	0.21	0.25		25	18	0.21	0.75
	30	14	0.00	–		30	18	0.00	–		30	18	0.00	–
AI/AN	13	6	–	0.00	AI/AN	13	6	–	0.00	AI/AN	13	4	–	0.00
	16	6	–	0.00		16	6	–	0.00		16	4	–	0.00
	18	6	–	0.00		18	6	–	0.00		18	4	–	0.00
	21	6	0.50	0.00		21	6	0.00	0.00		21	4	0.00	0.00
	25	6	0.00	0.00		25	6	0.00	0.00		25	4	0.00	0.00
	30	6	0.00	0.00		30	6	0.00	0.00		30	4	0.00	0.00
NH/PI	13	2	–	0.00	NH/PI	13	2	–	0.00	NH/PI	13	2	–	0.00
	16	2	–	0.00		16	2	–	0.00		16	2	–	0.00
	18	2	–	0.00		18	2	–	0.00		18	2	–	0.00
	21	2	1.00	–		21	2	1.00	–		21	2	1.00	–
	25	2	0.00	–		25	2	0.00	–		25	2	0.50	–
	30	2	0.00	–		30	2	0.00	–		30	2	0.00	–
OT	13	6	–	0.00	OT	13	6	–	0.00	OT	13	6	–	0.00
	16	6	–	0.00		16	6	–	0.00		16	6	–	0.00
	18	6	0.50	0.00		18	6	0.50	0.00		18	6	0.50	0.00
	21	6	0.00	0.00		21	6	0.00	0.00		21	6	0.00	0.00
	25	6	0.00	0.00		25	6	0.00	0.00		25	6	0.00	0.00
	30	6	0.00	0.00		30	6	0.00	0.00		30	6	0.00	0.00
MR	13	21	–	0.00	MR	13	23	–	0.00	MR	13	20	–	0.00
	16	21	–	0.00		16	23	–	0.00		16	20	–	0.00
	18	21	–	0.00		18	23	–	0.09		18	20	–	0.05
	21	21	0.67	0.08		21	23	0.25	0.00		21	20	0.20	0.00
	25	21	0.00	0.12		25	23	0.00	0.11		25	20	0.17	0.25
	30	21	0.00	0.00		30	23	0.00	0.00		30	20	0.06	0.00

Glasses					**KN95**					**Groucho**				
Race	Thr.	*N*	FPR	FNR	Race	Thr.	*N*	FPR	FNR	Race	Thr.	*N*	FPR	FNR
AS	13	92	–	0.00	AS	13	90	–	0.00	AS	13	84	–	0.00
	16	92	–	0.00		16	90	–	0.01		16	84	–	0.00
	18	92	1.00	0.01		18	90	1.00	0.05		18	84	0.75	0.18
	21	92	0.41	0.28		21	90	0.37	0.25		21	84	0.04	0.50
	25	92	0.18	0.25		25	90	0.15	0.44		25	84	0.02	0.56
	30	92	0.10	0.50		30	90	0.03	0.75		30	84	0.01	1.00
WH	13	76	–	0.00	WH	13	60	–	0.00	WH	13	62	–	0.00
	16	76	0.50	0.00		16	60	0.50	0.08		16	62	0.50	0.00
	18	76	0.09	0.09		18	60	0.24	0.21		18	62	0.11	0.20
	21	76	0.09	0.19		21	60	0.16	0.24		21	62	0.13	0.12
	25	76	0.04	0.00		25	60	0.05	0.19		25	62	0.11	0.22
	30	76	0.03	0.00		30	60	0.06	0.31		30	62	0.00	0.10
BL/AA	13	12	–	0.00	BL/AA	13	12	–	0.00	BL/AA	13	10	–	0.00
	16	12	–	0.00		16	12	–	0.00		16	10	–	0.10
	18	12	–	0.00		18	12	–	0.08		18	10	–	0.40
	21	12	0.33	0.00		21	12	0.50	0.00		21	10	0.00	0.00
	25	12	0.17	0.00		25	12	0.00	0.00		25	10	0.00	0.25
	30	12	0.00	0.00		30	12	0.00	0.00		30	10	0.00	0.50
HI/LA	13	19	–	0.00	HI/LA	13	20	–	0.00	HI/LA	13	19	–	0.00
	16	19	–	0.00		16	20	–	0.05		16	19	–	0.05
	18	19	–	0.05		18	20	–	0.20		18	19	–	0.37
	21	19	0.10	0.33		21	20	0.20	0.40		21	19	0.20	0.78
	25	19	0.13	0.50		25	20	0.12	0.75		25	19	0.07	1.00
	30	19	0.00	–		30	20	0.00	–		30	19	0.00	–

Continued on next page

Yoti FPR/FNR by condition, race, and threshold (continued)

Race	Thr.	*N*	FPR	FNR	Race	Thr.	*N*	FPR	FNR	Race	Thr.	*N*	FPR	FNR
ME/NA	13	2	–	0.00	ME/NA	13	2	–	0.00	ME/NA	13	2	–	0.00
	16	2	–	0.00		16	2	–	0.00		16	2	–	0.00
	18	2	–	0.00		18	2	–	0.00		18	2	–	0.00
	21	2	0.00	–		21	2	0.00	–		21	2	0.00	–
	25	2	0.00	–		25	2	0.00	–		25	2	0.00	–
	30	2	0.00	–		30	2	0.00	–		30	2	0.00	–
AI/AN	13	6	–	0.00	AI/AN	13	4	–	0.00	AI/AN	13	5	–	0.00
	16	6	–	0.00		16	4	–	0.00		16	5	–	0.00
	18	6	–	0.00		18	4	–	0.00		18	5	–	0.40
	21	6	0.00	0.00		21	4	1.00	0.00		21	5	0.00	0.00
	25	6	0.00	0.00		25	4	0.00	0.00		25	5	0.00	0.33
	30	6	0.00	0.00		30	4	0.00	0.00		30	5	0.00	0.33
NH/PI	13	2	–	0.00	NH/PI	13	2	–	0.00	NH/PI	13	2	–	0.00
	16	2	–	0.00		16	2	–	0.00		16	2	–	0.00
	18	2	–	0.00		18	2	–	0.00		18	2	–	0.00
	21	2	1.00	–		21	2	0.50	–		21	2	0.00	–
	25	2	0.50	–		25	2	0.00	–		25	2	0.00	–
	30	2	0.00	–		30	2	0.00	–		30	2	0.00	–
OT	13	6	–	0.00	OT	13	4	–	0.00	OT	13	5	–	0.00
	16	6	–	0.00		16	4	–	0.00		16	5	–	0.00
	18	6	0.00	0.00		18	4	0.50	0.00		18	5	0.00	0.00
	21	6	0.00	0.00		21	4	0.00	0.00		21	5	0.00	0.00
	25	6	0.00	0.00		25	4	0.00	0.00		25	5	0.00	0.00
	30	6	0.00	0.00		30	4	0.00	0.00		30	5	0.00	0.00
MR	13	26	–	0.00	MR	13	22	–	0.00	MR	13	23	–	0.00
	16	26	–	0.00		16	22	–	0.05		16	23	–	0.04
	18	26	–	0.04		18	22	–	0.09		18	23	–	0.30
	21	26	0.33	0.21		21	22	0.30	0.33		21	23	0.20	0.38
	25	26	0.12	0.00		25	22	0.00	0.50		25	23	0.07	0.56
	30	26	0.00	0.00		30	22	0.05	0.00		30	23	0.00	0.33

Eyepatch					**Wig**					**Costume Mask**				
Race	Thr.	*N*	FPR	FNR	Race	Thr.	*N*	FPR	FNR	Race	Thr.	*N*	FPR	FNR
AS	13	91	–	0.00	AS	13	90	–	0.00	AS	13	70	–	0.00
	16	91	–	0.00		16	90	–	0.00		16	70	–	0.00
	18	91	0.75	0.05		18	90	1.00	0.02		18	70	1.00	0.00
	21	91	0.31	0.25		21	90	0.27	0.15		21	70	1.00	0.00
	25	91	0.14	0.34		25	90	0.17	0.25		25	70	1.00	0.00
	30	91	0.03	0.25		30	90	0.05	0.50		30	70	1.00	–
WH	13	68	–	0.00	WH	13	70	–	0.00	WH	13	56	–	0.00
	16	68	0.55	0.00		16	70	0.54	0.02		16	56	1.00	0.00
	18	68	0.16	0.14		18	70	0.19	0.12		18	56	1.00	0.00
	21	68	0.15	0.21		21	70	0.10	0.10		21	56	1.00	0.00
	25	68	0.09	0.19		25	70	0.08	0.00		25	56	1.00	0.00
	30	68	0.02	0.00		30	70	0.02	0.19		30	56	1.00	0.00
BL/AA	13	12	–	0.00	BL/AA	13	12	–	0.00	BL/AA	13	8	–	0.00
	16	12	–	0.00		16	12	–	0.00		16	8	–	0.00
	18	12	–	0.08		18	12	–	0.00		18	8	–	0.00
	21	12	0.33	0.00		21	12	0.00	0.00		21	8	1.00	0.00
	25	12	0.17	0.00		25	12	0.00	0.00		25	8	1.00	0.00
	30	12	0.00	0.17		30	12	0.00	0.00		30	8	1.00	0.00
HI/LA	13	18	–	0.00	HI/LA	13	18	–	0.00	HI/LA	13	12	–	0.00
	16	18	–	0.00		16	18	–	0.00		16	12	–	0.00
	18	18	–	0.06		18	18	–	0.00		18	12	–	0.00
	21	18	0.20	0.38		21	18	0.30	0.12		21	12	1.00	0.00
	25	18	0.12	0.50		25	18	0.31	1.00		25	12	1.00	0.00
	30	18	0.00	–		30	18	0.00	–		30	12	1.00	–
ME/NA	13	2	–	0.00	ME/NA	13	2	–	0.00	ME/NA	13	1	–	0.00
	16	2	–	0.00		16	2	–	0.00		16	1	–	0.00
	18	2	–	0.00		18	2	–	0.00		18	1	–	0.00
	21	2	0.00	–		21	2	0.00	–		21	1	1.00	–
	25	2	0.00	–		25	2	0.00	–		25	1	1.00	–
	30	2	0.00	–		30	2	0.00	–		30	1	1.00	–
AI/AN	13	5	–	0.00	AI/AN	13	6	–	0.00	AI/AN	13	3	–	0.00
	16	5	–	0.00		16	6	–	0.00		16	3	–	0.00
	18	5	–	0.20		18	6	–	0.00		18	3	–	0.00
	21	5	0.00	0.00		21	6	0.00	0.00		21	3	1.00	0.00
	25	5	0.00	0.00		25	6	0.00	0.00		25	3	1.00	0.00
	30	5	0.00	0.33		30	6	0.00	0.00		30	3	1.00	0.00

Continued on next page

Yoti FPR/FNR by condition, race, and threshold (continued)

Race	Thr.	*N*	FPR	FNR	Race	Thr.	*N*	FPR	FNR	Race	Thr.	*N*	FPR	FNR
NH/PI	13	2	–	0.00	NH/PI	13	2	–	0.00	NH/PI	13	2	–	0.00
	16	2	–	0.00		16	2	–	0.00		16	2	–	0.00
	18	2	–	0.00		18	2	–	0.00		18	2	–	0.00
	21	2	1.00	–		21	2	0.00	–		21	2	1.00	–
	25	2	0.50	–		25	2	0.00	–		25	2	1.00	–
	30	2	0.00	–		30	2	0.00	–		30	2	1.00	–
OT	13	4	–	0.00	OT	13	6	–	0.00	OT	13	3	–	0.00
	16	4	–	0.00		16	6	–	0.00		16	3	–	0.00
	18	4	0.00	0.00		18	6	0.50	0.00		18	3	–	0.00
	21	4	0.00	0.00		21	6	0.00	0.00		21	3	–	0.00
	25	4	0.00	0.00		25	6	0.00	0.00		25	3	–	0.00
	30	4	0.00	0.00		30	6	0.00	0.00		30	3	–	0.00
MR	13	23	–	0.00	MR	13	22	–	0.00	MR	13	19	–	0.00
	16	23	–	0.00		16	22	–	0.00		16	19	–	0.00
	18	23	–	0.22		18	22	–	0.00		18	19	–	0.00
	21	23	0.33	0.27		21	22	0.20	0.08		21	19	1.00	0.00
	25	23	0.25	0.57		25	22	0.14	0.00		25	19	1.00	0.00
	30	23	0.00	0.33		30	22	0.00	0.00		30	19	1.00	0.00

W FNR and FPR by test condition and race, PRIVATELY

Table 23: PRIVATELY FPR and FNR by condition, race, and age threshold. Rates are proportions.

Baseline					Open-mouth					Frown				
Race	Thr.	*N*	FPR	FNR	Race	Thr.	*N*	FPR	FNR	Race	Thr.	*N*	FPR	FNR
AS	21	46	0.41	0.59	AS	21	35	0.43	0.59	AS	21	33	0.35	0.56
	25	46	0.12	0.75		25	35	0.09	0.78		25	33	0.21	0.70
	30	46	0.04	1.00		30	35	0.03	1.00		30	33	0.11	1.00
WH	21	38	0.21	0.52	WH	21	24	0.35	0.53	WH	21	26	0.29	0.55
	25	38	0.04	0.55		25	24	0.00	0.50		25	26	0.22	0.53
	30	38	0.03	0.57		30	24	0.00	0.56		30	26	0.05	0.56
BL/AA	21	6	0.00	0.50	BL/AA	21	5	0.00	0.50	BL/AA	21	6	0.40	0.50
	25	6	0.00	0.50		25	5	0.00	0.60		25	6	0.00	0.50
	30	6	0.00	0.75		30	5	0.00	0.75		30	6	0.00	0.75
HI/LA	21	10	0.29	0.62	HI/LA	21	4	0.33	0.67	HI/LA	21	5	0.40	1.00
	25	10	0.11	0.67		25	4	0.00	1.00		25	5	0.00	1.00
	30	10	0.00	–		30	4	0.00	–		30	5	0.00	–
ME/NA	21	1	0.00	–	AI/AN	21	3	0.00	0.50	AI/AN	21	3	0.50	0.50
	25	1	0.00	–		25	3	0.00	0.50		25	3	0.00	0.50
	30	1	0.00	–		30	3	0.00	0.67		30	3	0.00	0.67
AI/AN	21	3	0.00	0.50	NH/PI	21	1	0.00	–	NH/PI	21	1	0.50	–
	25	3	0.00	0.50		25	1	0.00	–		25	1	0.00	–
	30	3	0.00	0.67		30	1	0.00	–		30	1	0.00	–
NH/PI	21	1	0.00	–	Other	21	3	0.00	0.50	Other	21	3	0.00	0.50
	25	1	0.00	–		25	3	0.00	0.50		25	3	0.00	0.50
	30	1	0.00	–		30	3	0.00	0.50		30	3	0.00	0.50
Other	21	3	0.00	0.50	MR	21	8	0.20	0.57	MR	21	10	0.38	0.56
	25	3	0.00	0.50		25	8	0.00	0.50		25	10	0.00	0.57
	30	3	0.00	0.50		30	8	0.00	0.67		30	10	0.11	0.67
MR	21	13	0.25	0.54										
	25	13	0.00	0.56										
	30	13	0.08	0.67										

Smile					Mustache					Mustache				
Race	Thr.	*N*	FPR	FNR	Race	Thr.	*N*	FPR	FNR	Race	Thr.	*N*	FPR	FNR
AS	21	35	0.35	0.60	AS	21	5	–	0.71	AS	21	38	0.44	0.59
	25	35	0.09	0.78		25	5	0.00	0.67		25	38	0.14	0.78
	30	35	0.03	–		30	5	0.00	–		30	38	0.00	1.00
WH	21	27	0.17	0.52	WH	21	2	–	0.50	WH	21	34	0.29	0.54
	25	27	0.00	0.56		25	2	0.00	0.50		25	34	0.15	0.58
	30	27	0.00	0.50		30	2	0.00	1.00		30	34	0.04	0.54
BL/AA	21	6	0.00	0.50	HI/LA	21	1	–	1.00	BL/AA	21	6	0.25	0.50
	25	6	0.00	0.60		25	1	0.00	–		25	6	0.00	0.50
	30	6	0.00	0.75		30	1	0.00	–		30	6	0.00	0.75
HI/LA	21	7	0.00	0.75	MR	21	1	–	1.00	HI/LA	21	9	0.00	0.67
	25	7	0.00	1.00		25	1	0.00	–		25	9	0.00	1.00
	30	7	0.00	–		30	1	0.00	–		30	9	0.00	–
AI/AN	21	3	0.00	0.50						AI/AN	21	3	0.00	0.50
	25	3	0.00	0.67							25	3	0.00	0.50
	30	3	0.00	0.67							30	3	0.00	0.67
NH/PI	21	1	0.00	–						NH/PI	21	1	0.50	–
	25	1	0.00	–							25	1	0.00	–
	30	1	0.00	–							30	1	0.00	–
Other	21	3	0.00	0.50						Other	21	3	0.00	0.50
	25	3	0.00	0.50							25	3	0.00	0.50
	30	3	0.00	0.50							30	3	0.00	0.50
MR	21	11	0.00	0.60						MR	21	12	0.25	0.50
	25	11	0.00	0.67							25	12	0.12	0.62
	30	11	0.00	0.67							30	12	0.00	0.67

Beard					Glasses					KN95				
Race	Thr.	*N*	FPR	FNR	Race	Thr.	*N*	FPR	FNR	Race	Thr.	*N*	FPR	FNR
AS	21	37	0.42	0.57	AS	21	46	0.43	0.55	AS	21	45	0.32	0.71
	25	37	0.08	0.74		25	46	0.30	0.67		25	45	0.04	0.90
	30	37	0.00	1.00		30	46	0.08	0.67		30	45	0.00	1.00
WH	21	34	0.44	0.52	WH	21	38	0.24	0.50	WH	21	37	0.30	0.64
	25	34	0.15	0.55		25	38	0.13	0.57		25	37	0.00	0.67
	30	34	0.04	0.54		30	38	0.03	0.53		30	37	0.00	0.62
BL/AA	21	5	0.40	0.50	BL/AA	21	6	0.25	0.50	BL/AA	21	6	0.25	0.50
	25	5	0.00	0.50		25	6	0.00	0.50		25	6	0.00	0.60

Continued on next page

Table 23 – continued from previous page

Race	Thr.	*N*	FPR	FNR	Race	Thr.	*N*	FPR	FNR	Race	Thr.	*N*	FPR	FNR
	30	5	0.00	0.50		30	6	0.00	0.60		30	6	0.00	1.00
HI/LA	21	9	0.29	0.50	HI/LA	21	10	0.29	0.56	HI/LA	21	10	0.00	0.71
	25	9	0.00	1.00		25	10	0.20	1.00		25	10	0.00	1.00
	30	9	0.00	–		30	10	0.00	–		30	10	0.00	–
AI/AN	21	2	0.50	0.50	ME/NA	21	1	0.50	–	ME/NA	21	1	0.00	–
	25	2	0.00	0.50		25	1	0.00	–		25	1	0.00	–
	30	2	0.00	0.50		30	1	0.00	–		30	1	0.00	–
NH/PI	21	1	0.50	–	AI/AN	21	3	0.00	0.50	AI/AN	21	3	0.00	0.50
	25	1	0.00	–		25	3	0.00	0.50		25	3	0.00	0.67
	30	1	0.00	–		30	3	0.00	0.50		30	3	0.00	0.67
Other	21	3	0.50	0.50	NH/PI	21	1	0.50	–	NH/PI	21	1	0.50	–
	25	3	0.00	0.50		25	1	0.00	–		25	1	0.00	–
	30	3	0.00	0.50		30	1	0.00	–		30	1	0.00	–
MR	21	10	0.38	0.56	Other	21	3	0.00	0.50	Other	21	3	0.00	0.50
	25	10	0.00	0.67		25	3	0.00	0.50		25	3	0.00	0.50
	30	10	0.00	0.50		30	3	0.00	0.50		30	3	0.00	0.50
					MR	21	13	0.33	0.50	MR	21	12	0.17	0.64
						25	13	0.27	0.62		25	12	0.00	0.83
						30	13	0.08	0.50		30	12	0.00	0.67

	Groucho					**Eyepatch**					**Wig**			
Race	Thr.	*N*	FPR	FNR	Race	Thr.	*N*	FPR	FNR	Race	Thr.	*N*	FPR	FNR
AS	21	42	0.39	0.58	AS	21	46	0.30	0.61	AS	21	45	0.21	0.68
	25	42	0.04	0.85		25	46	0.12	0.64		25	45	0.13	0.75
	30	42	0.00	1.00		30	46	0.00	1.00		30	45	0.04	1.00
WH	21	34	0.39	0.50	WH	21	36	0.09	0.56	WH	21	36	0.12	0.58
	25	34	0.08	0.61		25	36	0.04	0.52		25	36	0.00	0.58
	30	34	0.00	0.58		30	36	0.00	0.50		30	36	0.03	0.53
BL/AA	21	5	0.25	0.50	BL/AA	21	6	0.00	0.50	BL/AA	21	6	0.00	0.50
	25	5	0.00	1.00		25	6	0.00	0.50		25	6	0.00	0.75
	30	5	0.00	1.00		30	6	0.00	0.75		30	6	0.00	0.75
HI/LA	21	10	0.29	0.56	HI/LA	21	9	0.00	0.80	HI/LA	21	9	0.00	0.67
	25	10	0.00	1.00		25	9	0.00	1.00		25	9	0.00	1.00
	30	10	0.00	–		30	9	0.00	–		30	9	0.00	–
ME/NA	21	1	0.00	–	ME/NA	21	1	0.00	–	ME/NA	21	1	0.00	–
	25	1	0.00	–		25	1	0.00	–		25	1	0.00	–
	30	1	0.00	–		30	1	0.00	–		30	1	0.00	–
AI/AN	21	3	0.50	0.50	AI/AN	21	3	0.00	0.50	AI/AN	21	3	0.00	0.50
	25	3	0.00	0.67		25	3	0.00	0.50		25	3	0.00	0.67
	30	3	0.00	0.67		30	3	0.00	0.67		30	3	0.00	0.67
NH/PI	21	1	0.00	–	NH/PI	21	1	0.00	–	NH/PI	21	1	0.00	–
	25	1	0.00	–		25	1	0.00	–		25	1	0.00	–
	30	1	0.00	–		30	1	0.00	–		30	1	0.00	–
Other	21	3	0.00	0.50	Other	21	3	0.00	0.50	Other	21	3	0.00	0.50
	25	3	0.00	0.50		25	3	0.00	0.50		25	3	0.00	0.50
	30	3	0.00	0.50		30	3	0.00	0.50		30	3	0.00	0.50
MR	21	12	0.38	0.50	MR	21	12	0.14	0.60	MR	21	11	0.00	0.67
	25	12	0.00	0.71		25	12	0.11	0.57		25	11	0.00	0.67
	30	12	0.00	0.67		30	12	0.00	0.67		30	11	0.10	0.67

	Costume Mask													
Race	Thr.	*N*	FPR	FNR	Race	Thr.	*N*	FPR	FNR	Race	Thr.	*N*	FPR	FNR
AS	21	42	0.50	0.50										
	25	42	0.50	0.50										
	30	42	0.50	0.50										
WH	21	36	0.50	0.50										
	25	36	0.50	0.50										
	30	36	0.50	0.50										
BL/AA	21	4	0.50	0.50										
	25	4	0.50	0.50										
	30	4	0.50	0.50										
HI/LA	21	8	0.50	0.50										
	25	8	0.50	0.50										
	30	8	0.50	–										
ME/NA	21	1	0.50	–										
	25	1	0.50	–										
	30	1	0.50	–										
AI/AN	21	2	0.50	0.50										
	25	2	0.50	0.50										
	30	2	0.50	0.50										
NH/PI	21	1	0.50	–										
	25	1	0.50	–										
	30	1	0.50	–										

Continued on next page

Table 23 – continued from previous page

Race	Thr.	*N*	FPR	FNR	Race	Thr.	*N*	FPR	FNR	Race	Thr.	*N*	FPR	FNR
Other	21	2	–	0.50										
	25	2	–	0.50										
	30	2	–	0.50										
MR	21	11	0.50	0.50										
	25	11	0.50	0.50										
	30	11	0.50	0.50										

X Joint age and liveness prediction, YOTI

Table 24: Joint YOTI age-estimation and liveness performance by age threshold, age group, and test condition. FPR and FNR are reported as proportions.

Thr.	Age Group	N	FPR	FNR	Thr.	Age Group	N	FPR	FNR	Thr.	Age Group	N	FPR	FNR
		Baseline					**Open-mouth**					**Frown**		
13	13-16	26	–	0.42	13	13-16	18	–	0.89	13	13-16	18	–	0.39
	17-20	62	–	0.34		17-20	40	–	0.75		17-20	41	–	0.34
	21-24	34	–	0.18		21-24	24	–	0.54		21-24	21	–	0.43
	25-28	30	–	0.20		25-28	22	–	0.91		25-28	24	–	0.46
	29-32	16	–	0.50		29-32	12	–	0.83		29-32	11	–	0.64
	33-44	6	–	0.50		33-44	4	–	0.50		33-44	4	–	0.25
	45-64	14	–	0.50		45-64	10	–	1.00		45-64	10	–	0.80
16	13-16	26	0.21	0.42	16	13-16	18	0.12	0.90	16	13-16	18	0.12	0.30
	17-20	62	–	0.34		17-20	40	–	0.75		17-20	41	–	0.34
	21-24	34	–	0.18		21-24	24	–	0.54		21-24	21	–	0.43
	25-28	30	–	0.20		25-28	22	–	0.91		25-28	24	–	0.46
	29-32	16	–	0.50		29-32	12	–	0.83		29-32	11	–	0.64
	33-44	6	–	0.50		33-44	4	–	0.50		33-44	4	–	0.25
	45-64	14	–	0.50		45-64	10	–	1.00		45-64	10	–	0.80
18	13-16	26	0.19	–	18	13-16	18	0.06	–	18	13-16	18	0.33	–
	17-20	62	0.00	0.47		17-20	40	0.00	0.84		17-20	41	0.50	0.38
	21-24	34	–	0.18		21-24	24	–	0.54		21-24	21	–	0.43
	25-28	30	–	0.20		25-28	22	–	0.91		25-28	24	–	0.46
	29-32	16	–	0.50		29-32	12	–	0.83		29-32	11	–	0.64
	33-44	6	–	0.50		33-44	4	–	0.50		33-44	4	–	0.25
	45-64	14	–	0.50		45-64	10	–	1.00		45-64	10	–	0.80
21	13-16	26	0.00	–	21	13-16	18	0.00	–	21	13-16	18	0.11	–
	17-20	62	0.11	–		17-20	40	0.10	–		17-20	41	0.22	–
	21-24	34	–	0.38		21-24	24	–	0.62		21-24	21	–	0.57
	25-28	30	–	0.23		25-28	22	–	0.95		25-28	24	–	0.54
	29-32	16	–	0.50		29-32	12	–	0.83		29-32	11	–	0.64
	33-44	6	–	0.50		33-44	4	–	0.50		33-44	4	–	0.25
	45-64	14	–	0.50		45-64	10	–	1.00		45-64	10	–	0.80
25	13-16	26	0.00	–	25	13-16	18	0.00	–	25	13-16	18	0.00	–
	17-20	62	0.03	–		17-20	40	0.02	–		17-20	41	0.05	–
	21-24	34	0.26	–		21-24	24	0.12	–		21-24	21	0.14	–
	25-28	30	–	0.40		25-28	22	–	1.00		25-28	24	–	0.58
	29-32	16	–	0.50		29-32	12	–	0.83		29-32	11	–	0.64
	33-44	6	–	0.50		33-44	4	–	0.50		33-44	4	–	0.25
	45-64	14	–	0.50		45-64	10	–	1.00		45-64	10	–	0.80
30	13-16	26	0.00	–	30	13-16	18	0.00	–	30	13-16	18	0.00	–
	17-20	62	0.00	–		17-20	40	0.00	–		17-20	41	0.00	–
	21-24	34	0.00	–		21-24	24	0.00	–		21-24	21	0.00	–
	25-28	30	0.03	–		25-28	22	0.00	–		25-28	24	0.17	–
	29-32	16	0.38	0.62		29-32	12	0.00	1.00		29-32	11	0.17	0.80
	33-44	6	–	0.67		33-44	4	–	0.50		33-44	4	–	0.25
	45-64	14	–	0.50		45-64	10	–	1.00		45-64	10	–	0.80
		Smile					**Mustache**					**Beard**		
13	13-16	18	–	0.44	13	13-16	26	–	1.00	13	13-16	26	–	1.00
	17-20	43	–	0.40		17-20	54	–	0.98		17-20	54	–	1.00
	21-24	26	–	0.35		21-24	26	–	1.00		21-24	26	–	1.00
	25-28	25	–	0.52		25-28	23	–	1.00		25-28	24	–	1.00
	29-32	9	–	0.44		29-32	14	–	1.00		29-32	14	–	1.00
	33-44	6	–	0.50		33-44	4	–	1.00		33-44	2	–	1.00
	45-64	10	–	0.60		45-64	14	–	1.00		45-64	14	–	1.00
16	13-16	18	0.38	0.30	16	13-16	26	0.00	1.00	16	13-16	26	0.00	1.00
	17-20	43	–	0.40		17-20	54	–	0.98		17-20	54	–	1.00
	21-24	26	–	0.35		21-24	26	–	1.00		21-24	26	–	1.00
	25-28	25	–	0.52		25-28	23	–	1.00		25-28	24	–	1.00
	29-32	9	–	0.44		29-32	14	–	1.00		29-32	14	–	1.00
	33-44	6	–	0.50		33-44	4	–	1.00		33-44	2	–	1.00
	45-64	10	–	0.60		45-64	14	–	1.00		45-64	14	–	1.00
18	13-16	18	0.39	–	18	13-16	26	0.00	–	18	13-16	26	0.00	–
	17-20	43	0.50	0.46		17-20	54	0.00	0.98		17-20	54	0.00	1.00
	21-24	26	–	0.35		21-24	26	–	1.00		21-24	26	–	1.00
	25-28	25	–	0.52		25-28	23	–	1.00		25-28	24	–	1.00
	29-32	9	–	0.44		29-32	14	–	1.00		29-32	14	–	1.00
	33-44	6	–	0.50		33-44	4	–	1.00		33-44	2	–	1.00

Continued on next page

Table 24 – continued from previous page

Thr.	Age Group	N	FPR	FNR	Thr.	Age Group	N	FPR	FNR	Thr.	Age Group	N	FPR	FNR
	45-64	10	–	0.60		45-64	14	–	1.00		45-64	14	–	1.00
21	13-16	18	0.06	–	21	13-16	26	0.00	–	21	13-16	26	0.00	–
	17-20	43	0.28	–		17-20	54	0.02	–		17-20	54	0.00	–
	21-24	26	–	0.54		21-24	26	–	1.00		21-24	26	–	1.00
	25-28	25	–	0.56		25-28	23	–	1.00		25-28	24	–	1.00
	29-32	9	–	0.44		29-32	14	–	1.00		29-32	14	–	1.00
	33-44	6	–	0.50		33-44	4	–	1.00		33-44	2	–	1.00
	45-64	10	–	0.60		45-64	14	–	1.00		45-64	14	–	1.00
25	13-16	18	0.00	–	25	13-16	26	0.00	–	25	13-16	26	0.00	–
	17-20	43	0.07	–		17-20	54	0.00	–		17-20	54	0.00	–
	21-24	26	0.08	–		21-24	26	0.00	–		21-24	26	0.00	–
	25-28	25	–	0.76		25-28	23	–	1.00		25-28	24	–	1.00
	29-32	9	–	0.44		29-32	14	–	1.00		29-32	14	–	1.00
	33-44	6	–	0.50		33-44	4	–	1.00		33-44	2	–	1.00
	45-64	10	–	0.60		45-64	14	–	1.00		45-64	14	–	1.00
30	13-16	18	0.00	–	30	13-16	26	0.00	–	30	13-16	26	0.00	–
	17-20	43	0.00	–		17-20	54	0.00	–		17-20	54	0.00	–
	21-24	26	0.00	–		21-24	26	0.00	–		21-24	26	0.00	–
	25-28	25	0.00	–		25-28	23	0.00	–		25-28	24	0.00	–
	29-32	9	0.00	0.67		29-32	14	0.00	1.00		29-32	14	0.00	1.00
	33-44	6	–	0.50		33-44	4	–	1.00		33-44	2	–	1.00
	45-64	10	–	0.60		45-64	14	–	1.00		45-64	14	–	1.00
	Glasses					**KN95**					**Groucho**			
13	13-16	26	–	0.12	13	13-16	22	–	1.00	13	13-16	22	–	1.00
	17-20	62	–	0.23		17-20	53	–	0.98		17-20	58	–	1.00
	21-24	33	–	0.27		21-24	34	–	1.00		21-24	29	–	1.00
	25-28	30	–	0.27		25-28	30	–	1.00		25-28	30	–	1.00
	29-32	16	–	0.38		29-32	16	–	1.00		29-32	12	–	1.00
	33-44	6	–	0.67		33-44	5	–	1.00		33-44	4	–	1.00
	45-64	14	–	0.50		45-64	12	–	1.00		45-64	10	–	1.00
16	13-16	26	0.50	0.08	16	13-16	22	0.00	1.00	16	13-16	22	0.00	1.00
	17-20	62	–	0.23		17-20	53	–	1.00		17-20	58	–	1.00
	21-24	33	–	0.27		21-24	34	–	1.00		21-24	29	–	1.00
	25-28	30	–	0.27		25-28	30	–	1.00		25-28	30	–	1.00
	29-32	16	–	0.38		29-32	16	–	1.00		29-32	12	–	1.00
	33-44	6	–	0.67		33-44	5	–	1.00		33-44	4	–	1.00
	45-64	14	–	0.50		45-64	12	–	1.00		45-64	10	–	1.00
18	13-16	26	0.19	–	18	13-16	22	0.00	–	18	13-16	22	0.00	–
	17-20	62	0.00	0.33		17-20	53	0.00	1.00		17-20	58	0.00	1.00
	21-24	33	–	0.27		21-24	34	–	1.00		21-24	29	–	1.00
	25-28	30	–	0.27		25-28	30	–	1.00		25-28	30	–	1.00
	29-32	16	–	0.38		29-32	16	–	1.00		29-32	12	–	1.00
	33-44	6	–	0.67		33-44	5	–	1.00		33-44	4	–	1.00
	45-64	14	–	0.50		45-64	12	–	1.00		45-64	10	–	1.00
21	13-16	26	0.04	–	21	13-16	22	0.00	–	21	13-16	22	0.00	–
	17-20	62	0.15	–		17-20	53	0.00	–		17-20	58	0.00	–
	21-24	33	–	0.64		21-24	34	–	1.00		21-24	29	–	1.00
	25-28	30	–	0.47		25-28	30	–	1.00		25-28	30	–	1.00
	29-32	16	–	0.38		29-32	16	–	1.00		29-32	12	–	1.00
	33-44	6	–	0.67		33-44	5	–	1.00		33-44	4	–	1.00
	45-64	14	–	0.50		45-64	12	–	1.00		45-64	10	–	1.00
25	13-16	26	0.00	–	25	13-16	22	0.00	–	25	13-16	22	0.00	–
	17-20	62	0.05	–		17-20	53	0.00	–		17-20	58	0.00	–
	21-24	33	0.21	–		21-24	34	0.00	–		21-24	29	0.00	–
	25-28	30	–	0.53		25-28	30	–	1.00		25-28	30	–	1.00
	29-32	16	–	0.38		29-32	16	–	1.00		29-32	12	–	1.00
	33-44	6	–	0.67		33-44	5	–	1.00		33-44	4	–	1.00
	45-64	14	–	0.50		45-64	12	–	1.00		45-64	10	–	1.00
30	13-16	26	0.00	–	30	13-16	22	0.00	–	30	13-16	22	0.00	–
	17-20	62	0.00	–		17-20	53	0.00	–		17-20	58	0.00	–
	21-24	33	0.00	–		21-24	34	0.00	–		21-24	29	0.00	–
	25-28	30	0.20	–		25-28	30	0.00	–		25-28	30	0.00	–
	29-32	16	0.25	0.38		29-32	16	0.00	1.00		29-32	12	0.00	1.00
	33-44	6	–	0.67		33-44	5	–	1.00		33-44	4	–	1.00
	45-64	14	–	0.50		45-64	12	–	1.00		45-64	10	–	1.00
	Eyepatch					**Wig**					**Costume Mask**			
13	13-16	23	–	1.00	13	13-16	25	–	1.00	13	13-16	17	–	1.00
	17-20	60	–	1.00		17-20	59	–	1.00		17-20	52	–	1.00
	21-24	34	–	1.00		21-24	34	–	1.00		21-24	23	–	1.00
	25-28	28	–	1.00		25-28	28	–	1.00		25-28	22	–	1.00
	29-32	15	–	1.00		29-32	16	–	1.00		29-32	8	–	1.00
	33-44	6	–	1.00		33-44	6	–	1.00		33-44	3	–	1.00

Continued on next page

Table 24 – continued from previous page

Thr.	Age Group	N	FPR	FNR	Thr.	Age Group	N	FPR	FNR	Thr.	Age Group	N	FPR	FNR
	45-64	12	–	1.00		45-64	14	–	1.00		45-64	10	–	0.90
16	13-16	23	0.00	1.00	16	13-16	25	0.00	1.00	16	13-16	17	0.00	1.00
	17-20	60	–	1.00		17-20	59	–	1.00		17-20	52	–	1.00
	21-24	34	–	1.00		21-24	34	–	1.00		21-24	23	–	1.00
	25-28	28	–	1.00		25-28	28	–	1.00		25-28	22	–	1.00
	29-32	15	–	1.00		29-32	16	–	1.00		29-32	8	–	1.00
	33-44	6	–	1.00		33-44	6	–	1.00		33-44	3	–	1.00
	45-64	12	–	1.00		45-64	14	–	1.00		45-64	10	–	0.90
18	13-16	23	0.00	–	18	13-16	25	0.00	–	18	13-16	17	0.00	–
	17-20	60	0.00	1.00		17-20	59	0.00	1.00		17-20	52	0.00	1.00
	21-24	34	–	1.00		21-24	34	–	1.00		21-24	23	–	1.00
	25-28	28	–	1.00		25-28	28	–	1.00		25-28	22	–	1.00
	29-32	15	–	1.00		29-32	16	–	1.00		29-32	8	–	1.00
	33-44	6	–	1.00		33-44	6	–	1.00		33-44	3	–	1.00
	45-64	12	–	1.00		45-64	14	–	1.00		45-64	10	–	0.90
21	13-16	23	0.00	–	21	13-16	25	0.00	–	21	13-16	17	0.00	–
	17-20	60	0.00	–		17-20	59	0.00	–		17-20	52	0.00	–
	21-24	34	–	1.00		21-24	34	–	1.00		21-24	23	–	1.00
	25-28	28	–	1.00		25-28	28	–	1.00		25-28	22	–	1.00
	29-32	15	–	1.00		29-32	16	–	1.00		29-32	8	–	1.00
	33-44	6	–	1.00		33-44	6	–	1.00		33-44	3	–	1.00
	45-64	12	–	1.00		45-64	14	–	1.00		45-64	10	–	0.90
25	13-16	23	0.00	–	25	13-16	25	0.00	–	25	13-16	17	0.00	–
	17-20	60	0.00	–		17-20	59	0.00	–		17-20	52	0.00	–
	21-24	34	0.00	–		21-24	34	0.00	–		21-24	23	0.00	–
	25-28	28	–	1.00		25-28	28	–	1.00		25-28	22	–	1.00
	29-32	15	–	1.00		29-32	16	–	1.00		29-32	8	–	1.00
	33-44	6	–	1.00		33-44	6	–	1.00		33-44	3	–	1.00
	45-64	12	–	1.00		45-64	14	–	1.00		45-64	10	–	0.90
30	13-16	23	0.00	–	30	13-16	25	0.00	–	30	13-16	17	0.00	–
	17-20	60	0.00	–		17-20	59	0.00	–		17-20	52	0.00	–
	21-24	34	0.00	–		21-24	34	0.00	–		21-24	23	0.00	–
	25-28	28	0.00	–		25-28	28	0.00	–		25-28	22	0.00	–
	29-32	15	0.00	1.00		29-32	16	0.00	1.00		29-32	8	0.00	1.00
	33-44	6	–	1.00		33-44	6	–	1.00		33-44	3	–	1.00
	45-64	12	–	1.00		45-64	14	–	1.00		45-64	10	–	0.90

Y Joint age and liveness prediction, FACETEC

Table 25: Joint FACETEC age estimation and liveness performance by age threshold, age group, and test condition. FPR and FNR are reported as proportions.

Thr.	Age Group	*N*	FPR	FNR	Thr.	Age Group	*N*	FPR	FNR	Thr.	Age Group	*N*	FPR	FNR
	Baseline					**Open-mouth**					**Frown**			
13	13-16	26	–	0.62	13	13-16	26	–	0.77	13	13-16	26	–	0.69
	17-20	62	–	0.98		17-20	62	–	1.00		17-20	62	–	1.00
	21-24	34	–	1.00		21-24	34	–	1.00		21-24	34	–	1.00
	25-28	30	–	1.00		25-28	30	–	1.00		25-28	30	–	1.00
	29-32	16	–	1.00		29-32	16	–	1.00		29-32	16	–	1.00
	33-44	6	–	1.00		33-44	6	–	1.00		33-44	6	–	1.00
	45-64	14	–	1.00		45-64	14	–	1.00		45-64	14	–	1.00
16	13-16	26	0.07	0.75	16	13-16	26	0.07	0.75	16	13-16	26	0.07	0.58
	17-20	62	–	0.79		17-20	62	–	0.89		17-20	62	–	0.71
	21-24	34	–	1.00		21-24	34	–	1.00		21-24	34	–	1.00
	25-28	30	–	1.00		25-28	30	–	1.00		25-28	30	–	1.00
	29-32	16	–	1.00		29-32	16	–	1.00		29-32	16	–	1.00
	33-44	6	–	1.00		33-44	6	–	1.00		33-44	6	–	1.00
	45-64	14	–	1.00		45-64	14	–	1.00		45-64	14	–	1.00
18	13-16	26	0.00	–	18	13-16	26	0.04	–	18	13-16	26	0.08	–
	17-20	62	0.00	0.88		17-20	62	0.00	0.95		17-20	62	0.00	0.83
	21-24	34	–	0.68		21-24	34	–	0.91		21-24	34	–	0.85
	25-28	30	–	0.63		25-28	30	–	1.00		25-28	30	–	0.80
	29-32	16	–	0.62		29-32	16	–	0.75		29-32	16	–	0.81
	33-44	6	–	1.00		33-44	6	–	1.00		33-44	6	–	1.00
	45-64	14	–	0.50		45-64	14	–	0.86		45-64	14	–	0.71
21	13-16	26	0.00	–	21	13-16	26	0.00	–	21	13-16	26	0.04	–
	17-20	62	0.08	–		17-20	62	0.02	–		17-20	62	0.03	–
	21-24	34	–	0.79		21-24	34	–	0.91		21-24	34	–	0.94
	25-28	30	–	0.73		25-28	30	–	0.97		25-28	30	–	0.90
	29-32	16	–	0.69		29-32	16	–	0.75		29-32	16	–	0.81
	33-44	6	–	1.00		33-44	6	–	1.00		33-44	6	–	1.00
	45-64	14	–	0.50		45-64	14	–	1.00		45-64	14	–	0.64
25	13-16	26	0.00	–	25	13-16	26	0.00	–	25	13-16	26	0.00	–
	17-20	62	0.02	–		17-20	62	0.02	–		17-20	62	0.05	–
	21-24	34	0.03	–		21-24	34	0.06	–		21-24	34	0.09	–
	25-28	30	–	0.90		25-28	30	–	1.00		25-28	30	–	0.87
	29-32	16	–	0.69		29-32	16	–	0.88		29-32	16	–	0.88
	33-44	6	–	1.00		33-44	6	–	1.00		33-44	6	–	1.00
	45-64	14	–	0.50		45-64	14	–	0.93		45-64	14	–	0.64
30	13-16	26	0.00	–	30	13-16	26	0.00	–	30	13-16	26	0.00	–
	17-20	62	0.00	–		17-20	62	0.00	–		17-20	62	0.00	–
	21-24	34	0.00	–		21-24	34	0.00	–		21-24	34	0.00	–
	25-28	30	0.03	–		25-28	30	0.00	–		25-28	30	0.03	–
	29-32	16	0.12	0.88		29-32	16	0.00	0.88		29-32	16	0.00	0.88
	33-44	6	–	1.00		33-44	6	–	1.00		33-44	6	–	1.00
	45-64	14	–	0.57		45-64	14	–	1.00		45-64	14	–	0.64
	Smile					**Squint**					**Mustache**			
13	13-16	26	–	0.96	13	13-16	26	–	1.00	13	13-16	26	–	0.88
	17-20	62	–	1.00		17-20	62	–	1.00		17-20	62	–	1.00
	21-24	34	–	1.00		21-24	34	–	1.00		21-24	34	–	1.00
	25-28	30	–	1.00		25-28	30	–	1.00		25-28	30	–	1.00
	29-32	16	–	1.00		29-32	16	–	1.00		29-32	16	–	1.00
	33-44	6	–	1.00		33-44	6	–	1.00		33-44	6	–	1.00
	45-64	14	–	1.00		45-64	14	–	1.00		45-64	14	–	1.00
16	13-16	26	0.00	1.00	16	13-16	26	0.00	1.00	16	13-16	26	0.07	0.75
	17-20	62	–	0.97		17-20	62	–	1.00		17-20	62	–	0.94
	21-24	34	–	1.00		21-24	34	–	1.00		21-24	34	–	1.00
	25-28	30	–	1.00		25-28	30	–	1.00		25-28	30	–	1.00
	29-32	16	–	1.00		29-32	16	–	1.00		29-32	16	–	1.00
	33-44	6	–	1.00		33-44	6	–	1.00		33-44	6	–	1.00
	45-64	14	–	1.00		45-64	14	–	1.00		45-64	14	–	1.00
18	13-16	26	0.00	–	18	13-16	26	0.00	–	18	13-16	26	0.08	–
	17-20	62	0.00	0.98		17-20	62	0.00	1.00		17-20	62	0.00	0.92
	21-24	34	–	1.00		21-24	34	–	1.00		21-24	34	–	0.97
	25-28	30	–	1.00		25-28	30	–	1.00		25-28	30	–	0.90
	29-32	16	–	0.94		29-32	16	–	1.00		29-32	16	–	0.94
	33-44	6	–	1.00		33-44	6	–	1.00		33-44	6	–	1.00

Continued on next page

Table 25 – continued from previous page

Thr.	Age Group	N	FPR	FNR	Thr.	Age Group	N	FPR	FNR	Thr.	Age Group	N	FPR	FNR
	45-64	14	–	0.93		45-64	14	–	1.00		45-64	14	–	0.86
21	13-16	26	0.00	–	21	13-16	26	0.00	–	21	13-16	26	0.04	–
	17-20	62	0.00	–		17-20	62	0.00	–		17-20	62	0.05	–
	21-24	34	–	1.00		21-24	34	–	1.00		21-24	34	–	1.00
	25-28	30	–	1.00		25-28	30	–	1.00		25-28	30	–	1.00
	29-32	16	–	0.94		29-32	16	–	1.00		29-32	16	–	0.94
	33-44	6	–	1.00		33-44	6	–	1.00		33-44	6	–	1.00
	45-64	14	–	0.93		45-64	14	–	1.00		45-64	14	–	0.86
25	13-16	26	0.00	–	25	13-16	26	0.00	–	25	13-16	26	0.00	–
	17-20	62	0.00	–		17-20	62	0.00	–		17-20	62	0.00	–
	21-24	34	0.00	–		21-24	34	0.00	–		21-24	34	0.00	–
	25-28	30	–	1.00		25-28	30	–	1.00		25-28	30	–	1.00
	29-32	16	–	0.94		29-32	16	–	1.00		29-32	16	–	0.88
	33-44	6	–	1.00		33-44	6	–	1.00		33-44	6	–	1.00
	45-64	14	–	0.93		45-64	14	–	1.00		45-64	14	–	0.86
30	13-16	26	0.00	–	30	13-16	26	0.00	–	30	13-16	26	0.00	–
	17-20	62	0.00	–		17-20	62	0.00	–		17-20	62	0.00	–
	21-24	34	0.00	–		21-24	34	0.00	–		21-24	34	0.00	–
	25-28	30	0.00	–		25-28	30	0.00	–		25-28	30	0.00	–
	29-32	16	0.00	1.00		29-32	16	0.00	1.00		29-32	16	0.00	1.00
	33-44	6	–	1.00		33-44	6	–	1.00		33-44	6	–	1.00
	45-64	14	–	0.93		45-64	14	–	1.00		45-64	14	–	0.86
	Beard					**Glasses**					**KN95**			
13	13-16	26	–	1.00	13	13-16	26	–	0.92	13	13-16	26	–	1.00
	17-20	62	–	1.00		17-20	62	–	1.00		17-20	62	–	1.00
	21-24	34	–	1.00		21-24	34	–	1.00		21-24	34	–	1.00
	25-28	30	–	1.00		25-28	30	–	1.00		25-28	30	–	1.00
	29-32	16	–	1.00		29-32	16	–	1.00		29-32	16	–	1.00
	33-44	6	–	1.00		33-44	6	–	1.00		33-44	6	–	1.00
	45-64	14	–	1.00		45-64	14	–	1.00		45-64	14	–	1.00
16	13-16	26	0.00	1.00	16	13-16	26	0.07	0.75	16	13-16	26	0.00	1.00
	17-20	62	–	0.98		17-20	62	–	0.76		17-20	62	–	1.00
	21-24	34	–	1.00		21-24	34	–	1.00		21-24	34	–	1.00
	25-28	30	–	1.00		25-28	30	–	1.00		25-28	30	–	1.00
	29-32	16	–	1.00		29-32	16	–	1.00		29-32	16	–	1.00
	33-44	6	–	1.00		33-44	6	–	1.00		33-44	6	–	1.00
	45-64	14	–	1.00		45-64	14	–	1.00		45-64	14	–	1.00
18	13-16	26	0.00	–	18	13-16	26	0.00	–	18	13-16	26	0.00	–
	17-20	62	0.00	0.98		17-20	62	0.00	0.82		17-20	62	0.00	1.00
	21-24	34	–	1.00		21-24	34	–	0.79		21-24	34	–	1.00
	25-28	30	–	0.83		25-28	30	–	0.77		25-28	30	–	1.00
	29-32	16	–	1.00		29-32	16	–	0.94		29-32	16	–	1.00
	33-44	6	–	1.00		33-44	6	–	0.83		33-44	6	–	1.00
	45-64	14	–	1.00		45-64	14	–	0.64		45-64	14	–	1.00
21	13-16	26	0.00	–	21	13-16	26	0.00	–	21	13-16	26	0.00	–
	17-20	62	0.02	–		17-20	62	0.10	–		17-20	62	0.00	–
	21-24	34	–	1.00		21-24	34	–	0.76		21-24	34	–	1.00
	25-28	30	–	0.97		25-28	30	–	0.80		25-28	30	–	1.00
	29-32	16	–	1.00		29-32	16	–	0.94		29-32	16	–	1.00
	33-44	6	–	1.00		33-44	6	–	1.00		33-44	6	–	1.00
	45-64	14	–	1.00		45-64	14	–	0.71		45-64	14	–	1.00
25	13-16	26	0.00	–	25	13-16	26	0.00	–	25	13-16	26	0.00	–
	17-20	62	0.00	–		17-20	62	0.05	–		17-20	62	0.00	–
	21-24	34	0.00	–		21-24	34	0.06	–		21-24	34	0.00	–
	25-28	30	–	0.97		25-28	30	–	0.90		25-28	30	–	1.00
	29-32	16	–	1.00		29-32	16	–	0.88		29-32	16	–	1.00
	33-44	6	–	1.00		33-44	6	–	1.00		33-44	6	–	1.00
	45-64	14	–	1.00		45-64	14	–	0.64		45-64	14	–	1.00
30	13-16	26	0.00	–	30	13-16	26	0.00	–	30	13-16	26	0.00	–
	17-20	62	0.00	–		17-20	62	0.02	–		17-20	62	0.00	–
	21-24	34	0.00	–		21-24	34	0.00	–		21-24	34	0.00	–
	25-28	30	0.00	–		25-28	30	0.07	–		25-28	30	0.00	–
	29-32	16	0.00	1.00		29-32	16	0.00	1.00		29-32	16	0.00	1.00
	33-44	6	–	1.00		33-44	6	–	1.00		33-44	6	–	1.00
	45-64	14	–	1.00		45-64	14	–	0.64		45-64	14	–	1.00
	Groucho					**Eyepatch**					**Wig**			
13	13-16	26	–	1.00	13	13-16	26	–	1.00	13	13-16	26	–	0.62
	17-20	62	–	1.00		17-20	62	–	1.00		17-20	62	–	1.00
	21-24	34	–	1.00		21-24	34	–	1.00		21-24	34	–	1.00
	25-28	30	–	1.00		25-28	30	–	1.00		25-28	30	–	1.00
	29-32	16	–	1.00		29-32	16	–	1.00		29-32	16	–	1.00
	33-44	6	–	1.00		33-44	6	–	1.00		33-44	6	–	1.00

Continued on next page

Table 25 – continued from previous page

Thr.	Age Group	N	FPR	FNR	Thr.	Age Group	N	FPR	FNR	Thr.	Age Group	N	FPR	FNR
	45-64	14	–	1.00		45-64	14	–	1.00		45-64	14	–	1.00
16	13-16	26	0.00	1.00	16	13-16	26	0.00	1.00	16	13-16	26	0.00	0.75
	17-20	62	–	1.00		17-20	62	–	1.00		17-20	62	–	0.66
	21-24	34	–	1.00		21-24	34	–	1.00		21-24	34	–	1.00
	25-28	30	–	1.00		25-28	30	–	1.00		25-28	30	–	1.00
	29-32	16	–	1.00		29-32	16	–	1.00		29-32	16	–	1.00
	33-44	6	–	1.00		33-44	6	–	1.00		33-44	6	–	1.00
	45-64	14	–	1.00		45-64	14	–	1.00		45-64	14	–	1.00
18	13-16	26	0.00	–	18	13-16	26	0.00	–	18	13-16	26	0.00	–
	17-20	62	0.00	1.00		17-20	62	0.00	1.00		17-20	62	0.00	0.87
	21-24	34	–	1.00		21-24	34	–	1.00		21-24	34	–	0.82
	25-28	30	–	1.00		25-28	30	–	1.00		25-28	30	–	0.90
	29-32	16	–	1.00		29-32	16	–	1.00		29-32	16	–	0.69
	33-44	6	–	1.00		33-44	6	–	1.00		33-44	6	–	1.00
	45-64	14	–	1.00		45-64	14	–	1.00		45-64	14	–	0.71
21	13-16	26	0.00	–	21	13-16	26	0.00	–	21	13-16	26	0.00	–
	17-20	62	0.00	–		17-20	62	0.00	–		17-20	62	0.03	–
	21-24	34	–	1.00		21-24	34	–	1.00		21-24	34	–	0.91
	25-28	30	–	1.00		25-28	30	–	1.00		25-28	30	–	0.93
	29-32	16	–	1.00		29-32	16	–	1.00		29-32	16	–	0.62
	33-44	6	–	1.00		33-44	6	–	1.00		33-44	6	–	1.00
	45-64	14	–	1.00		45-64	14	–	1.00		45-64	14	–	0.79
25	13-16	26	0.00	–	25	13-16	26	0.00	–	25	13-16	26	0.00	–
	17-20	62	0.00	–		17-20	62	0.00	–		17-20	62	0.00	–
	21-24	34	0.00	–		21-24	34	0.00	–		21-24	34	0.09	–
	25-28	30	–	1.00		25-28	30	–	1.00		25-28	30	–	0.90
	29-32	16	–	1.00		29-32	16	–	1.00		29-32	16	–	0.75
	33-44	6	–	1.00		33-44	6	–	1.00		33-44	6	–	1.00
	45-64	14	–	1.00		45-64	14	–	1.00		45-64	14	–	0.71
30	13-16	26	0.00	–	30	13-16	26	0.00	–	30	13-16	26	0.00	–
	17-20	62	0.00	–		17-20	62	0.00	–		17-20	62	0.00	–
	21-24	34	0.00	–		21-24	34	0.00	–		21-24	34	0.00	–
	25-28	30	0.00	–		25-28	30	0.00	–		25-28	30	0.00	–
	29-32	16	0.00	1.00		29-32	16	0.00	1.00		29-32	16	0.00	0.88
	33-44	6	–	1.00		33-44	6	–	1.00		33-44	6	–	1.00
	45-64	14	–	1.00		45-64	14	–	1.00		45-64	14	–	0.71

Costume mask

13	13-16	26	–	1.00
	17-20	62	–	1.00
	21-24	34	–	1.00
	25-28	30	–	1.00
	29-32	16	–	1.00
	33-44	6	–	1.00
	45-64	14	–	1.00
16	13-16	26	0.00	1.00
	17-20	62	–	1.00
	21-24	34	–	1.00
	25-28	30	–	1.00
	29-32	16	–	1.00
	33-44	6	–	1.00
	45-64	14	–	1.00
18	13-16	26	0.00	–
	17-20	62	0.00	1.00
	21-24	34	–	1.00
	25-28	30	–	1.00
	29-32	16	–	1.00
	33-44	6	–	1.00
	45-64	14	–	1.00
21	13-16	26	0.00	–
	17-20	62	0.00	–
	21-24	34	–	1.00
	25-28	30	–	1.00
	29-32	16	–	1.00
	33-44	6	–	1.00
	45-64	14	–	1.00
25	13-16	26	0.00	–
	17-20	62	0.00	–
	21-24	34	0.00	–
	25-28	30	–	1.00
	29-32	16	–	1.00
	33-44	6	–	1.00
	45-64	14	–	1.00

Continued on next page

Table 25 – continued from previous page

Thr.	Age Group	*N*	FPR	FNR	Thr.	Age Group	*N*	FPR	FNR	Thr.	Age Group	*N*	FPR	FNR
30	13-16	26	0.00	–										
	17-20	62	0.00	–										
	21-24	34	0.00	–										
	25-28	30	0.00	–										
	29-32	16	0.00	1.00										
	33-44	6	–	1.00										
	45-64	14	–	1.00										

Z Gender x skin tone MAE, Yoti

Table 26: Mean absolute error (MAE) in Yoti age estimates by self-reported gender identity, skin tone, and test condition. Results pool iPhone and Android age estimates.

Gender	Skin Tone	*N*	MAE	Gender	Skin Tone	*N*	MAE	Gender	Skin Tone	*N*	MAE
	Baseline				**Gawk**				**Frown**		
Female	5	18	2.11	Female	5	14	2.57	Female	5	14	2.21
	6	28	2.82		6	28	4.07		6	28	3.71
	7	18	1.39		7	10	3.30		7	12	1.58
	8	4	4.25		8	4	5.25		8	4	5.50
Male	5	4	2.25	Male	5	4	1.75	Male	5	3	2.33
	6	16	1.50		6	10	1.60		6	9	3.11
	7	20	1.70		7	13	2.92		7	13	1.92
Non-binary	7	2	1.00	Non-binary	7	2	1.50	Non-binary	7	2	0.00
	8	2	1.50		8	2	1.00		8	2	2.00
Transgender	7	2	0.00	Transgender	7	2	0.00	Transgender	7	2	0.50
	Smile				**Mustache**				**Beard**		
Female	5	14	2.64	Female	5	18	2.22	Female	5	18	1.83
	6	28	4.07		6	28	3.18		6	28	3.68
	7	10	3.60		7	18	2.28		7	16	2.09
	8	4	4.75		8	4	2.50		8	2	7.50
Male	5	4	1.25	Male	5	4	2.50	Male	5	4	3.00
	6	9	1.89		6	16	2.00		6	16	2.56
	7	13	1.62		7	18	3.00		7	20	4.50
Non-binary	7	2	0.50	Non-binary	7	2	1.00	Non-binary	7	2	1.00
	8	2	3.00		8	2	1.50		8	2	1.00
Transgender	7	2	0.50	Transgender	7	2	0.50	Transgender	7	2	0.50
	Glasses				**KN95**				**Groucho**		
Female	5	18	2.33	Female	5	18	1.83	Female	5	18	3.06
	6	27	2.70		6	26	3.62		6	25	4.44
	7	18	2.28		7	13	2.54		7	13	2.46
	8	4	4.75		8	4	2.00		8	4	5.50
Male	5	4	3.00	Male	5	4	5.75	Male	5	2	3.00
	6	16	2.06		6	14	2.43		6	10	1.10
	7	20	2.80		7	16	4.94		7	15	1.53
Non-binary	7	2	1.00	Non-binary	8	2	4.00	Non-binary	7	2	0.50
	8	2	0.50						8	2	1.00
				Transgender	7	1	2.00				
Transgender	7	2	0.00					Transgender	7	2	0.00
	Eyepatch				**Wig**				**Costume mask**		
Female	5	15	2.47	Female	5	18	2.67	Female	5	13	47.46
	6	27	3.63		6	28	3.43		6	19	40.21
	7	18	1.72		7	16	2.44		7	13	46.23
	8	4	7.50		8	4	4.00		8	2	49.00
Male	5	4	1.25	Male	5	4	1.75	Male	5	2	42.00
	6	16	1.69		6	16	1.38		6	12	43.92
	7	19	1.53		7	20	2.20		7	14	37.93
Non-binary	7	1	1.00	Non-binary	7	1	1.00	Non-binary	7	2	51.00
	8	2	0.50		8	2	1.00		8	2	50.00
Transgender	7	2	0.50	Transgender	7	2	0.50	Transgender	7	2	49.00

AA Liveness x skin tone pass rates, Yoti

Table 27: Yoti liveness performance by skin tone and test condition. Pass proportions pool iPhone and Android observations; null/no-result and missing liveness responses are treated as failures.

Skin Tone	*N*	Pass	Fail	Pass Prop.	Skin Tone	*N*	Pass	Fail	Pass Prop.	Skin Tone	*N*	Pass	Fail	Pass Prop.
Baseline					**Open-mouth**					**Frown**				
5	22	14	8	0.64	5	22	5	17	0.23	5	22	11	11	0.50
6	44	29	15	0.66	6	44	8	36	0.18	6	44	22	22	0.50
7	42	25	17	0.60	7	42	10	32	0.24	7	42	13	29	0.31
8	6	2	4	0.33	8	6	0	6	0.00	8	6	1	5	0.17
Smile					**Squint**					**Mustache**				
5	22	13	9	0.59	5	22	0	22	0.00	5	22	0	22	0.00
6	44	20	24	0.45	6	44	0	44	0.00	6	44	0	44	0.00
7	42	15	27	0.36	7	42	0	42	0.00	7	42	1	41	0.02
8	6	3	3	0.50	8	6	0	6	0.00	8	6	0	6	0.00
Beard					**Glasses**					**KN95**				
5	22	0	22	0.00	5	22	17	5	0.77	5	22	1	21	0.05
6	44	0	44	0.00	6	44	32	12	0.73	6	44	0	44	0.00
7	42	0	42	0.00	7	42	29	13	0.69	7	42	0	42	0.00
8	6	0	6	0.00	8	6	0	6	0.00	8	6	0	6	0.00
Groucho					**Eyepatch**					**Wig**				
5	22	0	22	0.00	5	22	0	22	0.00	5	22	0	22	0.00
6	44	0	44	0.00	6	44	0	44	0.00	6	44	0	44	0.00
7	42	0	42	0.00	7	42	0	42	0.00	7	42	0	42	0.00
8	6	0	6	0.00	8	6	0	6	0.00	8	6	0	6	0.00
Costume Mask														
5	22	0	22	0.00										
6	44	0	44	0.00										
7	42	1	41	0.02										
8	6	0	6	0.00										

AB Liveness x skin tone pass rates, FACETEC

Table 28: FACETEC liveness performance by skin tone and test condition. Pass proportions pool iPhone and Android observations; null responses are treated as failures.

Skin Tone	*N*	Pass	Fail	Pass Prop.	Skin Tone	*N*	Pass	Fail	Pass Prop.	Skin Tone	*N*	Pass	Fail	Pass Prop.
Baseline					**Open-mouth**					**Frown**				
5	22	7	15	0.32	5	22	5	17	0.23	5	22	5	17	0.23
6	44	11	33	0.25	6	44	9	35	0.20	6	44	15	29	0.34
7	42	17	25	0.40	7	42	5	37	0.12	7	42	17	25	0.40
8	6	1	5	0.17	8	6	2	4	0.33	8	6	2	4	0.33
Smile					**Squint**					**Mustache**				
5	22	0	22	0.00	5	22	0	22	0.00	5	22	0	22	0.00
6	44	2	42	0.05	6	44	0	44	0.00	6	44	4	40	0.09
7	42	3	39	0.07	7	42	0	42	0.00	7	42	6	36	0.14
8	6	0	6	0.00	8	6	0	6	0.00	8	6	1	5	0.17
Beard					**Glasses**					**KN95**				
5	22	0	22	0.00	5	22	4	18	0.18	5	22	0	22	0.00
6	44	0	44	0.00	6	44	11	33	0.25	6	44	0	44	0.00
7	42	1	41	0.02	7	42	10	32	0.24	7	42	0	42	0.00
8	6	0	6	0.00	8	6	0	6	0.00	8	6	0	6	0.00
Groucho					**Eyepatch**					**Wig**				
5	22	0	22	0.00	5	22	0	22	0.00	5	22	10	12	0.45
6	44	0	44	0.00	6	44	0	44	0.00	6	44	17	27	0.39
7	42	0	42	0.00	7	42	0	42	0.00	7	42	15	27	0.36
8	6	0	6	0.00	8	6	0	6	0.00	8	6	4	2	0.67
Costume Mask														
5	22	0	22	0.00										
6	44	0	44	0.00										
7	42	0	42	0.00										
8	6	0	6	0.00										

AC Age band x skin tone MAE, Yoti

Table 29: Mean absolute error by age bin, skin tone, and test condition for Yoti.

Condition	Age	Skin	*N*	MAE	Condition	Age	Skin	*N*	MAE	Condition	Age	Skin	*N*	MAE
Baseline	13–16	5	4	1.50	Frown	13–16	5	4	2.25	Mustache	13–16	5	4	2.25
	13–16	6	8	1.88		13–16	6	6	2.83		13–16	6	8	1.88
	13–16	7	6	2.17		13–16	7	4	3.00		13–16	7	6	2.00
	17–20	5	10	2.10		17–20	5	6	2.00		17–20	5	10	2.70
	17–20	6	12	2.08		17–20	6	10	3.00		17–20	6	12	1.58
	17–20	7	20	0.70		17–20	7	15	0.93		17–20	7	18	1.39
	17–20	8	4	3.50		17–20	8	4	5.00		17–20	8	4	2.75
	21–24	5	4	2.25		21–24	5	3	2.33		21–24	5	4	2.00
	21–24	6	10	2.60		21–24	6	8	2.50		21–24	6	10	3.40
	21–24	7	4	3.00		21–24	7	2	4.50		21–24	7	4	3.00
	25–28	6	6	3.17		25–28	6	6	4.83		25–28	6	6	3.83
	29–32	5	4	2.75		29–32	5	4	2.50		29–32	5	4	1.50
	29–32	6	4	0.50		29–32	6	3	2.00		29–32	6	4	1.00
	29–32	7	2	2.50		29–32	7	2	1.50		29–32	7	2	4.00
	33–44	7	2	1.00		33–44	7	2	0.50		33–44	7	2	3.50
	33–44	8	2	3.00		33–44	8	2	3.00		33–44	8	2	1.00
	45–64	6	4	4.00		45–64	6	4	7.50		45–64	6	4	6.50
	45–64	7	8	1.88		45–64	7	4	1.50		45–64	7	8	4.25
Open-mouth	13–16	5	4	2.50	Smile	13–16	5	4	2.50	Beard	13–16	5	4	2.75
	13–16	6	6	2.17		13–16	6	6	2.67		13–16	6	8	2.12
	13–16	7	4	3.75		13–16	7	4	3.50		13–16	7	6	2.17
	17–20	5	6	1.33		17–20	5	6	4.00		17–20	5	10	2.00
	17–20	6	10	2.40		17–20	6	10	3.40		17–20	6	12	2.67
	17–20	7	13	1.38		17–20	7	15	1.33		17–20	7	18	2.00
	17–20	8	4	3.00		17–20	8	4	3.50		17–20	8	4	4.25
	21–24	5	4	1.00		21–24	5	4	0.75		21–24	5	4	2.50
	21–24	6	8	3.50		21–24	6	8	3.38		21–24	6	10	3.00
	21–24	7	2	5.00		21–24	7	2	3.50		21–24	7	4	5.75
	25–28	6	6	5.50		25–28	6	6	3.00		25–28	6	6	4.00
	29–32	5	4	5.25		29–32	5	4	1.25		29–32	5	4	1.00
	29–32	6	4	1.00		29–32	6	3	1.00		29–32	6	4	1.25
	29–32	7	2	2.00		29–32	7	–	–		29–32	7	2	1.50
	33–44	7	2	5.50		33–44	7	2	6.50		33–44	7	2	5.25
	33–44	8	2	5.50		33–44	8	2	5.50		33–44	8	–	–
	45–64	6	4	7.00		45–64	6	4	8.25		45–64	6	4	9.00
	45–64	7	4	4.00		45–64	7	4	1.25		45–64	7	8	5.12
Glasses	13–16	5	4	1.50	KN95	13–16	5	4	4.50	Eyepatch	13–16	5	2	1.00
	13–16	6	8	2.00		13–16	6	6	2.50		13–16	6	7	1.43
	13–16	7	6	2.83		13–16	7	4	2.50		13–16	7	6	2.67
	17–20	5	10	1.50		17–20	5	10	1.40		17–20	5	10	2.20
	17–20	6	12	1.83		17–20	6	10	2.20		17–20	6	12	1.92
	17–20	7	20	1.10		17–20	7	13	2.31		17–20	7	19	0.68
	17–20	8	4	3.50		17–20	8	4	3.50		17–20	8	4	4.50
	21–24	5	4	5.00		21–24	5	4	3.75		21–24	5	4	1.75
	21–24	6	9	2.44		21–24	6	10	2.60		21–24	6	10	3.10
	21–24	7	4	3.50		21–24	7	4	4.00		21–24	7	4	3.50
	25–28	6	6	4.17		25–28	6	6	3.17		25–28	6	6	4.33
	29–32	5	4	3.25		29–32	5	4	2.25		29–32	5	3	3.67
	29–32	6	4	1.25		29–32	6	4	2.00		29–32	6	4	0.75
	29–32	7	2	4.50		29–32	7	2	1.50		29–32	7	2	2.00
	33–44	7	2	2.00		33–44	7	1	6.00		33–44	7	2	2.00
	33–44	8	2	3.00		33–44	8	2	1.00		33–44	8	2	6.50
	45–64	6	4	4.00		45–64	6	4	9.50		45–64	6	4	8.00
	45–64	7	8	4.12		45–64	7	6	8.17		45–64	7	7	1.57
Groucho	13–16	5	2	2.50	Wig	13–16	5	4	1.75	Costume mask	13–16	5	4	47.75
	13–16	6	8	1.88		13–16	6	8	2.12		13–16	6	3	54.00
	13–16	7	6	2.33		13–16	7	6	2.67		13–16	7	5	54.80
	17–20	5	10	1.20		17–20	5	10	3.20		17–20	5	6	49.50
	17–20	6	10	1.90		17–20	6	12	2.42		17–20	6	12	48.67
	17–20	7	18	1.11		17–20	7	17	0.82		17–20	7	17	47.41
	17–20	8	4	0.50		17–20	8	4	1.25		17–20	8	4	49.50
	21–24	5	4	2.75		21–24	5	4	2.50		21–24	5	2	48.00
	21–24	6	7	1.71		21–24	6	10	3.90		21–24	6	4	44.50
	21–24	7	2	1.50		21–24	7	4	3.00		21–24	7	4	44.00
	25–28	6	6	6.67		25–28	6	6	2.67		25–28	6	5	43.20
	29–32	5	4	8.25		29–32	5	4	1.50		29–32	5	3	39.00

Continued on next page

Table 29 – continued from previous page

Condition	Age	Skin	*N*	MAE	Condition	Age	Skin	*N*	MAE	Condition	Age	Skin	*N*	MAE
	29–32	6	–	–		29–32	6	4	1.00		29–32	6	3	33.67
	29–32	7	2	5.00		29–32	7	2	2.50		29–32	7	–	–
	33–44	7	–	–		33–44	7	2	6.00		33–44	7	1	29.00
	33–44	8	2	11.00		33–44	8	2	6.50		33–44	8	–	–
	45–64	6	4	9.00		45–64	6	4	3.25		45–64	6	4	12.50
	45–64	7	4	2.25		45–64	7	8	3.25		45–64	7	4	11.75

AD MAE by race and test condition, YOTI

Table 30: Mean absolute error by self-reported racial group and test condition for YOTI.

Condition	Race	*N*	MAE	Condition	Race	*N*	MAE	Condition	Race	*N*	MAE
	AI/AN	6	1.835		AI/AN	6	1.83		AI/AN	6	1.33
	AS	92	1.965		AS	64	2.55		AS	75	2.54
	BL/AA	12	2.5		BL/AA	12	3		BL/AA	12	2.42
	HI/LA	20	2.3		HI/LA	10	3.6		HI/LA	18	1.78
Baseline	ME/NA	2	–	Frown	ME/NA	–	–	Mustache	ME/NA	–	–
	MR	26	2.27		MR	19	2.68		MR	23	1.93
	NH/PI	2	–		NH/PI	2	–		NH/PI	2	–
	OT	6	3.335		OT	6	3.17		OT	6	3.5
	WH	76	1.945		WH	50	2.52		WH	67	2.2
	AI/AN	6	3.17		AI/AN	4	1		AI/AN	6	1.5
	AS	68	2.57		AS	74	2.51		AS	92	2.91
	BL/AA	10	3.7		BL/AA	10	5		BL/AA	12	3.75
	HI/LA	8	4.75		HI/LA	18	2.56		HI/LA	19	2.26
Open-mouth	ME/NA	–	–	Beard	ME/NA	–	–	Glasses	ME/NA	2	–
	MR	15	2.73		MR	20	2.6		MR	26	1.85
	NH/PI	2	–		NH/PI	2	7.5		NH/PI	2	–
	OT	6	5.33		OT	6	3.33		OT	6	3.83
	WH	47	2.66		WH	68	2.86		WH	76	1.93
	AI/AN	6	3		AI/AN	4	2.5		AI/AN	5	3
	AS	68	2.46		AS	90	2.89		AS	91	2.4
	BL/AA	12	2.83		BL/AA	12	4.67		BL/AA	12	2.92
	HI/LA	14	2.21		HI/LA	20	2.9		HI/LA	18	2.22
Smile	ME/NA	–	–	KN95	ME/NA	2	–	Eyepatch	ME/NA	2	–
	MR	21	2.43		MR	22	3.05		MR	23	3.48
	NH/PI	2	–		NH/PI	2	–		NH/PI	2	–
	OT	6	3.5		OT	4	5.5		OT	4	3
	WH	52	2.52		WH	60	3.68		WH	68	2.57
	AI/AN	5	5.2		AI/AN	6	3.17		AI/AN	3	38.33
	AS	84	3.44		AS	90	2.36		AS	70	45.8
	BL/AA	10	3.2		BL/AA	12	2.83		BL/AA	8	38.88
	HI/LA	19	3.53		HI/LA	18	2.33		HI/LA	12	49.67
Groucho	ME/NA	2	–	Wig	ME/NA	2	–	Costume mask	ME/NA	1	–
	MR	23	3.35		MR	22	2.09		MR	19	44.47
	NH/PI	2	–		NH/PI	2	–		NH/PI	2	–
	OT	5	3		OT	6	4.33		OT	3	11.67
	WH	62	2.68		WH	70	2.21		WH	56	43.84