QualitEye: Public and Privacy-preserving Gaze Data Quality Verification

Mayar Elfares
mayar.elfares@vis.uni-stuttgart.de
Institute of Information Security, Collaborative Artificial Intelligence Group, University of Stuttgart
Stuttgart, Germany

Pascal Reisert
Institute of Information Security, University of Stuttgart
Stuttgart, Germany
pascal.reisert@sec.uni-stuttgart.de

Ralf Küsters
Institute of Information Security, University of Stuttgart
Stuttgart, Germany
ralf.kuesters@sec.uni-stuttgart.de

Andreas Bulling
Collaborative Artificial Intelligence Group, University of Stuttgart
Stuttgart, Germany
andreas.bulling@vis.uni-stuttgart.de

ABSTRACT

Gaze-based applications are increasingly advancing with the availability of large datasets but ensuring data quality presents a substantial challenge when collecting data at scale. It further requires different parties to collaborate, therefore, privacy concerns arise. We propose QualitEye—the first method for verifying image-based gaze data quality. QualitEye employs a new semantic representation of eye images that contains the information required for verification while excluding irrelevant information for better domain adaptation. QualitEye covers a public setting where parties can freely exchange data and a privacy-preserving setting where parties cannot reveal their raw data nor derive gaze features/labels of others with adapted private set intersection protocols. We evaluate QualitEye on the MPIIFaceGaze and GazeCapture datasets and achieve a high verification performance (with a small overhead in runtime for privacy-preserving versions). Hence, QualitEye paves the way for new gaze analysis methods at the intersection of machine learning, human-computer interaction, and cryptography.

KEYWORDS

Gaze, quality, privacy, private set intersection

1 INTRODUCTION

Eye tracking has seen widespread adoption for numerous applications, such as for gaze-based human-computer interaction [32, 59, 60, 69, 80], for understanding the human visual system [1, 9], measuring user experience [28] or for computational user modelling [2, 18, 27]. With eye tracking becoming pervasive [17] and increasingly integrated into personal devices [19, 47, 49], recent years have also seen a significant increase in the availability of large-scale gaze datasets [81, 85, 94, 96, 98]. Traditionally, these datasets have been collected in research contexts but are now increasingly collected and shared by private individuals and commercial enterprises [34, 35].

A challenge amplified by these advances that has largely been neglected in the gaze community so far is verifying the quality of the acquired, collected, or shared gaze data. Although a few prior works [7, 8, 36, 44, 53, 68] focused on gaze data quality (e.g., evaluating the eye-tracking systems' accuracy, signal-to-noise ratio, or robustness), the verification aspect of the acquired gaze data is largely neglected, especially for image-based gaze data (e.g. features or labels). In this work, we verify the quality of the eye images and their compliance with the respective labels (e.g. gaze direction and head pose) while ignoring irrelevant cross-user features (e.g. appearance). The quality verification ensures that similar eye features correspond to similar labels, by comparing the data samples with a relatively reliable source (e.g. a publicly available dataset or a trusted party). This is particularly important as gaze data quality can be subject to inaccurate labels or inconsistent features due to technical problems with the recording setup, choice of data preprocessing methods, or calibration and systematic errors.

Gaze data quality concerns can be found in numerous eye tracking applications, including (i) Data collection and cleaning where data collectors can verify the quality of their data against other reliable sources to detect and correct (or remove) corrupt or inaccurate data samples [12], (ii) Auto-labelling of new data from a small subset of reliable labelled data [78], (iii) Remote learning/analytics setups, e.g. federated learning/analytics [34, 35], where a gaze-based model is trained at different data sources.

The need for data quality verification is further aggravated in setups where in-the-clear (public) access to the raw gaze data is not allowed due to privacy concerns, and therefore, the data needs to be processed in a privacy-preserving manner, e.g. [34, 35, 58]. Therefore, we present QualitEye, the first work to study this question and propose a computational method for gaze data quality verification. QualitEyecovers two setups:

(1) **Public verification:** that can be used locally to (i) verify the within-dataset consistency or (ii) compare the local gaze dataset against the publicly available datasets.
(2) **Privacy-preserving verification:** Since the within-dataset verification can be subject to systematic errors and biases, and public datasets are not always available (e.g. due to privacy concerns)), this can be used to remotely verify the dataset quality against a private dataset owned by a different party without direct access to the data.

QualitEye, first, disentangles the gaze direction and head pose features that correspond to the data labels, ignoring the cross-party irrelevant features (e.g. appearance) for a high domain adaptation performance instead of the raw pixel-wise data comparison. Then,

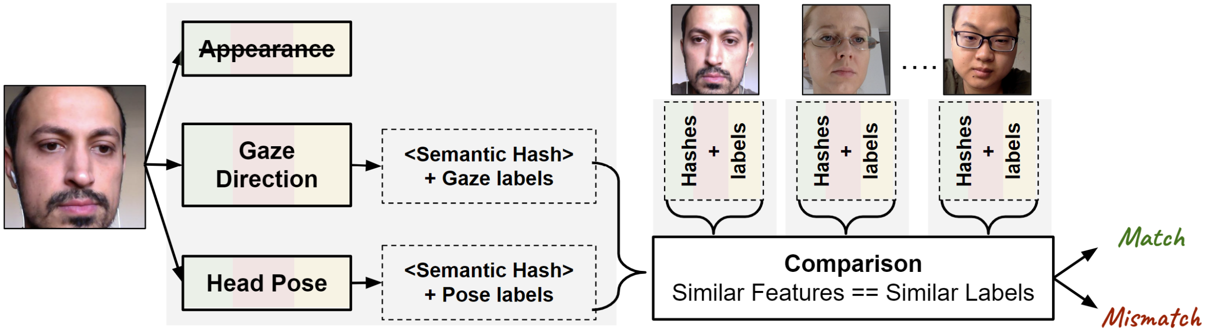

Figure 1: To verify the gaze data quality, data owners compute a hashed semantic representation of the eye image that includes the respective gaze direction and head pose while ignoring the cross-users' irrelevant features (e.g. appearance). Then, they compare the hash values and corresponding labels with a reference party (e.g. a reliable source) to find the (mis)matching data samples.

features are semantically hashed for an efficient bit-wise comparison to obtain data-independent (i.e. not domain-specific), deterministic (i.e. produces the same outputs for similar semantics), and generative (i.e. learns the gaze data distribution) representations. Finally, representations are compared against different data samples under public and privacy-preserving setups, as shown in Figure 1.

To evaluate QualitEye, we present appearance-based gaze estimation as our guiding example since it is the basic building block of gaze-based applications and has well-established publicly available datasets for evaluation. Nonetheless, QualitEye does not restrict how the dataset should be processed after the quality verification; hence, QualitEye is domain-, task-, and model-agnostic. We thereby validate our method through extensive experiments on the well-established full-face appearance-based gaze estimation datasets, MPIIFaceGaze and GazeCapture, and achieved a gaze quality metric (Matthew Correlation Coefficient [70]) of 0.92 and 0.94, respectively, with a small overhead in runtime for the privacy-preserving setups. The Matthews Correlation Coefficient (MCC) is a metric used to assess the quality of binary classifications (i.e. match or mismatch) and it is particularly useful in cases where there is a class imbalance (e.g., when one class is much more frequent than the other), making it more reliable than metrics like accuracy in these situations (cf. Section 4).

In summary, this work makes the following contributions:

- QualitEye is the first work to investigate the problem of image-based gaze data quality verification.
- Instead of the raw pixel-wise data comparison, we propose a new generic hashed representation learning model that disentangles the gaze-specific features and ignores the cross-users' irrelevant features (e.g. appearance).
- We propose methods for public and privacy-preserving gaze data quality verification. For the latter, we extend existing privacy-preserving protocols with semantic similarities and label matching to handle the different privacy requirements.

2 RELATED WORK

In the following, we summarise previous work that is most closely related to our method for gaze data quality verification. QualitEye focuses on the verification of (i) gaze data quality through (ii) unsupervised gaze representation learning to compare relevant features for both public and (iii) private gaze data verification.

Gaze data quality. The availability of large-scale and diverse datasets to study and address the variability in eye data across users, tasks, and settings remains a key limitation in the eye tracking research [34, 42]. Nonetheless, with the recent advances in gaze data acquisition, sharing, and processing, adequate quality standards required for high-performing gaze-based models and analytics are highly neglected [45]. The main reasons are (i) the time-consuming process of collecting and labelling large datasets, (ii) the unwillingness of data owners to share their private eye data, and (iii) the challenge for single parties (e.g. companies or research groups) to collect such diverse data at scale [84]. Recently, Deborah et al. [53] highlighted the importance of the gaze data quality reporting by including metadata, e.g. sampling rate, tracked eye(s), filter settings, date and time, total recording duration, and display resolution to insure accessible, re-usable and interoperable data and comparable, reproducible, and transparent implementations. Few recent works have evaluated eye-tracking systems' accuracy and signal-to-noise ratio [7, 8, 36, 68]. In contrast, we perform an automatic quality verification for appearance-based gaze data without manual reports, both in the clear (publicly) and without direct access/data sharing (privacy-preserving verification).

Unsupervised gaze representation learning. Previously, supervised representation learning techniques have been used to learn task-specific eye features, e.g. gaze estimation [94, 96] or eye contact detection [95]. However, these methods assume carefully labelled data and do not generalise well to other tasks, unlike our method, which does not restrict specific tasks or the availability of correct labels. Alternatively, self-supervised learning (SSL) methods enable deep learning models to learn the underlying image features by obtaining supervisory signals directly from the data without task-specific labels, which yield more subtle representations, especially for less common semantics [41], such as the heterogeneous person-specific differences in appearance-based methods [34]. SSL methods first started with joint embedding architectures (e.g. siamese networks) that try to learn the joint embeddings of two inputs [67, 86, 93], such methods collapse in practice and theory, i.e. the network produces identical embeddings ignoring the matching/non-matching inputs. Therefore, contrastive methods were introduced. They rely on creating pairs of positive and negative samples and learning the similar/dissimilar embeddings. However, enumerating all possible positive-negative pairs and all possible predictions in computer vision is an intractable problem, and current solutions introduce bias through hand-picked examples (e.g. [55]). Other non-contrastive methods [23, 25, 89] require the (theoretical) optimisation of the capacity of the latent variable, i.e. maximising the likelihood estimation on the intractable marginal log-likelihood. Nonetheless, in practice, a "fuzzy" latent can solve these limitations; therefore, in this work, to learn gaze representations, we propose a variational auto-encoder (VAE)-based model that is generative, non-contrastive, and uses a fuzzy latent variable. It uses a neural network as an amortised optimisation across the gaze data points.

Privacy-sensitive gaze data. Gaze data might contain personal (e.g. identifiers [22, 79, 91] and confidential attributes [46, 48, 83, 90]), or business-related information (e.g. the used devices or information about the participants [16]) that cannot be shared. Privacy threats and formal privacy solutions have been so far under-explored in the eye tracking and security communities. One reason for the (missing) development of these research branches is that cryptographic and privacy-preserving techniques (PPTs) [6] for a long time were considered impractical due to their significantly large computational costs. However, recent privacy-preserving technologies, such as federated learning (FL) [34], differential privacy (DP) [15, 65, 66, 84], and secure multi-party computation (MPC) [35], have been efficiently discussed as viable solutions for gaze-based tasks. Nonetheless, QualitEye is the first work to tackle the issues related to gaze data quality verification while also taking privacy concerns into account.

Privacy-preserving techniques. Secure data comparison techniques [3, 13, 30, 51, 64, 71–73, 87] aim to identify similarities between datasets while preserving privacy. Traditional methods rely on general-purpose cryptographic techniques, such as homomorphic encryption (HE) [38], secure multiparty computation (MPC) [40, 92], and oblivious transfer [77], to ensure that no sensitive data is exposed during the comparison process. Recent approaches leverage private set intersection (PSI) protocols [11, 20, 21, 29], which allow two or more parties to compare encrypted data sets and identify common elements without revealing the actual contents of their datasets. Unlike general-purpose privacy-preserving techniques, PSI outputs only the intersection (or its cardinality), performs operations only on one party's input, and exchanges cryptographic hashes or encoded elements, significantly reducing computational and communication overhead. Secure comparison techniques are applied across diverse domains, including biometric verification [13, 26, 37, 87], data linkage [3, 33, 72, 73], fraud or abuse detection [30, 51, 71, 99], and genomics [5, 61, 64]. These applications usually require significant error tolerance to address discrepancies such as name variations, formatting differences, typos, or outdated records, and to account for factors like lighting, orientation, physiological changes, or genomic mutations. In contrast, gaze-based applications are particularly sensitive to subtle variations in gaze direction and head pose, necessitating a novel privacy-preserving comparison tailored to these unique challenges, as proposed in this work. It also intentionally excludes appearance-based factors, concentrating solely on the application-relevant data features, thereby reducing the risk of data leakage and minimizing data transfer.

3 QUALITEYE

Gaze data quality verification is the process of computationally assessing the consistency and ensuring compliance with the requirements and standards of gaze data.

Problem statement. We consider a setup in which (i) a reference party R owns a validation dataset and (ii) a data owner O owns a gaze dataset. Each dataset includes the raw eye images with features x_i^j along with their corresponding labels y_i^j, where i denotes the data sample index and $j \in \{R, O\}$ denotes the party's id. In this paper, we present a most common leading gaze example - appearance-based gaze estimation with eye images and their corresponding gaze direction and head pose labels[1]. We assume a horizontal data distribution where each party's dataset includes different data samples, e.g. data of different participants and different numbers of samples. The data owner wishes to verify the quality of the gaze dataset with respect to the reference dataset to overcome the data collection and labelling problems mentioned above.

We define the gaze dataset quality as the number of data samples (the image features and the corresponding labels) that comply with the reference dataset, i.e. the cardinality of the intersection set. More formally, the set of mismatching data samples would be $\{(x_i^O, y_i^O) | \exists (x_i^R, y_i^R) : x_i^O = x_i^R \land y_i^O \neq y_i^R\}$.

3.1 Semantic gaze representations

Since the direct pixel-wise comparisons of the gaze images only capture exact matches. To obtain more meaningful results, we increase the efficiency of comparisons by first encoding the images in a semantic representation.

Our goal is to learn the semantic representations of the data samples that reflect the similarity of the gaze-based information (e.g. gaze direction and head pose) while ignoring other cross-party irrelevant features (e.g. appearance). In addition, this representation should be deterministic (i.e. produces the same outputs for similar semantics at different parties), generative (i.e. learns the gaze data distribution), and domain-agnostic (i.e. not dataset-specific) to be able to generalize well to different (unseen) datasets at different parties.

Variational auto-encoder (VAE). As shown in Figure 2, our end-to-end VAE-based implementation is composed of four main components:

- **Top-down encoder** E_ϕ that maps the input x to a latent z corresponding to a variational distribution and outputs the parameters

[1] Our approach is not limited to gaze estimation and can be extended to other gaze-based applications. Nonetheless, we choose gaze estimation as our guiding example since it is the basic building block of gaze-based applications and has well-established publicly available datasets for evaluation.

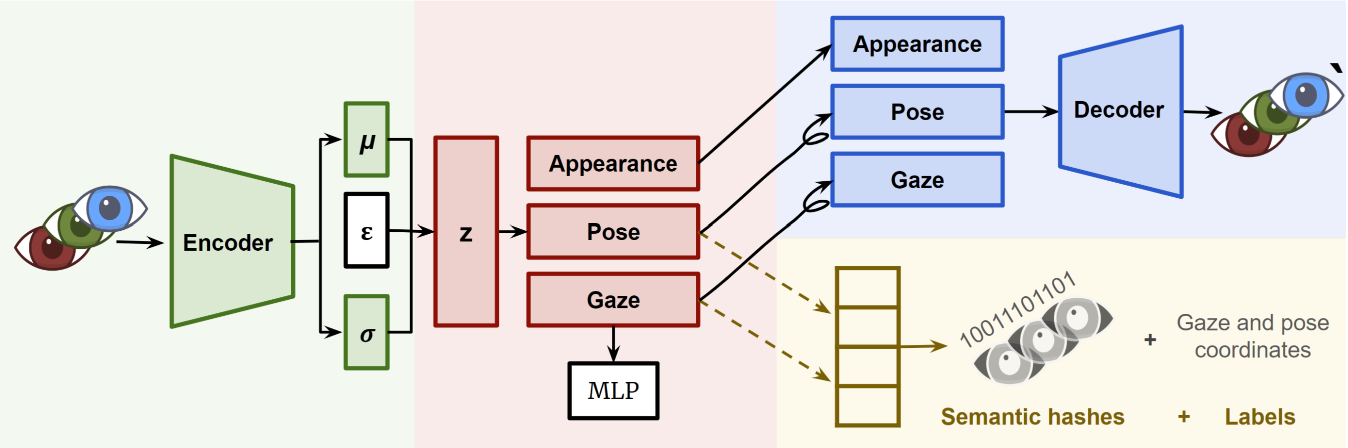

Figure 2: To obtain the hashed semantic representations used for label-based data comparison, *1- green:* eye images are encoded into a latent vector z. *2- red:* Then, the gaze and head pose information are disentangled from the appearance via some transformations and a multi-layer perception (MLP). *3- blue:* During training, the appearance, gaze, and pose are passed as a latent vector to a decoder that reconstructs the transformed eye images. *4- yellow:* Finally, the gaze and head pose latent codes are hashed and are passed as inputs to the PSI protocol along with the corresponding labels.

of the distribution (e.g. mean μ and variance σ). In our case, we use a multivariate Gaussian distribution $N(x|\mu, \sigma)$ due to the nature of RGB eye images and to estimate the average of the gaze data distribution (with likelihood $p_\theta(x|z)$) according to the *central limit theorem* for a better cross-domain adaptation. Then the prior $p_\theta(x)$ is a mixture of Gaussian distributions, and the posterior distribution $p_\theta(z|x)$ can be approximated to $q_\phi(z|x)$ (i.e. the amortized optimization).

- **Bottom-up decoder** D_θ maps the latent to the input space. Hence, our problem is to compute the conditional likelihood distribution $p_\theta(x|z)$ by the probabilistic decoder and the approximated posterior distribution $q_\phi(z|x)$ by the probabilistic encoder.
- **Feature disentanglement:** For a better domain adaptation, our model explicitly learns to disentangle the gaze direction and head pose representations as equivalent input rotations. This is achieved by training the model in a person-independent manner [67] and splitting z into three sub-vectors, (i) gaze direction, (ii) head pose, and (iii) appearance (more specifically, all other information found in the image), similar to [75], and rotating the latent sub-vectors using rotation matrices to the frontal angle and then to certain yaw and pitch angles. That is, for the same person, the model learns to transform the gaze direction and head pose of one image into the other by multiplying the sub-codes by a rotation matrix and optimising a pixel-wise $L1$ loss function over the entire encoder-decoder and an MLP regression for the gaze sub-code.

  The aim of disentangling gaze direction and head pose from the rest of the image is to (i) only compare the information that is relevant to the ground-truth labels and (ii) to minimize the amount of information exchanged in the cross-party setup (i.e. data minimization) for better privacy, runtime, and communication.
- **Hashing:** Once the latent vector is disentangled, the gaze direction and head pose sub-codes are hashed with a locality-sensitive hash function. Locality-sensitive hashing (LSH) is a fuzzy hashing technique that maps similar inputs to the same hash value with a certain probability. Such hashes (i) reduce the dimensionality of the semantic representations, which likelyimproves efficiency (e.g. computational runtime and communication), (ii) allow efficient comparisons of the bit-wise representations (i.e. comparing the hashed values in the hamming space instead of the latent space), (iii) are data-independent (i.e. not domain-specific to generate the same hash for different inputs at different parties), and most importantly, (iv) produce identical hashes for images with similar features (i.e. to tolerate small systematic errors in data acquisition commonly found in eye tracking data).

VAEs are specifically interesting for gaze-based data because of (i) the inherent mutual information in appearance-based eye data (higher mutual information yields better disentanglement [24]), (ii) the independence between the latent variables (e.g. gaze direction and head pose) encourages interpretability yielding better semantics [24, 43], and (iii) they are more generalisable[2] [4].

3.2 Quality verification in the public setting (QualitEye-V_0)

As the raw pixel-wise eye image comparison cannot be used for quality verification given the cross-user variations in appearance, the hashed disentangled gaze direction and head pose representation can be used instead. Hence, a gaze data owner O can verify the within-dataset consistency by checking that similar representations of gaze direction or head pose have similar respective labels. Additionally, in case of systematic or calibration errors (e.g. resulting from changes in user position during data collection), O can compare the collected dataset against the publicly available datasets (as the reference) to either check for possible errors (i.e. correcting the non-compliant data samples) or for auto-labelling the dataset.

[2] The generalisation capability makes VAEs more robust against adversarial attacks [4]. This increases privacy but remains out of the scope of this paper.

However, in other scenarios, the datasets might not be available at one party, therefore, O might need to verify the dataset quality against a different dataset owned by another reference party R. A (public) solution would be for one party (e.g. R) to send their data as hashed disentangled representations along with the labels to the other party (e.g. O) for comparison.

3.3 Privacy-Preserving Quality Verification

The (public) solution mentioned above does not guarantee privacy as O can perform a dictionary attack (i.e. tries all possible hashes of the input space) and recover the plain representations, especially when the dictionary has a computationally reasonable size, e.g. in the case of gaze estimation. Note that, even if the plain representations do not include the appearance and the raw images cannot be reconstructed [35] (via our data minimization step of disentanglement), O can still deduce information beyond the (mis)matching samples such as the number of samples in R's dataset, the semantic meaning of all other samples, the plaintext labels, the kind of error (if any)... etc. Therefore, a better solution is to use a cryptographic solution with formal provable guarantees – private set intersection (PSI), where both parties interactively compute the intersection (i.e. one party cannot compute the intersection without the other party's help, e.g. mitigating dictionary attacks).

We assume a semi-honest (a.k.a. honest-but-curious) security model, i.e. parties will not deviate from the defined protocol; however, they might try to learn possible information from the legitimately received messages. Note that, in this work, an adversary refers to the main parties R and O.

To get high-quality data, one promising solution is to collect data across many data owners [34, 35] and compare the different data sources. However, in scenarios where gaze data privacy is a concern, in the lack of an alternative for a long time, it was common practice to protect the shared data by contractual agreements. However, this practice has several disadvantages, e.g. parties have to trust each other to honour the contract and users might not grant transfer of their data between parties. Instead, cryptographic solutions can be used, but they usually come with a computational overhead or a drop in utility. We, therefore, present several versions of QualitEye with different tradeoffs between efficiency (i.e. runtime and communication), privacy, and utility.

3.3.1 *Preliminaries.*

We, first, introduce essential concepts and terminologies. These preliminaries provide the necessary context and background for understanding the subsequent privacy-preserving gaze data verification methods.

Private set intersection (PSI). PSI is a secure multiparty computation (MPC) protocol that allows two (or more) parties to compare elements in their sets by computing the intersection without revealing any information beyond this intersection. Recently, PSI constructs [11, 20, 21, 29] included different adversarial models, efficiency tradeoffs, and security guarantees. In this paper, we focus on the semi-honest (a.k.a. honest-but-curious) adversarial model [39]. This assumes that different data owners have a mutual interest in working together and improving the quality of their own datasets. We therefore assume that they stick to the rules and follow a previously agreed protocol. However, they nevertheless are happy to gain any information leaked by the protocol. Such data owners are called honest-but-curious. QualitEye uses different cryptographic primitives to ensure that a curious data owner gains only minimal knowledge about other parties' data, e.g. oblivious transfer.

PSI constructions can include different cryptographic primitives. QualitEye relies on the following primitives:

- **Key agreement protocols:** In cryptography, key agreement protocols allow two or more parties to agree on a cryptographic key. Among these protocols, the Diffie–Hellman protocol [31] is one of the earliest practical protocols. More specifically, we base our privacy-preserving constructions of QualitEye$_{v1,v12,v3}$ on a Diffie-Hellman-based PSI protocol where, as shown in Figure 3, the two parties O and R hash all their data samples x_i^j [3], and raise the hashed values to their private keys. Then, these values are exchanged and compared by O. A match means that both x_i^O and x_i^R are similar. A mismatch reveals no information about the inputs[4]. This way, dictionary attacks are not possible since R is committed to certain values in the first message, and the intersection requires the interaction between both parties to get both private keys (i.e. O will not remain available for R to try all different dictionary entries). Note that an eavesdropper can intercept the communication but does not have access to the private keys; therefore, the eavesdropper cannot infer the intersection [50].
- **Oblivious transfer (extension):** Oblivious transfer (OT) [77] and Oblivious transfer extension (OTe) [10, 52] are cryptographic protocols where one party sends one of many pieces of information to a receiver, but remains oblivious as to what exact piece has been sent.
- **Oblivious pseudorandom functions (OPRF):** OPRF is a cryptographic protocol where two parties jointly compute a pseudorandom function (PRF), i.e. a function which emulates a random oracle. Similar to OT, the sender does not learn any information about the other party's input, i.e. the sender is oblivious to what exact values has been sent. More specifically, we base our PSI construction of QualitEye$_{v4}$ on an OPRF-based construction [62]. As shown in Figure 4, both parties hash their inputs to two hash values. R only selects one of the two hash values (using cuckoo hashing [74]). O sends all his inputs (as two PRF outputs per input corresponding to the two hash values) to R. R compares these outputs to his to compute the intersection. This way, R only learns the matching elements while all other elements in O's dataset look random to him, and O learns nothing about R's inputs.

[3] Parties get a cryptographic hash of the semantic hash values using their private keys K_O and K_R. Samples are hashed to a primitive root modulo p where p is a large prime number that parties agree on, i.e. $H(x_i^j)^{K_j} \mod p$. In modular arithmetic, a number g is a primitive root modulo n if every number a coprime to n is congruent to a power of g modulo n. That is, g is a primitive root modulo n if for every integer a coprime to n, there exists some integer k for which $g^k \equiv a(modn)$. In other words, every invertible number is of the form g^k for some integer k.

[4] As proved by Diffie-Hellman [31], raising $H(x_i)$ to an exponent (i.e. the secret key) makes it indistinguishable from random (i.e. the protocol hides the inputs when there is no match), even if the exponent is used elsewhere (i.e. it is safe to reuse $H(x_i)^{K_j}$ as in QualitEye$_{v3}$).

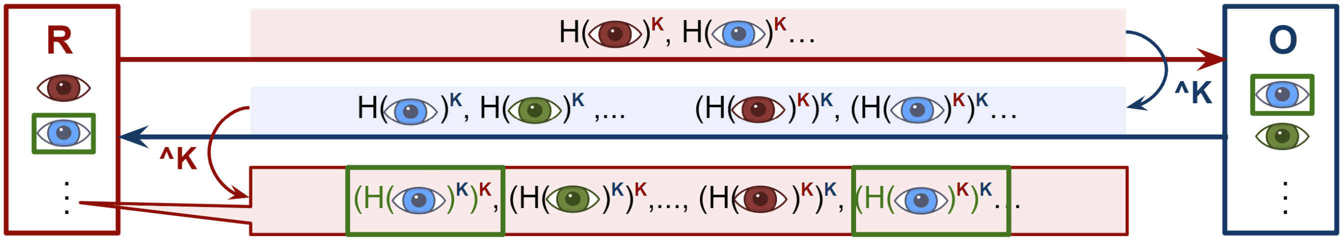

Figure 3: In QualitEye-V$_1$, both parties *O* and *R* exchange their elements as hashed values raised to their private keys. Then, they raise the other party's received values again to their private keys. *R* then start the comparison to find the matching inputs. Note that, we further send the eye labels (e.g. gaze direction and head pose) as an additional encrypted payload to each element. In QualitEye-V$_2$, *O* shuffles the second message to only reveal the cardinality of the intersection. In QualitEye-V$_3$, *R* includes the first message, e.g. as a package, before the start of the protocol. In all versions, only the private keys are secret, and all other values are sent in the clear. Security is still guaranteed due to the hardness of the 'the discrete logarithm problem', i.e. it is hard to infer the private keys [31].

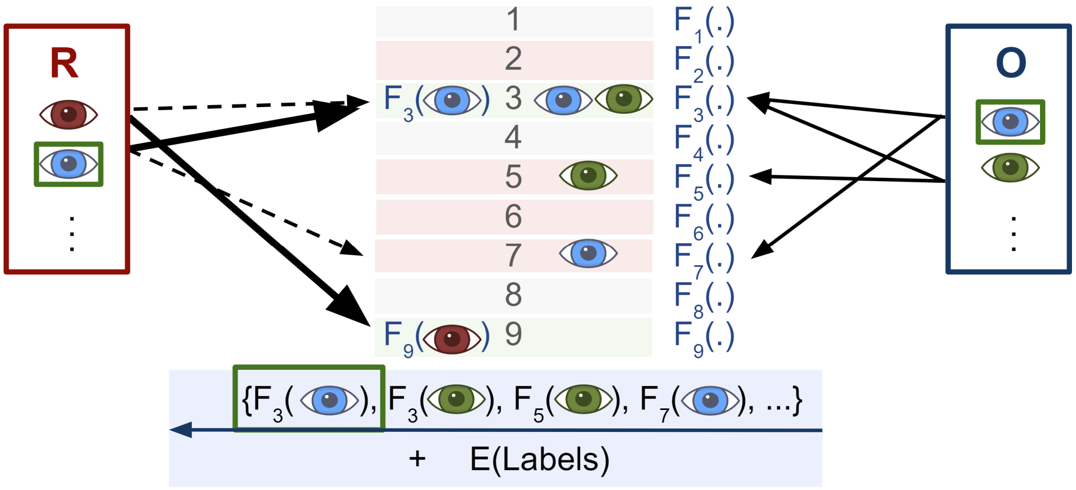

Figure 4: In QualitEye-V$_4$, each party computes two values per element (i.e. the PRF output of the hashed semantic representations). *R* only selects one value per element (via cuckoo hashing). *O* send his values (along with the encrypted blinded labels) to *R*. *R* sends back all received information along with her encrypted blinded labels to *O*, who makes the comparison and finds the (mis)matching elements.

3.3.2 Adaptation of existing PSI protocols. PSI protocols typically operate on input messages to find matches between two (or more) sets. In our case, three additional challenges arise:

(1) The comparison of the raw eye images only reflects the exact similarity. We solve this issue by comparing the hashed semantic representations of the inputs in each dataset (generated in subsection 3.1).
(2) In addition to finding matching elements in the parties' sets, we further need to check the compliance of the labels. In the following, we solve this problem by extending the protocols with additional payloads.
(3) The accuracy of some labels (e.g. gaze direction) in the reference gaze-based data can include some error, currently >3 degrees for gaze direction [88]. Therefore, we allow some error tolerance for the semantic representations and their respective payloads.

Therefore, QualitEye uses the aforementioned preliminaries in privacy-preserving gaze data verification to offer different versions according to different efficiency tradeoffs. The different versions cover: (i) Dataset sizes: hundreds vs thousands of samples, (ii) difference in datasets size: symmetric (i.e. when both parties have the same amount of data) vs asymmetric (i.e. when one party has a relatively larger dataset) dataset distributions, (iii) resulting information: the intersection set vs its cardinality, (iv) receiver of this information: one vs both parties.

- **QualitEye-V$_1$** can be used when both parties have relatively small datasets (i.e. tens to hundreds of samples) and would like to know the exact (mis)matching samples. More specifically, QualitEye-V$_1$ is based on the Diffie-Hellman key exchange protocol, as shown in Figure 3.
- **QualitEye-V$_2$** can be used when both parties have relatively small datasets (i.e. tens to hundreds of samples) but are only

interested in knowing the cardinality of the intersection (i.e. the size of the (mis)matching samples and not the exact samples). Similarly, QualitEye-V$_1$ can be extended to only reveal the cardinality of the intersection (i.e. the intersection size and not the exact samples). In QualitEye-V$_1$, R sends $M2b$ (Figure 3) in the same order sent by O so that O can find the corresponding elements in her set. In QualitEye-V$_2$, R shuffles those elements before sending them to O. Hence, O can only count the number of matching elements but cannot map them to the raw elements (i.e. elements appear uniform).

- **QualitEye-V$_3$** can be used when one party has a relatively small dataset while the reference party owns a larger dataset (i.e. thousands to billions of samples). In this case, $M1$ in QualitEye-V$_1$ and QualitEye-V$_2$ (Figure 3) can be sent in advance (e.g. offline as a built-in package in an eye-tracking software or as a publicly-published data) as it does not depend on the other party's input. Note that this does not break privacy and can be shared with multiple parties (or protocol instances) [5].
- **QualitEye-V$_4$** can be used when one or both parties own large datasets where the previous versions are highly inefficient (c.f. 4). As shown in Figure 4, QualitEye-V$_4$ is based on a OPRF-based PSI protocol [62]. QualitEye-V$_4$ can be an alternative to QualitEye-V$_3$ when changes to the data of the reference party are frequently made.

For all versions, the final information can be revealed to (i) one party, depending on which party starts the protocol or (ii) both parties with an additional communication step containing the resulting information (i.e. the intersection set or its cardinality). We further propose the following adaptations: In addition to the hashed semantic representations of the gaze direction and head pose, both parties input the corresponding labels encrypted as Elgamal ciphertext (an asymmetric key encryption based on the Diffie-Hellman key exchange) under their private keys. For instance, in Figure 4, O sends the PRF outputs along with the corresponding encrypted (under O's key) labels. If R finds a matching PRF representation in his dataset, he encrypts his labels and sends both encrypted labels to O. If there is no match, R re-encrypts O's label. Then, O decrypts both labels to find the mismatching inputs. Note that O cannot distinguish between the cases where a matching representation does not exist in R's dataset and a matching representation exists with a matching label; he only learns the mis-matched labels (i.e. the non-compliant samples). On the other hand, O only learns the cardinality of the matching set. Additionally, parties do not learn the labels in the clear as labels are further blinded following [12].

Furthermore, to accommodate the (unavoidable) systematic errors in the labels (e.g. 3 degrees for gaze direction) in the hashing step, similar to the gaze representations, we adjust the probability that different latent codes are mapped to the same hash values and drop the least significant bits in the labels accordingly. The parties agree on the exact values that can be adjusted according to the data collection setups (controlled vs in-the-wild, remote vs near-eye cameras... etc) and the corresponding established error values (e.g. eye-tracker drift or calibration errors).

4 EXPERIMENTS

4.1 Datasets

To evaluate QualitEye, we use different appearance-based gaze estimation datasets covering different conditions: appearances (genders, ethnicities, glasses, and make-up), illumination (indoor and outdoor), and gaze direction and head pose distributions. Mainly, experiments were conducted on the full-face MPIIFaceGaze [97] and GazeCapture [63] datasets. The MPIIFaceGaze [97] dataset contains ~200 thousand full-face images collected in the wild from 15 participants. The GazeCapture [63] dataset contains ~2.5 million frames of ~1.5 thousand participants.

4.2 Results and implementation details

We train our VAE with an enhanced ResNet [88] backbone on the training set (80% of GazeCapture) in a person-specific fashion since the inter-subject anatomical differences are known to affect the performance of gaze-based tasks [63, 97]. We then test our model on the remaining 20% of GazeCapture and the full MPIIFaceGaze dataset.

4.2.1 *Training*.

Loss function. We use the well-established loss function [75] adapted to our disentanglement criteria:

$$L_{\text{full}} = \lambda_{recon} L_{\text{recon}} + \lambda_{EC} L_{\text{EC}} + \lambda_{gaze} L_{\text{gaze}} + \lambda_{KL} L_{\text{KL}}$$

where L_{recon} is the reconstruction loss that guides the encoding-decoding process pixel-wise, L_{EC} is the embedding consistency loss that ensures the embedding of the same appearance into the same features even with different (disentangled) gaze direction and head pose, L_{gaze} is the gaze direction loss between the estimated gaze of the MLP and the true gaze direction, and L_{KL} is the VAE Kullback-Leibler divergence loss that regularizes the model by approximating the prior distribution via the encoded distribution and penalising deviations from the model. For the coefficients, we use $\lambda_{recon} = 2$, $\lambda_{EC} = 1$, $\lambda_{gaze} = 0.1$, and $\lambda_{KL} = 1$ with a batch size of 128 and a learning rate of $5 \cdot 10^{-7}$. This yielded a reconstruction loss of 0.3859, a gaze angular error of 5.0543°, and a KL-divergence loss of 12.9531. The EC_{gaze} is 5.8974 and the EC_{pose} is 10.4678.

Feature disentanglement. As shown in Figure 2, to disentangle the appearance, gaze direction, and head pose features, the encoder processes the input image into three distinct latent codes of size 64, 2, and 16, respectively. As shown in Figure **??**, we intentionally neglect the appearance code for (i) data minimisation, i.e. sharing less information to enhance privacy and (ii) for transferability across different subjects. The head pose code captures rotation and orientation (i.e. pitch and yaw angles) and excludes gaze direction to ensure that the gaze features remain consistent even if head orientation changes. This is achieved by a transformation (i.e. a rotation matrix), proposed by Park et al. [75], that maps each gaze representation relative to a canonical (or frontal) head position, leading to successfully disentangling the gaze direction and head pose as shown in Figure **??**.

4.2.2 *Inference and gaze data quality verification*.

[5] Kales et al. [57] further proposed an efficient encoding mechanism that could be used with $M1$ to enhance efficiency.

Quality metric. Once the model is trained, we use it for the gaze data quality verification in the inference mode. We report the quality metric as a Matthews correlation coefficient (MCC) [70] to account for true positives (TP, the correct compliant matching samples), true negatives (TN, the correct non-compliant mismatching samples), false positives, and false negatives (FP and FN, the incorrect predictions), as:

$$MCC = \frac{(TP * TN - FP * FN)}{\sqrt{(TP + FP) * (TP + FN) * (TN + FP) * (TN + FN))}}$$

MCC is particularly interesting as it can be used for classes of different sizes [14], i.e. both symmetric and asymmetric scenarios. A coefficient of +1 indicates a perfect match, 0 represents a random prediction, and −1 is a total disagreement between the predicted match and the true match.

Experimental setup. We run experiments on the within- and cross-participant for the same domain, e.g. MPIIFaceGaze in Figure 7. We further report the average MCC with varying participant-based data splits to handle the unbalanced data distribution across data sources in Table 1, and cross-datasets in Table 2 quality verification. Note that, we use the available dataset's labels as our ground-truth which is subject to error since creating labels is a complex task in the first place given the eye-head interplay, eye registration error, occlusions, appearance biases... etc. Hence, we adapt our hashing step to compensate for such (unavoidable) errors in the data where the dimensionality of the target projected space is reduced to 80 bits with a collision probability of 0.05[6].

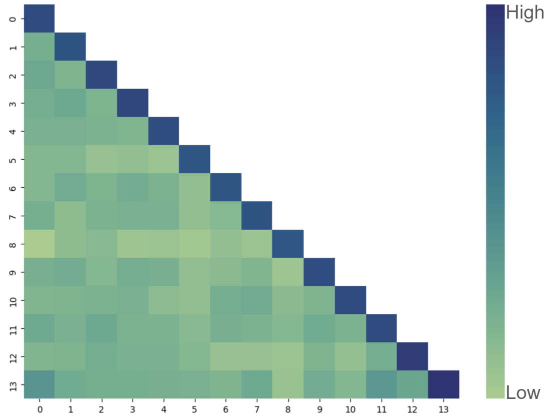

Figure 7: Cross-participant performance as normalised MCC on the MPIIFaceGaze dataset

Datasets' discrepencies. Although the GazeCapture and MPIIFaceGaze datasets are widely used in gaze research, they differ significantly in terms of data collection methodology, device settings, environmental conditions, and participant demographics. As shown in Table 1 and Figure 7, our method was able to successfully disentangle gaze direction and head poses within the same dataset, regardless of the appearance differences such as the different participants (1400 vs 15), demographics (diverse range of age vs university students), and background and lightning conditions (outdoor vs indoor) in GazeCapture and MPIIFaceGaze, respectively. The performance slightly degrades for head pose since the GazeCapture dataset was self-collected by participants using mobile devices (iOS phones and tablets), leading to a high variation in head pose due to uncontrolled conditions. It was also annotated with gaze points that are less precise due to mobile device limitations. Meanwhile, the MPIIFaceGaze dataset was recorded using laptops with a front-facing camera in indoor setups. This led to limited head pose variation and precise annotations due to controlled lab settings. Hence, the GazeCapture dataset is larger, more variable, and more general, and the same domain was used to train the VAE, hence the better performance. Nonetheless, QualitEye mainly relies on a reference dataset for comparison. Therefore, as shown in Table 2, QualitEye also succeeds at comparing different domains (i.e. datasets) but mainly fails in cases where a sample in one dataset does not have a matching sample in the other dataset when comparing ground-truth labels (i.e. outliers), as shown in Figure 8. Therefore, in practice, a careful consideration of the reference dataset that captures all the desired requirements is recommended.

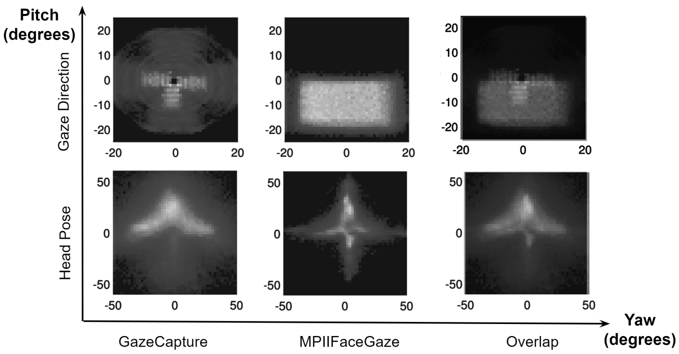

Figure 8: The pitch and yaw distribution of gaze direction and head pose on the GazeCapture (reference R) dataset, the MPIIFaceGaze (owner O) dataset, and the corresponding distribution overlap. QualitEye captures the overlapping samples and misses the samples in O that do not have a matching sample in R.

Privacy-preserving gaze data quality verification. The private cross-party gaze data verification setups maintain the same performance as the local public one in terms of data quality metrics. We use a computational security parameter $k = 128$ and a statistical security parameter of $\sigma = 40$ following [62].[7] However, privacy comes with a computational overhead in runtime and communication as shown in ??.

5 DISCUSSION

Our results show that the gaze data quality verification problem can be solved efficiently under different public and privacy-preserving

[6] Values are calculated according to the current SOTA remote gaze estimation models [88]. For instance, the normal gaze range is [-45,45] with a few outliers and an error of 3 and 5 degrees for gaze direction and head pose, respectively.

[7] $k = 128$, $\sigma = 40$ is a standard security parameter choice. Other values are possible. Generally, lower values might reduce the runtime of the cryptographic parts but can be more vulnerable to attacks. The security parameter choice does not affect accuracy.

Table 1: Within-dataset performance for gaze verification averaged over different participant-based data splits.

Dataset	TP	TN	FP	FN	**MCC**
Gaze direction:					
MPIIFaceGaze	0.9628	0.9967	0.0033	0.0372	0.9220
GazeCapture	0.9808	0.9966	0.0034	0.0192	0.9402
Head pose:					
MPIIFaceGaze	0.8381	0.9062	0.0938	0.1619	0.746
GazeCapture	0.9564	0.9289	0.0711	0.0436	0.8856

Table 2: Cross-domain performance for gaze quality verification on a TITAN X 12G GPU

Dataset	TP	TN	FP	FN	**MCC**	Runtime
MPIIFaceGaze vs GazeCapture	0.9740	0.9966	0.0360	0.0260	0.9331	152s

setups while achieving a good tradeoff between performance, runtime, and communication. As a pioneering work, QualitEye paves the way towards better gaze data quality and incentivises remote gaze-based data collection.

Results in **??** confirm our theoretical hypotheses in section 3: QualitEye$_{v1}$ can be used when both parties have relatively small datasets (i.e. tens to hundreds of samples) and would like to know the exact (mis)matching samples. QualitEye$_{v2}$ can be used when both parties have relatively small datasets (i.e. tens to hundreds of samples) and are only interested in knowing the cardinality of the intersection with an additional shuffling step. QualitEye$_{v3}$ can be used when one party has a relatively small dataset while the reference party owns a larger dataset (i.e. thousands to billions of samples) by offloading the larger dataset computation to an offline phase, i.e. asymmetric scenarios. QualitEye$_{v4}$ can be used when one or both parties own large datasets as the runtime with OPRF-based approach decreases significantly with respect to the DH-based approaches when the dataset size increases.

Relevance and Implications. Gaze data quality verification has the potential for numerous eye-tracking scenarios. We do not add restrictions on the datasets (e.g. size or content) nor on how the dataset should be processed after the quality verification (e.g. downstream tasks). Hence, QualitEye is domain-, task-, and model-agnostic. We further offered different privacy-preserving versions to accommodate these various scenarios (cf. Section 3). These application scenarios include:

(1) **Data collection and cleaning** where data collectors can verify the quality of their data against other reliable sources to detect and correct (or remove) corrupt or inaccurate data samples [12].
(2) **Auto-labelling** of new data from (i) a small subset of locally-labelled data or (ii) another reliable data source since data labelling is usually a tedious and time-consuming process [78].
(3) **Remote learning/ analytics setups**, e.g. federated learning/ analytics [34, 35], where different parties aim to locally train gaze-based models without sharing the raw data and only send the trained models (weights or gradients) to a central server. The quality of the remote private data, therefore, needs to be verified.
(4) **Try-before-you-buy dataset/model services** [82] where parties can verify the gaze data quality or the gaze-based model predictions (as labels) with a few user samples without getting access to the data before making a purchase.
(5) **Predictive benchmarks/leaderboards** (e.g. Kaggle [56]) to verify model predictions (as labels) against privately-held test sets without the need of sending a version of the gaze-based model to the leaderboard server (the current practice).

Limitations and future work. In this paper, we mainly focus on gaze angles and head poses as they are the most common labels in gaze-based datasets. Although no prior work has investigated this problem, we hypothesize that the gaze data quality can further be enhanced through diversity and inclusion. For example, further information (e.g. illumination conditions [54]) can be disentangled from the appearance to extend QualitEye. In addition, we focus exclusively on gaze-related tasks involving image data, recognizing that tasks in other data modalities (e.g., scanpaths or videos) also present rich and valuable opportunities for deep learning research. However, these modalities often require specialized deep learning models and privacy assumptions (e.g., temporal correlations) tailored to their unique characteristics and challenges. Furthermore, we presented a two-party computation (2PC) protocol; however, in general, 2PC protocols can be extended to multi-party computation (MPC) protocols and the gaze data quality verification can be improved with more data sources to reduce errors further. Moreover, QualitEye assumes a semi-honest setup – a common assumption in the security literature for newly developed problems. However, stronger security guarantees [76] can be achieved under the malicious setup (with different efficiency tradeoffs).

6 CONCLUSION

We presented QualitEye– the first work to investigate the problem of gaze data quality verification. We introduced a new generic hashed representation learning model that disentangles the gaze direction and head pose features for a high-domain adaptation performance, ignoring the cross-user irrelevant features (e.g. appearance) to allow for a label-specific comparison. Furthermore, we extended existing privacy-preserving interactive protocols with semantic similarities and labels matching to handle the different privacy and trust requirements. Our results show that QualitEye is efficient under different public and privacy-preserving setups in terms of performance, runtime, and communication.

REFERENCES

[1] Yomna Abdelrahman, Anam Ahmad Khan, Joshua Newn, Eduardo Velloso, Sherine Ashraf Safwat, James Bailey, Andreas Bulling, Frank Vetere, and Albrecht Schmidt. 2019. Classifying Attention Types with Thermal Imaging and Eye Tracking. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)* 3, 3 (2019), 1–27. https://doi.org/10.1145/3351227

[2] Ahmed Abdou, Ekta Sood, Philipp Müller, and Andreas Bulling. 2022. Gaze-enhanced Crossmodal Embeddings for Emotion Recognition. In *ACM ETRA (ETRA, Vol. 6)*. ACM, 1–18. https://doi.org/10.1145/3530879

[3] Allon Adir, Ehud Aharoni, Nir Drucker, Eyal Kushnir, Ramy Masalha, Michael Mirkin, and Omri Soceanu. 2022. Privacy-preserving record linkage using local sensitive hash and private set intersection. In *International Conference on Applied Cryptography and Network Security*. Springer, 398–424.

[4] Alexander A. Alemi, Ian Fischer, Joshua V. Dillon, and Kevin Murphy. 2016. Deep Variational Information Bottleneck. *CoRR* abs/1612.00410 (2016). arXiv:1612.00410 http://arxiv.org/abs/1612.00410

[5] Nour Almadhoun, Erman Ayday, and Özgür Ulusoy. 2020. Differential privacy under dependent tuples—the case of genomic privacy. *Bioinformatics* 36, 6 (2020), 1696–1703.

[6] David W. Archer, Borja de Balle Pigem, Dan Bogdanov, Mark Craddock, Adria Gascon, Ronald Jansen, Matjaž Jug, Kim Laine, Robert McLellan, Olga Ohrimenko, Mariana Raykova, Andrew Trask, and Simon Wardley. 2023. UN Handbook on Privacy-Preserving Computation Techniques. arXiv:2301.06167 [cs.CY]

[7] Samantha Aziz and Oleg Komogortsev. 2022. An assessment of the eye tracking signal quality captured in the HoloLens 2. In *2022 Symposium on eye tracking research and applications*. 1–6.

[8] Samantha Aziz, Dillon J Lohr, Lee Friedman, and Oleg Komogortsev. 2024. Evaluation of Eye Tracking Signal Quality for Virtual Reality Applications: A Case Study in the Meta Quest Pro. *arXiv preprint arXiv:2403.07210* (2024).

[9] Dale J Barr. 2008. Analyzing 'visual world'eyetracking data using multilevel logistic regression. *Journal of memory and language* 59, 4 (2008), 457–474.

[10] Donald Beaver. 1996. Correlated pseudorandomness and the complexity of private computations. In *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*. 479–488.

[11] Bellare, Namprempre, Pointcheval, and Semanko. 2003. The one-more-RSA-inversion problems and the security of Chaum's blind signature scheme. *Journal of Cryptology* 16 (2003), 185–215.

[12] E. Blass and F. Kerschbaum. 2023. Private Collaborative Data Cleaning via Non-Equi PSI. In *IEEE S&P*. 1419–1434.

[13] Remco Bloemen, Bryan Gillespie, Daniel Kales, Philipp Sippl, and Roman Walch. 2024. Large-scale MPC: Scaling private iris code uniqueness checks to millions of users. *arXiv preprint arXiv:2405.04463* (2024).

[14] Sabri Boughorbel, Fethi Jarray, and Mohammed El-Anbari. 2017. Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric. *PloS one* 12, 6 (2017), e0177678.

[15] Efe Bozkir, Onur Günlü, Wolfgang Fuhl, Rafael F Schaefer, and Enkelejda Kasneci. 2021. Differential privacy for eye tracking with temporal correlations. *Plos one* 16, 8 (2021), e0255979.

[16] Efe Bozkir, Süleyman Özdel, Mengdi Wang, Brendan David-John, Hong Gao, Kevin Butler, Eakta Jain, and Enkelejda Kasneci. 2023. Eye-tracked Virtual Reality: A Comprehensive Survey on Methods and Privacy Challenges. *arXiv preprint arXiv:2305.14080* (2023).

[17] Andreas Bulling and Hans Gellersen. 2010. Toward mobile eye-based human-computer interaction. *IEEE Pervasive Computing* 9, 4 (2010), 8–12. https://doi.org/10.1109/MPRV.2010.86

[18] Andreas Bulling, Jamie A. Ward, Hans Gellersen, and Gerhard Tröster. 2011. Eye Movement Analysis for Activity Recognition Using Electrooculography. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33, 4 (2011), 741–753. https://doi.org/10.1109/TPAMI.2010.86

[19] Mihai Bâce, Sander Staal, and Andreas Bulling. 2020. Quantification of Users' Visual Attention During Everyday Mobile Device Interactions. In *Proc. ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*. https://doi.org/10.1145/3313831.3376449

[20] Jan Camenisch, Markulf Kohlweiss, Alfredo Rial, and Caroline Sheedy. 2009. Blind and anonymous identity-based encryption and authorised private searches on public key encrypted data. In *Public Key Cryptography–PKC 2009: 12th International Conference on Practice and Theory in Public Key Cryptography, Irvine, CA, USA, March 18-20, 2009. Proceedings 12*. Springer, 196–214.

[21] Jan Camenisch and Gregory M Zaverucha. 2009. Private intersection of certified sets. In *Financial Cryptography and Data Security: 13th International Conference, FC 2009, Accra Beach, Barbados, February 23-26, 2009. Revised Selected Papers 13*. Springer, 108–127.

[22] Virginio Cantoni, Chiara Galdi, Michele Nappi, Marco Porta, and Daniel Riccio. 2015. GANT: Gaze analysis technique for human identification. *Pattern Recognition* 48, 4 (2015), 1027–1038.

[23] Mathilde Caron, Piotr Bojanowski, Julien Mairal, and Armand Joulin. 2019. Unsupervised pre-training of image features on non-curated data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2959–2968.

[24] Xi Chen, Yan Duan, Rein Houthooft, John Schulman, Ilya Sutskever, and Pieter Abbeel. 2016. Infogan: Interpretable representation learning by information maximizing generative adversarial nets. *Advances in neural information processing systems* 29 (2016).

[25] Xinlei Chen and Kaiming He. 2021. Exploring simple siamese representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 15750–15758.

[26] Jose Contreras and Hardik Gajera. 2022. DeV-IP: A k-out-n Decentralized and verifiable BFV for Inner Product evaluation. *Cryptology ePrint Archive* (2022).

[27] Antoine Coutrot, Janet H Hsiao, and Antoni B Chan. 2018. Scanpath modeling and classification with hidden Markov models. *Behavior Research Methods* 50, 1 (2018), 362–379.

[28] Benjamin Cowley, Marco Filetti, et al. 2016. The psychophysiology primer: a guide to methods and a broad review with a focus on human–computer interaction. *Found. Trends Hum.-Comput. Interact.* 9, 3-4 (2016), 151–308.

[29] Emiliano De Cristofaro and Gene Tsudik. 2012. Experimenting with fast private set intersection. In *International Conference on Trust and Trustworthy Computing*. Springer, 55–73.

[30] Pierpaolo Della Monica, Ivan Visconti, Andrea Vitaletti, and Marco Zecchini. 2024. Trust Nobody: Privacy-Preserving Proofs for Edited Photos with Your Laptop. In *2025 IEEE Symposium on Security and Privacy (SP)*. IEEE Computer Society, 14–14.

[31] Whitfield Diffie. 1976. New direction in cryptography. *IEEE Trans. Inform. Theory* 22 (1976), 472–492.

[32] Heiko Drewes. 2010. *Eye gaze tracking for human computer interaction.* Ph. D. Dissertation. LMU Munich. https://doi.org/10.5282/edoc.11591

[33] Thai Duong, Duong Hieu Phan, and Ni Trieu. 2020. Catalic: Delegated PSI cardinality with applications to contact tracing. In *International Conference on the Theory and Application of Cryptology and Information Security*. Springer, 870–899.

[34] Mayar Elfares, Zhiming Hu, Pascal Reisert, Andreas Bulling, and Ralf Küsters. 2022. Federated Learning for Appearance-based Gaze Estimation in the Wild. *NeurIPS-GMML* (2022). https://doi.org/10.48550/arXiv.2211.07330

[35] Mayar Elfares, Pascal Reisert, Wenwu Tang, Zhiming Hu, Ralf Küsters, and Andreas Bulling. 2024. PrivatEyes: Appearance-based Gaze Estimation Using Federated Secure Multi-Party Computation. *ACM ETRA* (2024).

[36] Anna Maria Feit, Shane Williams, Arturo Toledo, Ann Paradiso, Harish Kulkarni, Shaun Kane, and Meredith Ringel Morris. 2017. Toward everyday gaze input: Accuracy and precision of eye tracking and implications for design. In *Proceedings of the 2017 Chi conference on human factors in computing systems*. 1118–1130.

[37] Jesús García-Rodríguez, Stephan Krenn, and Daniel Slamanig. 2024. To pass or not to pass: Privacy-preserving physical access control. *Computers & Security* 136 (2024), 103566.

[38] Craig Gentry. 2009. Fully homomorphic encryption using ideal lattices. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*. 169–178.

[39] Oded Goldreich. 2004. *Foundations of Cryptography, Volume 2.* Cambridge university press Cambridge.

[40] Oded Goldreich, Silvio Micali, and Avi Wigderson. 2019. How to play any mental game, or a completeness theorem for protocols with honest majority. In *Providing Sound Foundations for Cryptography: On the Work of Shafi Goldwasser and Silvio Micali*. 307–328.

[41] Priya Goyal, Mathilde Caron, Benjamin Lefaudeux, Min Xu, Pengchao Wang, Vivek Pai, Mannat Singh, Vitaliy Liptchinsky, Ishan Misra, Armand Joulin, et al. 2021. Self-supervised pretraining of visual features in the wild. *arXiv preprint arXiv:2103.01988* (2021).

[42] Céline Gressel, Rebekah Overdorf, Inken Hagenstedt, Murat Karaboga, Helmut Lurtz, Michael Raschke, and Andreas Bulling. 2023. Privacy-Aware Eye Tracking: Challenges and Future Directions. *IEEE Pervasive Computing* 22, 1 (2023), 95–102. https://doi.org/10.1109/MPRV.2022.3228660

[43] Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. 2017. beta-vae: Learning basic visual concepts with a constrained variational framework. In *International conference on learning representations*.

[44] Kenneth Holmqvist, Marcus Nyström, and Fiona Mulvey. 2012. Eye tracker data quality: What it is and how to measure it. In *Proceedings of the symposium on eye tracking research and applications*. 45–52.

[45] Kenneth Holmqvist, Marcus Nyström, and Fiona Mulvey. 2012. Eye tracker data quality: what it is and how to measure it. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (Santa Barbara, California) *(ETRA '12)*. Association for Computing Machinery, New York, NY, USA, 45–52. https://doi.org/10.1145/2168556.2168563

[46] Sabrina Hoppe, Tobias Loetscher, Stephanie A Morey, and Andreas Bulling. 2018. Eye movements during everyday behavior predict personality traits. *Frontiers in Human Neuroscience* (2018), 105. https://doi.org/10.3389/fnhum.2018.00105

[47] Michael Xuelin Huang, Tiffany CK Kwok, Grace Ngai, Stephen CF Chan, and Hong Va Leong. 2016. Building a personalized, auto-calibrating eye tracker from user interactions. In *ACM CHI*. 5169–5179.

[48] Michael Xuelin Huang, Jiajia Li, Grace Ngai, and Hong Va Leong. 2016. Stressclick: Sensing stress from gaze-click patterns. In *Proceedings of the 24th ACM international conference on Multimedia*. 1395–1404.

[49] Michael Xuelin Huang, Jiajia Li, Grace Ngai, and Hong Va Leong. 2017. Screenglint: Practical, in-situ gaze estimation on smartphones. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 2546–2557.

[50] Bernardo A Huberman, Matt Franklin, and Tad Hogg. 1999. Enhancing privacy and trust in electronic communities. In *Proceedings of the 1st ACM conference on Electronic commerce*. 78–86.

[51] Apple Inc. 2021. CSAM Detection - Technical Summary. https://www.apple.com/child-safety/pdf/CSAM_Detection_Technical_Summary.pdf.

[52] Yuval Ishai, Joe Kilian, Kobbi Nissim, and Erez Petrank. 2003. Extending oblivious transfers efficiently. In *Annual International Cryptology Conference*. Springer, 145–161.

[53] Deborah N. Jakobi, Daniel G. Krakowczyk, and Lena A. Jäger. 2024. Reporting Eye-Tracking Data Quality: Towards a New Standard. In *Proceedings of the 2024 Symposium on Eye Tracking Research and Applications* (<conf-loc>, <city>Glasgow</city>, <country>United Kingdom</country>, </conf-loc>) *(ETRA '24)*. Association for Computing Machinery, New York, NY, USA, Article 47, 3 pages. https://doi.org/10.1145/3649902.3655658

[54] Kaiwen Jiang, Shu-Yu Chen, Hongbo Fu, and Lin Gao. 2023. NeRFFaceLighting: Implicit and Disentangled Face Lighting Representation Leveraging Generative Prior in Neural Radiance Fields. *ACM Transactions on Graphics* 42, 3 (2023), 1–18.

[55] Swati Jindal and Roberto Manduchi. 2023. Contrastive representation learning for gaze estimation. In *Annual Conference on Neural Information Processing Systems*. PMLR, 37–49.

[56] Kaggle. 2024. Kaggle: Your Machine Learning and Data Science Community. https://www.kaggle.com.

[57] Daniel Kales, Christian Rechberger, Thomas Schneider, Matthias Senker, and Christian Weinert. 2019. Mobile private contact discovery at scale. In *28th USENIX Security Symposium (USENIX Security 19)*. 1447–1464.

[58] Thivya Kandappu, Archan Misra, Shih-Fen Cheng, Randy Tandriansyah, and Hoong Chuin Lau. 2018. Obfuscation at-source: Privacy in context-aware mobile crowd-sourcing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (2018), 1–24.

[59] Mohamed Khamis, Daniel Buschek, Tobias Thieron, Florian Alt, and Andreas Bulling. 2017. EyePACT: Eye-Based Parallax Correction on Touch-Enabled Interactive Displays. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)* 1, 4 (2017), 1–18. https://doi.org/10.1145/3161168

[60] Mohamed Khamis, Ludwig Trotter, Ville Mäkelä, Emanuel von Zezschwitz, Jens Le, Andreas Bulling, and Florian Alt. 2018. CueAuth: Comparing Touch, Mid-Air Gestures, and Gaze for Cue-based Authentication on Situated Displays. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)* 2, 7 (2018), 1–22. https://doi.org/10.1145/3287052

[61] Miran Kim and Kristin Lauter. 2015. Private genome analysis through homomorphic encryption. In *BMC medical informatics and decision making*, Vol. 15. Springer, 1–12.

[62] Vladimir Kolesnikov, Ranjit Kumaresan, Mike Rosulek, and Ni Trieu. 2016. Efficient batched oblivious PRF with applications to private set intersection. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. 818–829.

[63] Kyle Krafka, Aditya Khosla, Petr Kellnhofer, Harini Kannan, Suchendra Bhandarkar, Wojciech Matusik, and Antonio Torralba. 2016. Eye tracking for everyone. In *IEEE ICPR*. 2176–2184.

[64] Huining Li, Xiaoye Qian, Ruokai Ma, Chenhan Xu, Zhengxiong Li, Dongmei Li, Feng Lin, Ming-Chun Huang, and Wenyao Xu. 2023. TherapyPal: Towards a Privacy-Preserving Companion Diagnostic Tool based on Digital Symptomatic Phenotyping. In *Proceedings of the 29th Annual International Conference on Mobile Computing and Networking*. 1–15.

[65] Jingjie Li, Amrita Roy Chowdhury, Kassem Fawaz, and Younghyun Kim. 2021. {Kalε ido}:{Real-Time} Privacy Control for {Eye-Tracking} Systems. In *30th USENIX Security Symposium*. 1793–1810.

[66] Ao Liu, Lirong Xia, Andrew Duchowski, Reynold Bailey, Kenneth Holmqvist, and Eakta Jain. 2019. Differential privacy for eye-tracking data. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*. 1–10.

[67] Gang Liu, Yuechen Yu, Kenneth Alberto Funes Mora, and Jean-Marc Odobez. 2018. A differential approach for gaze estimation with calibration.. In *BMVC*, Vol. 2. 6.

[68] Dillon J Lohr, Lee Friedman, and Oleg V Komogortsev. 2019. Evaluating the data quality of eye tracking signals from a virtual reality system: Case study using SMI's eye-tracking HTC vive. *arXiv preprint arXiv:1912.02083* (2019).

[69] Päivi Majaranta and Andreas Bulling. 2014. Eye tracking and eye-based human–computer interaction. *Advances in physiological computing* (2014), 39–65. https://doi.org/10.1007/978-1-4471-6392-3_3

[70] Brian W Matthews. 1975. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta (BBA)-Protein Structure* 405, 2 (1975), 442–451.

[71] Meta. 2019. Open-Sourcing Photo- and Video-Matching Technology to Make the Internet Safer. https://about.fb.com/news/2019/08/open-source-photo-video-matching/.

[72] Microsoft. 2015. PhotoDNA. https://www.microsoft.com/en-us/photodna.

[73] Dimitris Mouris, Daniel Masny, Ni Trieu, Shubho Sengupta, Prasad Buddhavarapu, and Benjamin Case. 2024. Delegated Private Matching for Compute. *Proceedings on Privacy Enhancing Technologies* (2024).

[74] Rasmus Pagh and Flemming Friche Rodler. 2001. Cuckoo hashing. In *European Symposium on Algorithms*. Springer, 121–133.

[75] Seonwook Park, Shalini De Mello, Pavlo Molchanov, Umar Iqbal, Otmar Hilliges, and Jan Kautz. 2019. Few-shot adaptive gaze estimation. In *Proceedings of the IEEE/CVF international conference on computer vision*. 9368–9377.

[76] Benny Pinkas, Mike Rosulek, Ni Trieu, and Avishay Yanai. 2020. PSI from PaXoS: fast, malicious private set intersection. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*. Springer, 739–767.

[77] Michael O Rabin. 2005. How to exchange secrets with oblivious transfer. *Cryptology ePrint Archive* (2005).

[78] Christoph Sager, Christian Janiesch, and Patrick Zschech. 2021. A survey of image labelling for computer vision applications. *Journal of Business Analytics* 4, 2 (2021), 91–110.

[79] Negar Sammaknejad, Hamidreza Pouretemad, Changiz Eslahchi, Alireza Salahirad, and Ashkan Alinejad. 2017. Gender classification based on eye movements: A processing effect during passive face viewing. *Advances in Cognitive Psychology* 13, 3 (2017), 232.

[80] Ryo Shimata, Yoshihiro Mitani, and Tsumoru Ochiai. 2015. A study of pupil detection and tracking by image processing techniques for a human eye-computer interaction system. In *2015 IEEE/ACIS 16th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*. IEEE, 1–4.

[81] Brian A Smith, Qi Yin, Steven K Feiner, and Shree K Nayar. 2013. Gaze locking: passive eye contact detection for human-object interaction. In *Proceedings of the 26th annual ACM symposium on User interface software and technology*. 271–280.

[82] Qiyang Song, Jiahao Cao, Kun Sun, Qi Li, and Ke Xu. 2021. Try before you buy: Privacy-preserving data evaluation on cloud-based machine learning data marketplace. In *Annual Computer Security Applications Conference*. 260–272.

[83] Julian Steil and Andreas Bulling. 2015. Discovery of everyday human activities from long-term visual behaviour using topic models. In *ACM UbiComp (UbiComp)*. ACM, 75–85. https://doi.org/10.1145/2750858.2807520

[84] Julian Steil, Inken Hagestedt, Michael Xuelin Huang, and Andreas Bulling. 2019. Privacy-aware eye tracking using differential privacy. In *ACM ETRA*. 1–9. https://doi.org/10.1145/3314111.3319915

[85] Yusuke Sugano, Yasuyuki Matsushita, and Yoichi Sato. 2014. Learning-by-synthesis for appearance-based 3d gaze estimation. In *IEEE ICPR*. 1821–1828.

[86] Yunjia Sun, Jiabei Zeng, Shiguang Shan, and Xilin Chen. 2021. Cross-encoder for unsupervised gaze representation learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 3702–3711.

[87] Erkam Uzun, Simon P Chung, Vladimir Kolesnikov, Alexandra Boldyreva, and Wenke Lee. 2021. Fuzzy labeled private set intersection with applications to private {Real-Time} biometric search. In *30th USENIX Security Symposium (USENIX Security 21)*. 911–928.

[88] Yunhan Wang, Xiangwei Shi, Shalini De Mello, Hyung Jin Chang, and Xucong Zhang. 2023. Investigation of Architectures and Receptive Fields for Appearance-based Gaze Estimation. arXiv:2308.09593 [cs.CV]

[89] Xueting Yan, Ishan Misra, Abhinav Gupta, Deepti Ghadiyaram, and Dhruv Mahajan. 2020. Clusterfit: Improving generalization of visual representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 6509–6518.

[90] Victoria Yaneva, Le An Ha, Sukru Eraslan, Yeliz Yesilada, and Ruslan Mitkov. 2018. Detecting autism based on eye-tracking data from web searching tasks. In *Proceedings of the 15th International Web for All Conference*. 1–10.

[91] Edwin Yang, Qiuye He, and Song Fang. 2022. WINK: Wireless Inference of Numerical Keystrokes via Zero-Training Spatiotemporal Analysis. In *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security (CCS' 2022)*. ACM, 3033–3047.

[92] Andrew C Yao. 1982. Protocols for secure computations. In *23rd annual symposium on foundations of computer science (sfcs 1982)*. IEEE, 160–164.

[93] Yu Yu and Jean-Marc Odobez. 2020. Unsupervised representation learning for gaze estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision*

and Pattern Recognition. 7314–7324.

[94] Xucong Zhang, Seonwook Park, Thabo Beeler, Derek Bradley, Siyu Tang, and Otmar Hilliges. 2020. Eth-xgaze: A large scale dataset for gaze estimation under extreme head pose and gaze variation. In *European Conference on Computer Vision*. Springer, 365–381.

[95] Xucong Zhang, Yusuke Sugano, and Andreas Bulling. 2017. Everyday eye contact detection using unsupervised gaze target discovery. In *Proceedings of the 30th annual ACM symposium on user interface software and technology*. 193–203.

[96] Xucong Zhang, Yusuke Sugano, Mario Fritz, and Andreas Bulling. 2015. Appearance-based Gaze Estimation in the Wild. In *IEEE CVPR*. 4511–4520. https://doi.org/10.1109/CVPR.2015.7299081

[97] Xucong Zhang, Yusuke Sugano, Mario Fritz, and Andreas Bulling. 2017. It's written all over your face: Full-face appearance-based gaze estimation. In *IEEE CVPR Workshops*. 51–60.

[98] Xucong Zhang, Yusuke Sugano, Mario Fritz, and Andreas Bulling. 2019. MPIIGaze: Real-World Dataset and Deep Appearance-Based Gaze Estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41, 1 (2019), 162–175. https://doi.org/10.1109/TPAMI.2017.2778103

[99] Ke Zhong and Sebastian Angel. 2024. Oryx: Private detection of cycles in federated graphs. *Cryptology ePrint Archive* (2024).