ABSTRACT
Eye-tracking is a critical source of information for understanding human behavior and developing future mixed-reality technology. Eye-tracking enables applications that classify user activity or predict user intent. However, eye-tracking datasets collected during common virtual reality tasks have also been shown to enable unique user identification, which creates a privacy risk. In this paper, we focus on the problem of user re-identification from eye-tracking features. We adapt standardized privacy definitions of \( k \)-anonymity and plausible deniability to protect datasets of eye-tracking features, and evaluate performance against re-identification by a standard biometric identification model on seven VR datasets. Our results demonstrate that re-identification goes down to chance levels for the privatized datasets, even as utility is preserved to levels higher than 72% accuracy in document type classification.

CCS CONCEPTS
- Computing methodologies → Image processing;
- Security and privacy → Privacy protections; Human and societal aspects of security and privacy;
- Human-centered computing → Ubiquitous and mobile computing; Human computer interaction (HCI).

KEYWORDS
privacy, eye tracking, biometrics, re-identification, \( k \)-anonymity, plausible deniability

ACM Reference Format:

1 INTRODUCTION
Re-identification attacks in literature have been extensively explored for social networks [Narayanan and Shmatikov 2009], location data [Primault et al. 2018], and medical data [El Emam et al. 2011]. Real-world re-identification attacks have been demonstrated to learn the medical prescriptions of a politician [Sweeney 2002] or reveal the Netflix preferences of half of a million users [Narayanan and Shmatikov 2009]. As a result of the Netflix dataset attack, a woman sued the company over the risk that her leaked viewing patterns would reveal her sexual orientation to her family [Singel 2009]. There are an increasing number of algorithms that can authenticate a user based on eye movement data [George and Routray 2016; Lohr et al. 2021; Schröder et al. 2020; Slaganovic et al. 2018]. Numerous datasets of eye-tracking data for virtual reality (VR) applications are publicly available [David-John et al. 2021a; Emery et al. 2021; Hu et al. 2021; Sitzmann et al. 2018; Steil et al. 2019; Xu et al. 2018]. Taken together, this means that re-identification attacks using eye movements are not only plausible, but imminent.

Do people care? Surveys by Adams et al. [2018] and Steil et al. [2019] have established that both users and developers have privacy concerns over VR and eye-tracking data collection and how they are applied to make inferences about the user. For example, VR developers have cited that they are aware of privacy concerns for users and share their sentiments; however, most developers are not privacy experts and there is a lack of standards for how to address topics like ethics or privacy issues. For users, survey participants have indicated that they would be willing to accept beneficial VR applications that collect eye-tracking data if they are sharing the data with trusted governmental health agencies or with a university for research purposes. The same users also responded that they would not share their data publicly or with private services, unless there were constraints in place for how the data was being used.

Is regulation the answer? Privacy laws in certain regions are designed to protect traditional biometric identifiers, such as iris patterns and face scans [Heller 2020]. However, legal scholars have pointed out that privacy laws rarely hold up in court, and would not apply to behavioral data streams due to ambiguous wording over what is considered a biometric [Roberg-Perez 2016]. A lack of enforceable privacy laws and data release standards implies that VR platforms could store or sell identities through eye-tracking and behavioral data captured alongside demographics, which are typically used for personalized ads on the web [Datta et al. 2015].

Scope and contributions. In this paper we propose two novel adaptations of privacy mechanisms to achieve \( k \)-anonymity and plausible deniability (PD) guarantees for datasets of eye-tracking features. We compared our mechanisms against the previously established Exponential mechanism for DP. We found that our \( k \)-same-select sequence approach defended against re-identification and achieved superior utility in document type recognition (≥72%).

2 RELATED WORK
Mechanisms that achieve formal privacy guarantees have been explored for protecting eye-tracking data against re-identification attacks for gaze samples [Li et al. 2020] and for features extracted...
from gaze data [Bozkir et al. 2021; Steil et al. 2019]. The un-shaded rows in Table 1 lists existing mechanisms that achieve formal privacy guarantees for eye-tracking data, type of input data, and how they were adapted to eye-tracking. The only formal privacy guarantee that has been explored is differential privacy (DP). While DP is popular in the privacy community due to the robust definition, there is an inevitable trade-off between increased DP privacy and lower data utility [Kifer and Machanavajjhala 2011].

We consider protecting eye-tracking datasets against re-identification attacks through alternative privacy guarantees. First, we explored $k$-anonymity to provide intuitive protection in that individual data cannot be distinguished from $k$-1 others. By adapting $k$-same-select [Gross et al. 2005], an upper bound on attack success is established while retaining utility. However, this approach releases $k$ copies of the same data values. From an eye-tracking perspective, releasing duplicate data is not a satisfying solution. We shifted to considering PD, which extends a similar intuition for synthetic data. Synthetic data retains utility by reproducing characteristics of the original data. We explored guarantees specific to re-identification, and found superior utility with $k$-anonymity and that synthetic data has promise for preserving privacy in eye-tracking datasets.

The presented mechanisms are intended to be applied to datasets prior to their release. In contrast, for real-time systems, methods such as a privacy-preserving API and real-time perturbations [David-John et al. 2021a; Li et al. 2020] will enable platforms to share samples, features and gaze-based metrics with third-party applications.

### 3 METHODOLOGY

We conducted an evaluation of re-identification attacks on eye-tracking features and apply privacy mechanisms to protect identity. This section describes the protocol for re-identification attacks, privacy mechanisms for processing features, datasets included in the evaluation, and the approach used for biometric classification.

#### 3.1 Threat Model

We assume that an adversary has access to a public eye-tracking dataset. The adversary trains an identification model to take eye-tracking feature vectors as input and output the associated identity. Given new eye-tracking data, from playing a VR game for example, the adversary can use the trained model to guess at the identity of the player. The re-identification attack is successful if the correct identity is returned.

#### 3.2 Proposed Solution

We propose two privacy mechanisms that can be applied to the eye tracking dataset prior to releasing it for public use. Thus, the adversary will train their model on privatized datasets. We assume that the adversary acquires un-privatized, i.e., raw data for the purposes of the re-identification attack, which is considered the test set. The privacy mechanisms are successful if they reduce the rate of re-identification to below chance levels.

#### 3.3 Privacy Mechanisms

In this section, we contribute two privacy mechanisms, one that satisfies $k$-anonymity and one that satisfies plausible deniability. We provide pseudocode for ease of re-implementation and publicly release code for $k$-same-select sequence. Both mechanisms are adaptations of prior work to consider eye-tracking features. For completeness, we provide pseudocode for our implementation of the DP-oriented mechanism defined by Steil et al. [2019].

The presented mechanisms are intended to be applied to datasets prior to their release. In contrast, for real-time systems, methods such as a privacy-preserving API and real-time perturbations [David-John et al. 2021a; Li et al. 2020] will enable platforms to share samples, features and gaze-based metrics with third-party applications.
the original $k$-same-select method and blue indicates our adapted steps. The data from all individuals are processed sequentially, i.e., the first feature vector of all individuals viewing a specific stimulus within a given task are randomly placed into groups of size $k$ to compute average values for release. The mechanism assumes that there is data from at least $k$ individuals available for grouping. The same groupings of individuals are used for each stimulus to achieve $k$-anonymity across the entire sequence of feature vectors. The adapted sequence mechanism is generalized by processing feature vectors in sequence; however, there is no guarantee that each individual had the same number of feature vectors per stimulus. Data are padded to repeat the last feature vector in the sequence for individuals with less features.

### Task-based Marginals Model (PD)

Plausible deniability (PD) is not a condition of a privacy mechanism, but instead a privacy criterion that is checked before data is released [Bindschaedler et al. 2017]. Any number of approaches can be applied to generate data that satisfies PD. A generative model takes a raw feature vector as input and PD establishes that at least $k$ other inputs from the original dataset could have plausibly generated the output synthetic feature. A parameter $y$ is used to control how close relative probabilities must be to be considered plausible, and $k$ controls the number of features from the original dataset that have to pass the privacy test before synthetic data can be released. The formal definition and steps to implement the privacy test are detailed in the Supplementary Material.

To achieve PD we applied the Marginals approach with publicly available code [Bindschaedler et al. 2017]. Marginals builds a distribution of discrete values for each feature column and releases synthetic data by randomly sampling each feature independently. The learned feature distributions are representative of each task. Resulting distributions are used to synthesize data by task and retain utility. We adapted this approach by binning each continuous feature into $B = 30$ discrete buckets over the range of values (blue lines of code).

The generated synthetic feature vectors consist of discrete values corresponding to buckets that cover the range of feature values. We sample values between the minimum and maximum value range from the corresponding bucket from a random uniform distribution to map synthetic data back into continuous feature values. The synthetic dataset is stratified to contain the same number of feature vectors from each individual for each task as the original dataset. The PD guarantee differs from $k$-anonymity, in that PD guarantees $k - 1$ other features from the original dataset could have generated the synthetic output, while $k$-anonymity guarantees that $k - 1$ other individuals could have generated a sequence of output features.

#### Exponential-DP Mechanism

The Exponential-DP noise mechanism was proven to be $\epsilon$-DP by Stein et al. [2019] and applies to each individual feature in the feature set. Exponential noise is sampled independently for each feature vector and depends on the range of each feature and the task duration. The first step in applying Exponential-DP is to compute the range $\delta_i$ for each feature $i$ as the maximum value minus the minimum value. The maximum number of feature vectors $t_{\text{max}}$ from any individual during viewing is used for padding the data from other individuals. The last feature vector recorded for an individual is repeated to ensure that each individual has $t_{\text{max}}$ total feature vectors. For each feature a value $y_i$ is sampled from an Exponential distribution with a scale of $\frac{1}{\delta_i}$, where $\lambda = \frac{e^{-\delta_i} \cdot y_i}{t_{\text{max}}}$.

The additive noise is then computed as $r = \frac{\log(y_i)}{t_{\text{max}}}$ and the positive or negative sign is randomly assigned. Values of $r$ are computed for every feature from the task, and are added to the original data to produce noisy feature vectors to release.

### 3.4 Datasets

We evaluate the above detailed privacy mechanisms on publicly available VR datasets of eye-tracking features. The datasets vary based on the number of individuals, amount of data available, task being performed, and type of stimulus being viewed. Table 2 summarizes the characteristics of datasets included in our evaluation.

### 3.5 Feature Sets

Six of the datasets listed in Table 2 release raw gaze sample data, while MPIIDPEye included both raw samples and a set of precomputed sliding windows of gaze-based features [Bulling et al. 2010]. To maintain consistency with past results from MPIIDPEye, we used their feature set in our analysis of this dataset. For all other datasets we used the feature set of the author for that dataset. We compare our results to published results in the literature.

---

1https://vbinds.ch/node/69

---

1Due to The Composition Theorem, Exponential-DP achieves a guarantee of $t$ times the number of features. For consistency with [Steil et al. 2019], we reference $\epsilon$ as the noise parameter for each feature, and not the composed guarantee.
datasets, we replicate the approach from David-John et al. [2021a] and extract features from fixation and saccade events detected using the I-SsT algorithm with default parameters [Agtzidis et al. 2019a]. The features extracted from fixation and saccades events leverage common statistics such as duration and amplitude, as well as the velocity and acceleration of gaze during the event [George and Routray 2016]. A feature set is generated for each type of event and a separate classification model is trained for each feature set.

### 3.6 Biometric Classifier

A Radial Basis Function (RBF) network is used to classify identity using feature vectors as input and is commonly used to identify users from eye-tracking data [David-John et al. 2021a; George and Routray 2016; Schröder et al. 2020]. An RBF network features a single hidden layer of nodes consisting of activation functions. The output of the activation functions is weighted to generate a probability that input is from each target class. The predicted class with the highest probability is considered the individual most likely to have produced the input feature vector, which is then used for biometric identification. Biometric identification relies on a set of features from an unknown individual viewing at least one stimulus. The feature vectors from all stimuli for an unknown individual are classified by the network, and the output scores are used to predict identity by averaging prediction scores.

As described in Section 3.5, the majority of datasets included in our evaluation use features extracted from both fixation and saccade events, requiring an RBF network trained independently on both features [George and Routray 2016]. The output identification scores are first averaged within each type of event, then a final classification is made with a weighted average between fixation and saccade scores. A weight of 0.4 was applied to the fixation scores with a weight of 0.6 for saccade scores, as saccade features provided a slightly higher accuracy in user identification. For MPIIDPeye the prediction scores from all inputs within a task are simply averaged before classifying identity.

### 4 RESULTS

In this section, we present privacy and utility metrics to evaluate the implemented privacy mechanisms from Section 3.3 for each dataset listed in Table 2. We compared our proposed privacy mechanisms with Exponential-DP as an established approach for DP. Section 4.1 presents identification rates for each privacy mechanism using a biometric identification model trained on processed data and tested on the original data. Section 4.2 presents utility results for document type recognition on the MPIIDPeye dataset.

### 4.1 Biometric Identification

Re-identification risk for eye-tracking data is evaluated by splitting eye-tracking features into training sets processed by privacy mechanisms and testing sets of unmodified data. Identification rates higher than chance, which is one divided by the number of individuals in a dataset, indicate that there is risk of re-identification

---

### Table 2: Characteristics of VR eye-tracking datasets.

<table>
<thead>
<tr>
<th>Dataset</th>
<th># Pts.</th>
<th>Chance Rate</th>
<th># Stim</th>
<th>Data Per Ppt</th>
<th>Stimuli Type</th>
<th>Task</th>
</tr>
</thead>
<tbody>
<tr>
<td>MPIIDPeye [Steil et al. 2019]</td>
<td>20</td>
<td>1/20 Ppts = 5.0%</td>
<td>3</td>
<td>30 mins</td>
<td>Documents</td>
<td>VR Reading</td>
</tr>
<tr>
<td>ET-DK2 [David-John et al. 2021a]</td>
<td>18</td>
<td>1/18 Ppts = 5.5%</td>
<td>50</td>
<td>21 mins</td>
<td>360° Images</td>
<td>Free Viewing</td>
</tr>
<tr>
<td>VR-Saliency [Sitzmann et al. 2018]</td>
<td>130</td>
<td>1/130 Ppts = 0.8%</td>
<td>8</td>
<td>4 mins</td>
<td>360° Images</td>
<td>Free Viewing</td>
</tr>
<tr>
<td>360_em [Agtzidis et al. 2019b]</td>
<td>13</td>
<td>1/13 Ppts = 7.7%</td>
<td>14</td>
<td>17 mins</td>
<td>360° Videos</td>
<td>Free Viewing</td>
</tr>
<tr>
<td>VR-EyeTracking [Xu et al. 2018]</td>
<td>43</td>
<td>1/43 Pts = 2.3%</td>
<td>208</td>
<td>Avg: 88 mins</td>
<td>360° Videos</td>
<td>Free Viewing</td>
</tr>
<tr>
<td>OpenEDS [Emery et al. 2021]</td>
<td>44</td>
<td>1/44 Pts = 2.3%</td>
<td>2</td>
<td>10 mins</td>
<td>3D Scene</td>
<td>Free Exploration</td>
</tr>
<tr>
<td>EHTask [Hu et al. 2021]</td>
<td>30</td>
<td>1/30 Ppts = 3.3%</td>
<td>15</td>
<td>30 mins</td>
<td>360° Videos</td>
<td>Free Viewing, Search, Saliency, Track</td>
</tr>
</tbody>
</table>

---

### Figure 1: Privacy evaluation for identification rate from eye-tracking features.

Privatizing the dataset with our presented mechanisms lowers all identification rates to chance for \(k = 8\) in \(k\)-same and Marginals, and \(\epsilon = 2\) for Exponential-DP. Chance identification rates demonstrate that identity is protected within a group of individuals. The different datasets contain eye-tracking data on tasks performed within a variety of VR environments (reading documents, 360° images, 360° videos, and 3D rendered scenes). Chance rates (1/#Ppts.) vary for each dataset based on the number of identities, and are listed in Table 2.
from released data. Figure 1 presents the identification rates for each dataset and mechanism. The ET-DK2 dataset produced the highest identification rate of all datasets with 100% identification with the original data. All datasets produced identification rates higher than chance prior to privacy mechanisms being applied.

When privacy mechanisms were applied, the identification rates of all datasets dropped to chance. The Exponential-DP and Marginals approaches degraded the identification rates to chance across all parameter values. The only exception was MPIIDPEye for Exponential-DP, which required a parameter value of $\varepsilon = 100$ for an identification rate of 6%, compared to a chance rate of 5%. $k$-same also reduces identification rates to chance, with a larger value of $k$ needed to bring ET-DK2 to chance (5.6%). Our results suggest that privacy mechanisms protect against re-identification attacks on eye-tracking features using a standard biometric identification model.

### 4.2 Utility Evaluation

Releasing a privacy-preserving dataset that is useful relies on achieving a practical level of utility. We evaluated utility for each privacy mechanism applied to the MPIIDPEye dataset to classify document type using gaze features.

Steil et al. [2019] first evaluated MPIIDPEye using an SVM model to classify document type as either Comic, Newspaper, or Textbook. The SVM used an RBF kernel, bias parameter $C$ set to one, and expressivity parameter $\gamma$ set to one divided by the number of features. The model was trained on data from each individual during the first half of reading that was processed by the privacy mechanism, and tested on data from the second half. Figure 2 presents feature-level model accuracy results for each mechanism. Each plot demonstrates utility relative to the original data and chance rate of guessing (33%). We observed that the Exponential-DP mechanism reduced accuracy to chance, or near chance rates. For Exponential-DP, accuracy started at 80% for $\varepsilon = 100$, and fell to chance at $\varepsilon = 10$. For Marginals, a low level of utility was retained as accuracy remained near 53% for all parameters. The $k$-same approach was stable across parameter values, with slightly lower accuracy for higher levels of $k$. $k$-same across all parameters maintained performance greater than 72%. This level of accuracy would be practical for an assistive reading interface that needs to identify the correct document type the majority of the time [Toyama et al. 2013].

### 5 CONCLUSION AND DISCUSSION

This paper addresses the open challenge of applying formal privacy definitions to behavioral data streams. Our work is the first to adapt the definitions of $k$-anonymity and PD to eye-tracking features. The definition of $k$-anonymity is intuitive as the theoretical risk of re-identification attacks are bounded above by $\frac{1}{k}$. The $k$-same-select sequence mechanism produced identification rates at chance while preserving model accuracy of 72% for document type classification. PD is a promising privacy criterion as it provides a clear interpretation with respect to re-identification, similar to $k$-anonymity; while using synthetic data to preserve privacy and retain utility. A Marginals mechanism for PD retains slight utility with an accuracy of 53% compared to a 33% guess rate. Deploying PD is computationally expensive, as a large-scale dataset of synthetic candidates are first generated before applying the privacy test. It took less than a minute to execute $k$-same and Exponential-DP, compared to roughly 30 minutes to generate and test synthetic data. Both $k$-same and Marginals mechanisms retain stable utility across their parameters, while the Exponential mechanism loses utility at the level of privacy needed for chance rates of identification.

**Implications.** The presented adaptations offer alternatives to DP, and demonstrate higher utility at chance rates for document type recognition. We recommend using $k$-same-select sequence for classification-based datasets to protect against re-identification as it is computationally efficient with an intuitive privacy guarantee.

**Limitations.** Our identification results were limited to an RBF network, although prior work explored random forest [Schröder et al. 2020], SVM [Miller et al. 2020], k-NNs [Bozkir et al. 2021] and deep network [Miller et al. 2021] models. In terms of DP, we only evaluated the Exponential-DP mechanism, although an alternative formulation of DP exists for time-series in the frequency domain [Bozkir et al. 2021], while limitations impact the generalization of our empirical results, it does not impact the theoretical framing and comparison of privacy definitions.
Future Work. Our work provides motivation to adapt privacy guarantees to VR behavioral data in the form eye tracking. It would be useful to explore how well privacy methods preserve utility for other classification-based applications, such as intent prediction [David-John et al. 2021b]. Beyond exploring additional datasets and utilities, the field of eye-tracking privacy would benefit from further development of approaches related to PD. Such techniques can achieve an intuitive definition of privacy while preserving utility through synthetic data that appears real. Our proposed privacy mechanisms can also be applied to a breadth of mixed-reality sensors, including head and hand tracking, EEG, and EMG data.

ACKNOWLEDGMENTS
Ekta Jain acknowledges funding from NSF Award #2026540. Brendan David-John acknowledges funding from a Google PhD Fellowship and the National Science Foundation GRFP (Awards DGE-#1315138 and DGE-#1842473). Kevin Butler acknowledges funding from NSF Award CNS-#1815883 and CNS-#1562485, and AFOSR award FA950-19-1-0169.

REFERENCES


Sharon Roberg-Perez. 2016. The future is now: Biometric information and data privacy. Antitrust 31 (2016), 60.


