Gaze3P: Gaze-Based Prediction of User-Perceived Privacy

Mayar Elfares
University of Stuttgart, Germany
mayar.elfares@vis.uni-stuttgart.de

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

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

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

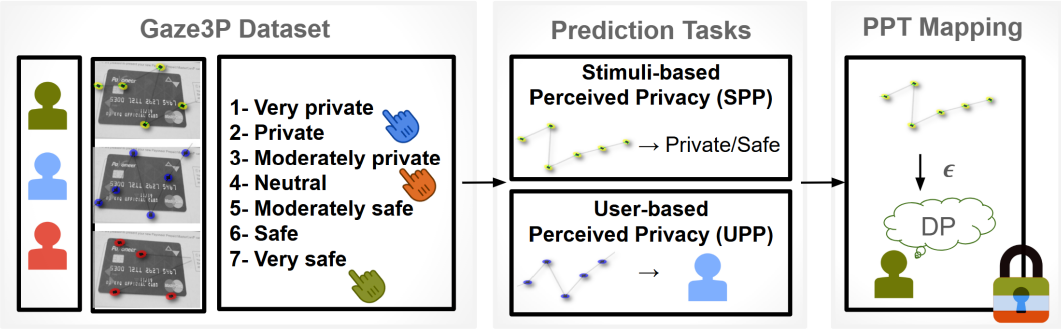


Figure 1: Users deal with various types of information in daily life that can have vastly different privacy requirements, e.g., personal photographs, passwords, or social media posts. Gaze3P is the first large-scale dataset that allows for the systematic study of user-perceived privacy. We report extensive experiments demonstrating the feasibility of predicting perceived privacy from human eye gaze. We also show how predicted privacy can be used to optimise the parameters of privacy-preserving techniques for data analysis and learning, such as Differential Privacy (DP), to better align them with user expectations.

Abstract

Privacy is a highly subjective concept and perceived variably by different individuals. Previous research on quantifying user-perceived privacy has primarily relied on questionnaires. Furthermore, applying user-perceived privacy to optimise the parameters of privacy-preserving techniques (PPT) remains insufficiently explored. To address these limitations, we introduce Gaze3P – the first dataset specifically designed to facilitate systematic investigations into user-perceived privacy. Our dataset comprises gaze data from 100 participants and 1,000 stimuli, encompassing a range of private and safe attributes. With Gaze3P we train a machine learning model to implicitly and dynamically predict perceived privacy from human eye gaze. Through comprehensive experiments, we show that the resulting models achieve high accuracy. Finally, we illustrate how predicted privacy can be used to optimise the parameters of differentially private mechanisms, thereby enhancing their alignment with user expectations.

Keywords

Gaze, Eye Tracking, Perceived Privacy, Privacy Quantification, Differential Privacy

1 Introduction

Privacy, particularly as it is perceived by individuals, is a complex and deeply subjective construct that varies significantly across contexts, cultures, and personal experiences [21, 28, 89]. Unlike technical privacy, which can be quantified through cryptographic guarantees or formal metrics, perceived privacy refers to an individual's internal judgment about the sensitivity or appropriateness of data sharing [9, 70]. Understanding and quantifying user-perceived privacy is essential because it directly influences users' willingness to engage with digital systems, share information, or consent to data sharing requests [11, 60, 89]. Therefore, accurately quantifying perceived privacy helps designers create user-aligned privacy mechanisms, improve transparency, and ultimately enhance user satisfaction and system usability [11, 56, 61, 74].

The ability to quantify user-perceived privacy levels also has significant potential for optimising the parameters of security protocols, such as Differential Privacy (DP) [31]. However, despite continuing discussions, the problem of how to map users' privacy perception to protocol parameters remains unsolved [21, 21, 28, 28, 67, 89, 102]. A key reason for this failure is the large number of factors that affect privacy perception, such as (i) oversight of the situational diversity [87], (ii) neglect of within- vs. between-subject

Proceedings on Privacy Enhancing Technologies 2026(1), 6–25

https://doi.org/10.56553/popets-2026-0002

variations [89], (iii) effects of biases, heuristics, or impulsivity on online user behaviour [67, 102], and, most importantly, (iv) the scarcity of the available behavioural data that encapsulates all relevant aspects. Previous work mainly relied on explicit feedback, such as questionnaires, which has been shown to not align well with users' satisfaction [101, 118], especially since user judgement dynamically changes depending on context, behaviour, or knowledge [55].

In this work, we explore a novel approach: The use of *human eye gaze as an implicit and dynamic source of information on user-perceived privacy*. In particular, our approach does not require a user to explicitly provide input or make active decisions, e,g, clicking a button or selecting a privacy setting. Naturally, the feedback derived from gaze is responsive to changing contexts, such as different content, tasks, or user states, i.e. as the user interacts with new stimuli, their gaze behaviour adapts, and our approach therefore continuously updates itself accordingly. Prior research has shown that eye gaze contains rich information about the user, such as identifiers [17], quasi-identifiers [113], confidential attributes (e.g. user activities [117], attentive [40, 120] and cognitive states [15, 58], or information about private situations [118]). Consequently, we try to answer our main research question RQ1:

RQ1: **Can human eye gaze be used as an implicit and dynamic indicator of user-perceived privacy?**

To this end, we present Gaze3P (**Gaze**-based **P**rediction of user-**P**erceived **P**rivacy) – the first large-scale dataset for studying user-perceived privacy from the perspective of human eye gaze. Gaze3P includes gaze data (i.e. where, when, and how a person looks) of 100 participants viewing 1,000 natural images showing different objects, some with private attributes (e.g. credit cards and medical history) as shown in Fig. 2. The dataset also provides user ratings of perceived privacy on a scale from 1 (very private) to 7 (very safe) for each image. The full dataset, including all annotations, is available at [Dataset Link] (cf. [8] for the implementation). Using our new dataset, we then explore:

RQ2: **How accurately can ML models predict privacy perceptions solely from gaze?**

We therefore present different tasks, focusing on the automatic prediction of users' perceived privacy solely from gaze behaviour. These tasks correspond to privacy-related problems or objectives that an algorithm is trained to address using our collected data. Once trained, the algorithm can be used to generalise its learned solution to apply to the same class of problems for previously unseen individuals. Since it is not feasible to a priori determine which features can be reliably extracted from gaze data—nor whether potentially confounding factors can be effectively disentangled—we employed machine learning algorithms to automatically identify privacy-related patterns associated with each task. These tasks include: **Stimuli-based Perceived Privacy (SPP)** tasks to infer how private a stimulus (e.g. image) is and **User-based Perceived Privacy (UPP)** tasks to infer information about the user (e.g. privacy expertise or identity). Our ML models demonstrate that human eye gaze provides accurate predictions of perceived privacy.

Hence, we explore a third research question:

RQ3: **Can gaze-based predictions of perceived privacy be integrated into privacy-preserving frameworks (such as Differential Privacy) to optimize utility while aligning with user expectations?**

We use the Gaze3P predictions of user-perceived privacy to optimise the parameters of differentially private mechanisms. Differentially private mechanisms obfuscate sensitive data samples such that only a limited amount of information about the private data can still be deduced from the obfuscated output of the mechanism. The exact amount of acceptable leakage depends on a privacy budget parameter $\varepsilon > 0$, which determines the obfuscating noise added by the mechanism. If ε is small, the privacy guarantee becomes stronger, but usually, the output of the mechanism is less accurate, and usability decreases. It is therefore important not to choose the privacy budget ε too small, i.e., to only add the minimal amount of noise that guarantees a target privacy level. The optimisation of DP-parameters has therefore seen much attention in recent years [11, 61, 98].

Our new dataset Gaze3P and the resulting ML model predictions provide a new way to determine ε, which reflects a user's perceived privacy. Depending on the actual use case, we propose different mappings from perceived privacy levels to ε-values. We evaluate how each mapping affects the utility of the obfuscated output dataset and show that our gaze-based approach outperforms previous work.

Contributions. In summary, our work makes the following contributions:

(1) We present Gaze3P – the first large-scale dataset for studying user-perceived privacy using human eye gaze, providing a dynamic and implicit user feedback.
(2) We propose several novel learning tasks focusing on predicting user-perceived privacy from human eye gaze. These tasks cover different aspects of privacy and also allow us to explore potential applications and limitations of gaze-based privacy perception.
(3) We demonstrate how gaze-based predictions can be used to optimise parameters of privacy-preserving techniques. Specifically, we introduce a novel approach that maps predicted privacy levels to DP's privacy parameter ε and show that aligning DP with user expectations improves the data utility in data analysis and learning.

2 Preliminaries

Eye Tracking. Gaze data is typically collected using eye-tracking devices that record the position and movement of a user's eyes relative to a visual stimulus or screen. Modern eye trackers employ infrared light to detect corneal reflection and pupil centre, enabling accurate estimation of gaze coordinates at high temporal resolutions. The raw gaze signal is then processed into interpretable features such as fixations, saccades, and pupil dilation:

- *Fixations* refer to time periods where the eye remains focused on a specific location, typically lasting 100–400 ms. They are indicative of visual attention and cognitive processing of that region.
- *Saccades* are rapid eye movements between fixations used to reposition the fovea to new visual targets, lasting 20–80 ms. These movements are ballistic, and their patterns can inform about scanning behaviour and search strategies.
- *Pupil dilation* is a physiological response modulated, amongst others, by both environmental lighting and cognitive load. Increased dilation was linked to heightened mental effort, emotional arousal, or attentional demand.

Together, these gaze features provide a rich, temporally fine-grained source of implicit user feedback (i.e. without requiring direct input or explicit interaction). We refer the reader to [53, 54, 95, 100] for details about eye tracking and gaze behaviour analysis.

By transforming the continuous, high-dimensional gaze signals into quantitative features—such as fixations, saccades, and pupil diameters—machine learning (ML) models are provided with informative inputs that capture essential characteristics of user interaction or cognitive states. These features are then passed into ML algorithms, enabling the models to learn underlying statistical patterns or associations within the data. This learned structure allows the models to perform specific tasks such as classification, regression, or clustering. Ultimately, this feature-to-model pipeline allows ML systems to generalize from training data and make accurate, data-driven inferences about new, unseen inputs.

Differential Privacy (DP). DP is a mathematical framework that ensures privacy by limiting the impact of any single data point on the output of a computation. A randomised algorithm M satisfies ε-DP if, for all datasets D and D' differing by at most one element, and for all measurable subsets S of the output space:

$$\Pr[M(D) \in S] \leq e^{\varepsilon} \cdot \Pr[M(D') \in S], \tag{1}$$

where $\varepsilon \geq 0$ is the privacy budget, controlling the privacy loss. A small ε means that D and D' are (almost) not distinguishable given a set of outputs S. From an adversarial perspective, an adversary $\mathcal{A}$ challenged to distinguish D and D' given an output set S will output the dataset which is more likely, e.g. D if $\Pr(D|M(D) \in S) \geq \frac{1}{2} \geq \Pr(D|M(D') \in S)$. In the most extreme case of Eq. (1) we have $\Pr(M(D) \in S) = e^{\varepsilon} \Pr(M(D') \in S)$ and hence $\Pr(D|M(D) \in S) = e^{\varepsilon} \Pr(D'|M(D) \in S) = e^{\varepsilon}(1 - \Pr(D|M(D) \in S)) \Rightarrow \Pr(D|M(D) \in S) = \frac{e^{\varepsilon}}{1+e^{\varepsilon}}$. Thus, the (absolute) advantage of an adversary is bounded by $\text{adv}_{\mathcal{A}} \leq 2\frac{e^{\varepsilon}}{1+e^{\varepsilon}} - 1 = \frac{e^{\varepsilon}-1}{e^{\varepsilon}+1}$ (cf. Section E for more details).

3 Related Work

User Perceived Privacy. As more data is being collected, shared, and processed, sensitive insights about the user's personality, intentions, and preferences are being leaked [63, 89]. Hence, to better protect user privacy, prior works have investigated the psychological mechanisms of privacy decision-making [89], self-disclosure [28], and the related cost-benefit analysis [21]. They showed that the user-perceived privacy dynamically changes according to the user's context, behaviour, and knowledge. Other works focused on privacy-in-context (i.e. contextual integrity) [97, 123] and further showed that user-perceived privacy is affected by culture, activities (e.g., online shopping vs. online banking), and platforms (desktop vs. mobile). Prior works [63, 79, 118] often used generic privacy mechanisms that remain static throughout the interaction. These mechanisms are typically predefined at the onset of a session (e.g., at the initiation of a protocol) and fail to account for the dynamic nature and context-dependent fluctuations of the users' privacy judgments [55].

Eye Gaze and Privacy Perception. The results of the aforementioned works usually rely on user questionnaires. However, when using questionnaires, users often fail to follow their own privacy preferences [93, 101, 107]. In this paper, we propose to use human eye gaze (instead of questionnaires) to implicitly capture the dynamics of user-specific privacy perception. Eye-tracking data is already widely used to study human behaviour and cognition [18, 35, 53, 54, 95, 100]. Prior research has demonstrated that gaze patterns can reflect cognitive processes such as risk perception and habituation [5, 48, 49]. For instance, studies have shown that increased risk is often associated with longer fixations and heightened visual scanning, indicating deeper cognitive engagement [52, 92, 119]. Conversely, habituation to repeated stimuli can lead to reduced gaze variability and decreased attention, even in the presence of sensitive information—a challenge for maintaining consistent privacy awareness [5]. Gaze3P builds on these insights and hypothesises that gaze can also be used as an indicator of user-perceived privacy. The idea to use gaze to detect privacy-sensitive situations is not completely new and has been explored in [118], where the users' eye movements and first-person video were recorded using an egocentric (head-mounted) camera. However, [118] focuses on detecting privacy-sensitive situations rather than quantifying privacy perception. It also only features a small set of 17 participants in a free-viewing task and relies on recording and processing the scene imagery, which might break privacy [101]. In this paper, we instead focus on the implicit and dynamic privacy perception feedback solely through gaze.

Other Privacy-Related Gaze Applications. Prior works at the intersection of eye tracking and privacy mainly focus on (i) eye-based authentication [78, 84], (ii) privacy considerations and guidelines [47, 64], (iii) secure AR/VR applications [13, 25–27, 75, 84], (iv) UI design for secure interactions [64], (v) secure gaze data sharing [12, 25–27, 37–39], and (vii) information leakage and attacks on gaze data [19, 38, 116, 122]. Apart from the different focus, Gaze3P, also differs in (i) the purpose of gaze, i.e., some prior work [75, 84, 122] uses gaze for explicit interaction (e.g., authentication [84] or gaze rays in VR [75]) while Gaze3P uses gaze as an implicit (passive) signal to infer perceived privacy, (ii) unlike prior work, we focus on the cognitive aspect of gaze in privacy, beyond the gaze location estimation [19, 122], (iii) privacy positioning, i.e. Gaze3P is user-centric and proactive (helping users protect their own privacy through implicit gaze behaviour) while prior works either focus on system-enforced protection for eye tracking [25–27, 37–39, 75], user authentication [78, 84], or demonstrating privacy vulnerabilities [64, 116].

Personalised Differential Privacy. In classical differential privacy (DP), the privacy parameter ε is chosen independently of the subjective privacy perception of an individual user but instead provides the same privacy guarantees to all users [61]. This is despite the fact that users have varying expectations about acceptable privacy levels. As a result, a certain DP-privacy level ε might not offer enough protection for some users while overprotecting others [42, 61]. Personalised Differential Privacy (PDP) [2, 4, 11, 24, 34, 61, 98] is an extension of standard DP that introduces flexibility by tailoring privacy protections based on individual preferences, allowing for a more nuanced balance between privacy and utility. Alaggan et al. [4] first introduced the theoretical concept of PDP through linear pre-processing (a.k.a stretching) of the input data. In their approach, the input data is scaled according to each individual's privacy preference before applying a differentially private mechanism. Cummings and Durfee [23] generalised the PDP framework to a broader class of mechanisms. They proposed a constructive method for implementing personalised privacy guarantees directly within the mechanism design. However, they demonstrated that computing an optimal personalised mechanism under these conditions is NP-hard. Later works [22, 41] used weighted moment estimation of each data point according to the privacy level. In these approaches, the privacy level specified by each user was used to assign a weight to their corresponding data contribution during the statistical estimation process. This weighting strategy ensured that data from users requiring stronger privacy protections (i.e., lower ε values) exert less influence on the aggregate statistics, while users with looser privacy requirements (higher ε values) contribute more significantly. Other works [10, 76] proposed partitioning the data into separate groups and then assigning different privacy levels. Jorgensen et al. [61], followed by Niu et al. [98] and Ebadi et al. [34], relied on excluding or sub-sampling some data samples according to their privacy levels. While all of these methods were designed for data analysis, Boesnisch et al. designed a privacy-preserving training mechanism for machine learning models that integrated individual privacy levels directly into the optimisation process [10, 11]. They adapted the gradient computation and noise addition to reflect user-specific privacy budgets.

Similarly, we determine a suitable privacy budget ε based on behavioural data rather than assigning it arbitrarily or using assumptions. Our work is the first to do so using human gaze data. This approach ensures that the privacy budgets more accurately reflect users' actual perceptions and expectations of privacy. Grounding ε in observed user behaviour not only improves the practical relevance and usability of privacy-preserving systems but also helps bridge the gap between formal privacy theory and human-centred privacy concerns, ultimately leading to more trustworthy and adaptive data handling practices.

4 Gaze3P Dataset

Due to the lack of available datasets, in this section, we present our new gaze-based dataset Gaze3P. Our large-scale data collection is essential for deriving statistically significant insights into users' perceptions of privacy and for empirically validating the use of gaze as a reliable indicator of perceived privacy. By offering a standardized dataset, Gaze3P aims to facilitate reproducible research and enable the development and evaluation of models that infer perceived privacy dynamically and without explicit feedback through gaze patterns.

Figure 2: Sample images from the VISPR dataset with safe (e.g. cat, colours) and private (e.g. credit card, political opinion) attributes

Eye tracker. For gaze data collection, we used an Eyelink eye tracker that provides binocular gaze data at a sampling rate of 2 kHz. As is common practice in laboratory eye tracking studies, we used a chin rest to stabilise participants' heads. Images were shown on a computer screen with a resolution of 1920x1080 pixels and a size of 545 mm x 303 mm. The eye-to-screen distance was 700 mm. The proportion of the calibrated area was set to 0.63 x 0.88 to stay within the trackable range of the system with an HV13-13 (horizontal/vertical 13 targets) calibration type for better spatial accuracy across the entire screen. The recorded gaze data was then processed into fixations, saccades, and pupil information (cf. Section 2). More details and visualisations can be found in Section A.

Stimuli. We randomly sampled a subset of images from the VISPR dataset [101] as our stimuli since it is the only publicly-available dataset that contains privacy-related attributes. See Fig. 2 for sample images. The dataset contains 68 attributes categorised into nine attribute groups. The attributes were compiled according to the guidelines for the EU Data Protection Directive (GDPR) 95/46/EC, the US Privacy Act of 1974, and the data sharing rules in online social networks. It further includes reliable and consistent attribute annotations by letting multiple annotators follow detailed labelling guidelines. We ensured a balanced distribution across the annotated privacy attributes. This stratified sampling approach was employed to maximise attribute coverage and mitigate sampling bias, given the practical constraint on the number of stimuli each participant could reasonably view during the experiment. Despite this limitation, the resulting dataset remains relatively large and representative, supporting robust analysis of visual privacy perception across a diverse range of attributes. Our dataset includes 1,000 images with corresponding private and safe (i.e. non-private) attributes. The attribute categories in VISPR [101] are:

- **Personal information:** e.g. age, gender, fingerprint, signature.
- **Documents:** e.g. credit card, passport, national ID.
- **Medical:** e.g. medical history, hospital tickets, physical disability.
- **Employment:** e.g. occupation, work occasion.
- **Life:** e.g. culture, religion, political opinion, sexual orientation.
- **Relationship:** e.g. personal, social, professional.
- **Whereabouts:** e.g. landmark, home address.
- **Online activity:** e.g. date/time of activity, username, password.
- **Automobile:** e.g. license plate, vehicle ownership.

Participants. We initially recruited 103 participants through university mailing lists and notice boards. We had to exclude three

participants due to calibration failures [71]. This resulted in a final group of 100 participants (37 females and 63 males). Participants were aged between 18 and 35 years, had different nationalities (25) and different (self-reported) privacy knowledge (12 experts)[1]. All participants had normal or corrected-to-normal vision. We refer to Section A for more demographic details.

Experiment design. After arriving in the lab, participants were first informed about the general purpose and procedures of the study. Following [114], we explicitly asked participants to consider the data as their own, e.g. their phone gallery. The experiment consisted of four blocks with 25 participants each. Each block included two tasks: free-viewing and a search task. In the former, participants were shown a stimulus (e.g. an image of a credit card) for five seconds and asked to rate its privacy-sensitivity for sharing on a scale from 1 (very private) to 7 (very safe) following [118]. Similarly, for the search task, participants were asked to search for a specific image (by a category, e.g. *document*) in a 2 x 2 image collage, and click the mouse and rate its sensitivity once found, following [114]. The same attributes were presented in both tasks, but each task featured different sets of images within those attributes. Each task included three practice trials and 50 recorded trials with randomized order of (non)private stimuli (50 %-50 % ratio). The resulting dataset includes a set of triplets of {stimulus, gaze patterns (timestamped coordinates), and privacy rating} for each participant in each task. Refer to Section A for more details on the dataset structure and data collection software.

Compliance with the privacy and ethics guidelines. The data was collected and processed according to the standards, guidelines, and approval of the ethical committee of the authors' institution, with participants' consent, remuneration, and pseudo-anonymisation procedures. In particular, the privacy and ethical guidelines of the Menlo Report [29] were satisfied.

5 Gaze3P Tasks

Our Gaze3P dataset enables new analyses that shed light on how users cognitively and behaviourally respond to privacy-relevant stimuli. Given the difficulty in determining beforehand which features can be meaningfully extracted from gaze data and whether confounding variables can be effectively separated, we adopted a machine learning-based approach to automatically uncover privacy-related patterns corresponding to each task. This method enables us to objectively evaluate our hypothesis: that gaze behaviour may function as a proxy for perceived privacy.

More specifically, in this section, we explore how well our gaze-based dataset Gaze3P is suited for two groups of learning tasks: (i) *stimuli-based perceived privacy* (SPP) tasks that infer how private a stimulus (e.g. image) is in Section 5.1 and (ii) *user-based perceived privacy (UPP)* tasks that infer information about the user (e.g. privacy expertise or identity) in Section 5.2. Each task entails learning a distinct mapping or pattern within the gaze data, such as predicting privacy ratings from gaze, classifying stimuli as private or non-private, or identifying user-specific privacy preferences based on demographic or behavioural features. For implementation, we used a basic Scikit-learn framework with default parameter settings. This provided a standardised and unbiased baseline for model training, avoiding manual feature selection or tuning that could skew results or introduce overfitting. It is important to note that all models in this study are trained on annotated gaze data, where user-specified privacy ratings serve as ground-truth labels to facilitate supervised learning and enable robust evaluation of model performance. The purpose of this training procedure is to allow the model to learn association patterns for each task. During testing and deployment, however, only the raw gaze data is provided as input to the model—without any accompanying user-specified ratings. This setup reflects a practical application scenario, wherein the model is expected to generalise to previously unseen users or stimuli and autonomously infer privacy-related judgments based solely on gaze behaviour. This not only facilitates the development of scalable and user-adaptive privacy-aware systems but also provides empirical insights into the underlying mechanisms of perceived privacy.

In the following, we describe the two training tasks in detail and evaluate the performance of our trained models w.r.t standard baselines, namely decision tree (DT), support vector machine (SVM), logistic regression (LR), random forest (RF), K-nearest neighbour (KNN), and transformer (TF) models (cf. Table 1). To study individual stimuli, we focused on the free-viewing data of our dataset.

5.1 Stimuli-based Perceived Privacy (SPP)

Quantifying perceived privacy levels helps understand how users feel about their privacy protections. Privacy perception is mainly influenced by the nature of the stimuli, e.g. their type (e.g. images) and the content being shared or observed (e.g. a credit card). A quantitative relation between the stimuli and the corresponding privacy perception can help in designing more context-aware and effective privacy-preserving mechanisms. We, therefore, propose four main SPP tasks:

(1) Binary Privacy Perception. Given solely the gaze data, the binary privacy perception task [118] aims to determine whether a user is exposed to or interacts with potentially sensitive information. In addition to the pioneering work of Steil et al. [118], we also want to determine how the different setups affect the binary perception, e.g. how the perception of a specific user (intra-user setting) varies for different stimuli in comparison to how the perception of many users (inter-user setting) varies.

(2) Privacy Level Perception. This task aims to map the gaze data as inputs to a privacy level as output. The dataset includes a ground truth of 7 different privacy levels, following Steil et al. [118], indicated by participants for each stimulus. Note that, unlike Steil et al. [118], we process all classes instead of combining them into two.

(3) Contextual Privacy Perception. Integrating contextual information such as demographics or user expertise can potentially improve predictions of user-perceived privacy levels[2] Contextual information is provided as additional features to the model, capturing the user's age, gender, nationality, and privacy expertise.

[1] All experts are MSc or PhD holders of information security degrees.

[2] Other contextual information like the type of application, time of the day, and the type of platform, has been shown to influence the user privacy perception too (cf. Section 3). To simplify the setup, we did not include this information in our dataset; however, our approach naturally extends to more detailed datasets.

(4) Private Attribute Recognition. Prior works [17, 113] showed that gaze is a good predictor of the user's private attributes, such as age and gender. Other works [101, 118] recognise the private attribute directly from images as inputs to their models. Here, we focus on predicting what private attribute the user is looking at, given the gaze data alone. The task becomes more challenging as semantics become more complex and diverse.

SPP applications. Before we discuss our results on tasks (1)–(4), we want to outline how stimuli-based perceived privacy results can be used in applications. For example, the quantified values in Tasks 1-3 (classified per attribute in Task 4, if needed) can be mapped to the corresponding parameters in privacy-preserving protocols. Such parameters can include ε values in DP [31, 32], model update perturbation or gradient clipping thresholds in federated learning [63, 88], and similar hand-picked parameters in K-anonymity [112], L-diversity [85], T-closeness [77], privacy auctions [44, 124], synthetic data generation [80], etc. Other stimuli-specific insights are important for multiple applications, such as access control models (to define who can access which piece of information, e.g. attribute-based access control (ABAC) [57] which uses attributes for dynamic access control and human-in-the-loop privacy controls (e.g. Instagram's 'Restrict' feature [104] for controlling interactions, e-mail spam filters that allow manual corrections [6], and marketing campaigns that respect the user's perceived privacy preferences [86]). We refer to Section 6 for sample applications and to Section C for further details.

SPP Baselines and Evaluation. As baselines, we ran the aforementioned basic models. We evaluate the accuracy of the models with cross-validation and test sets to ensure that the model generalises well to unseen data. Implementation details can be found in Section B. We also perform statistical hypothesis testing on the extracted eye-tracking features to examine whether the observed differences across experimental conditions or participant groups are statistically significant. These tests assess whether variations in metrics such as fixation duration, saccade amplitude, or pupil dilation are likely to reflect systematic effects rather than random noise or individual variability. To quantify the strength of evidence against the null hypothesis (i.e., that there is no meaningful difference between groups or conditions), we compute p-values. These values represent the probability of observing the given data, or something more extreme, under the assumption that the null hypothesis is true. A p-value below a conventional threshold (typically 0.05) is considered indicative of a statistically significant effect. Hence, we made the following key observations in relation to SPP Tasks 1-4:

Inter- and intra-user variations. Table 1shows the accuracy of the aforementioned models if all (meta)data is passed to the ML model. Our results demonstrate that gaze data can quantify perceived privacy since results exceed chance levels (0.5 for Task 1 and 0.14 for Tasks 2 and 3). We further analyse this in both person-specific and person-independent settings. In each case, models were trained on 80% of the data and tested on the remaining 20% with cross-validation. In the person-specific setting, models are tested on 20% of individual participants, accounting for personal gaze behaviours (with only one sample at a time independent of any user history), whereas in the person-independent approach, a generalizable model

Table 1: Accuracy of the SPP tasks Section 5.1.

	Binary Privacy Perception	Privacy Level Perception	Contextual Privacy Perception
Person-independent			
DT	0.54	0.19	0.23
SVM	0.64	0.34	0.37
LR	0.63	0.34	0.36
RF	0.60	0.29	0.32
KNN	0.56	0.21	0.24
TF	0.52	0.34	0.37
Person-specific			
DT	0.61	0.23	0.24
SVM	0.70	0.38	0.39
LR	0.75	0.40	0.40
RF	0.66	0.35	0.37
KNN	0.62	0.30	0.31
TF	0.81	0.52	0.54
Down-sampling			
DT	0.48	0.15	0.16
SVM	**0.57**	0.30	0.33
LR	0.55	0.28	0.29
RF	0.52	0.24	0.26
KNN	0.45	0.20	0.20
TF	0.46	0.30	0.31
Gaze + Stimuli (ours)			
DT	0.89	0.73	0.72
SVM	0.93	0.81	0.83
LR	0.92	0.80	0.81
RF	0.90	0.82	0.80
KNN	0.88	0.79	0.82
TF	0.97	0.87	0.90
Gaze + Stimuli (PrivacEye)			
DT	0.58	-	-
SVM	0.67	-	-
LR	0.66	-	-
RF	0.70	-	-
KNN	0.58	-	-
TF	0.75	-	-
Gaze Saliency Maps			
DT	0.45	0.17	0.21
SVM	0.52	0.30	0.31
LR	0.57	0.29	0.30
RF	0.52	0.23	0.27
KNN	0.50	0.18	0.18
TF	0.46	0.29	0.32
Residual Gaze Maps			
DT	0.23	0.11	0.13
SVM	0.21	0.15	0.22
LR	0.19	0.17	0.08
RF	0.27	0.19	0.21
KNN	0.09	0.10	0.23
TF	0.31	0.18	0.19

Figure 3: Qualitative example of the inter- and intra-person differences: Each row corresponds to a different participant, where red dots depict the areas that participants attended to the most (i.e. more fixations). The user-selected privacy levels l are consistent for each participant but different across participants. For example, the *political opinion* attribute is private (l between 1 and 2 - the private levels) for one user -the first row- and not private for the second (l between 4 and 6 - the safe levels). The gaze fixations are also denser on the private regions of interest (e.g. faces and politicians) for the *private* levels.

is trained across all participants and tested on 20% of each participant's data. Results in Table 1 and Fig. 3 indicate significant variations in inter-user differences (person-independent, leading to lower accuracy), whereas intra-user features exhibit greater consistency (person-specific, leading to higher accuracy). This suggests that each user possesses a distinct and individualised perception of privacy.

Fixation-based attention allocation. In general, fixation duration increases on private cues with fewer fixations on less private regions (H_0: *'There is no difference in the distribution of fixation durations across different stimuli regions'* tested with Kruskal-Wallis H-test and Dunn's test with Bonferroni correction for multiple comparisons with the different privacy levels, with $p\text{-}value = 0.04 < 0.05$ and Epsilon-squared effect size for Kruskal-Wallis H-test of 0.05). This can also be seen in Fig. 3. It further suggests that less fine-grained tracking is sufficient (i.e. relying on fixations alone). This was further supported by downsampling the gaze data to 30 Hz (30 FPS) to simulate commodity-level standard webcams, as shown in Table 1-downsampling. The results, again, suggest that high-resolution gaze data (Table 1-person-independent) may not be necessary for effective privacy perception prediction, facilitating a more widespread implementation and enhancing accessibility and usability for a broader audience [3].

Social influence: Users' privacy perception may be guided by social influence (e.g. demographics), as shown in the *Contextual Privacy Perception* results and Fig. 4. Our results indicate that demographics play a significant role in shaping outcomes (i.e. increasing accuracy).

[3]As this study represents exploratory research, we employed a high-resolution, constrained experimental setup. This choice was motivated by the initial uncertainty regarding whether a lower-resolution configuration would yield significant results and to what extent fine-grained details in the data would be necessary to capture relevant effects. The high-resolution setup ensured maximal data fidelity, allowing for comprehensive observation of potentially subtle phenomena during the early stages of research.

Figure 4: Qualitative example of social influence: By looking at the attribute *religion*, participants assign *private* ratings to stimuli that are more closely associated with or frequently encountered within their social background, probably due to their personal significance (c.f. Section 3).

Figure 5: Qualitative example of learning effect: Examples belong to the same participants on images of two different attributes *driver license* and *political opinion*, sampled from the first and last 5 stimuli. When a specific attribute is presented (e.g. in the first 5 stimuli) and repeated multiple times, it increases the participant's familiarity with the attribute through repeated exposure. Hence, the number of fixations, response time, and privacy-sensitivity decrease.

This suggests that demographic factors and social background contribute to identifiable patterns that machine learning models can detect. Consequently, these factors enhance the predictive power of the models, highlighting their relevance in understanding variations in the data.[4]

Long-term learning and adaptive privacy behaviour. Repeated exposure to private attributes changes user behaviour over time. As shown in Fig. 5, while gaze fixations decrease on repeated attributes over time (i.e. desensitisation to privacy risks), the privacy ratings increase (i.e. learning effect and growing privacy awareness), and response time decreases.

[4]It is important to note that our unbalanced demographics—an inherent consequence of the open and uncontrolled call for participation—limit our ability to make broad or generalizable claims about specific demographic groups or extrapolating results to a wider population. As a result, while our analysis may reveal meaningful trends, caution is required when interpreting demographic-specific conclusions using our dataset

Table 2: Evaluation of the private attribute recognition task as class-based mean average precision (C-MAP) (the average of the per-attribute average of the area under the Precision-Recall curve on all attributes), following [101]'s models (CaffeNet, GoogleNet, and ResNet) and more recent SOTA models (Multimodal fusion model: GazeFormer [91], Cross-modal transformer: ViLT [66], and self-supervised contrastive model: CLIP [109])

Model	Feature	VISPR [101] (stimuli)	Ours (gaze)	Ours (gaze + stimuli)
SVM	CaffeNet	41.34	29.80	58.22
SVM	GoogleNet	43.77	30.07	60.43
SVM	Resnet-50	44.21	32.45	62.37
E2E	CaffeNet	47.56	35.98	65.85
E2E	GoogleNet	47.72	34.08	67.00
E2E	Resnet-50	56.13	35.32	69.13
SOTA	GazeFormer	64.13	40.82	92.73
SOTA	ViLT	66.18	41.59	89.47
SOTA	CLIP	61.00	35.06	78.56

Figure 6: Qualitative example for visual privacy: Example stimuli corresponding to the gaze data that were misclassified

Visual Privacy. For private attribute recognition, as shown in Table 2, we compared our findings with the VISPR [101] evaluation as well as other SOTA models. While [101] only input the stimuli, we replicated their models and metrics on our dataset and inputted the gaze data alone vs the gaze and stimuli. Nonetheless, we observed that, as shown in Fig. 6 (left to right): (i) when using gaze as the only input, the models are able to identify objects that occupy a significant proportion of the image space while failing at identifying other subordinate attributes such as age or gender; (ii) when augmenting the stimulus with the gaze data, the models are able to identify more fine-grained details such as wedding rings and tattoos that the stimuli-alone version fails at (as long as the participants paid attention to such details), (iii) in all versions, models fail at identifying the relationship-based attributes (as also reported by [101]) since they require some reasoning capabilities that our basic models do not achieve.

Gaze as a supplementary input. We assess model performance with the inclusion of stimulus data and compare it to PrivacEye [118]. Results in Table 1-gaze+stimuli show that the stimulus features significantly improve the performance. Nonetheless, models trained solely on gaze data still yield meaningful results. This demonstrates that gaze behaviour alone carries informative signals related to perceived privacy, supporting its utility in scenarios where access to full visual content may be restricted or deliberately excluded for data minimization and privacy-preserving purposes.

Bottom-up and top-down attention. To minimise the confounding influence of bottom-up (saliency-driven) gaze, we attempted to control the visual composition of our stimuli by selecting the VISPR images that mostly contain a single dominant attribute. This design choice was intended to reduce the likelihood that gaze behaviour would be driven by reflexive visual attention to salient features, such as bright colours, known to automatically attract attention regardless of the observer's cognitive or emotional state. Such bottom-up attention is inherently stimulus-driven and unrelated to higher-order constructs like perceived privacy (i.e. top-down attention). However, this assumption may not hold in all scenarios, especially when stimuli are real-world images. To address this, we check if residual gaze patterns — i.e., the part of gaze behaviour not explained by bottom-up saliency — are more predictive of perceived privacy than raw gaze data. We use a saliency prediction model (DeepGaze [72]) to generate a saliency map for each stimulus image. This represents the expected gaze distribution if attention were purely bottom-up. Then we compute the residual gaze maps by subtracting the (normalised) saliency map from the original gaze map. We input both the original gaze maps as well as the residual gaze maps to the models, as shown in Table 1. Results are significantly lower for residual gaze maps, which may indicate that privacy-relevant content often coincides with visually salient regions. Additionally, the lack of variance in saliency features reduces the need for correction, leading to differences in performance between the two models, since the stimuli used in the study were relatively homogeneous in terms of saliency, mostly with one dominant object.

5.2 User-based Perceived Privacy (UPP)

Even for the same stimuli, different users develop different privacy perceptions. Hence, perceived privacy is subjective and varies substantially among participants. Understanding the user-specific features of privacy allows researchers and policymakers to design more effective, personalised, and user-centric privacy solutions. We, therefore, propose three UPP-tasks:

(1) Privacy Expertise Prediction. Prior works [14, 73, 81] showed that distinctive eye movement behaviours and gaze strategies correlate with domain-specific expertise. This task aims to develop a model that can distinguish between privacy experts and non-experts based on their gaze patterns during interactions with digital content or privacy-related tasks. The gaze data was collected from participants with varying levels of privacy expertise, potentially revealing how different levels of knowledge influence visual attention and decision-making in privacy-sensitive contexts. We address this task, therefore, by training a classifier to distinguish between the two groups based on the gaze features.

(2) User Privacy Profiling. The privacy expertise prediction task can be further extended to capture the gaze behaviour profile, e.g. a summary of the key gaze features that characterise each group [101], such as differences in attention to specific elements or gaze strategies. This is commonly used to adapt privacy preferences

according to different cohorts [45]. Hence, the task aims to cluster the different groups according to the gaze patterns.

(3) Privacy-aware Gaze Identification. Current gaze-based user identification methods often rely on specially designed visual stimuli or induced artificial gaze patterns. Prior works [1, 30] have investigated the feasibility of distinguishing individuals based on their natural gaze behaviour while freely viewing images and suggested that viewing different images, such as a personal vacation photo, elicits distinct emotional responses, which are reflected in gaze patterns and are unique to each individual. We extend this idea to privacy-aware gaze behaviour, proposing that individuals exhibit unique gaze patterns not only in response to personal relevance but also when assessing perceived privacy. Privacy perception is inherently subjective, shaped by personal experiences, cultural influences, and cognitive biases. By analysing how users visually explore and evaluate images with different privacy implications, we need to demonstrate in this task that privacy perception itself can serve as an implicit user identifier [33].

UPP applications. User-specific insights can, for example, be used for privacy nudges [59] (to encourage users to make privacy-conscious decisions according to their expertise or profiles, commonly used in social media platforms to prevent oversharing, e.g. Facebook's contextual privacy warnings [90] and Chrome browser-based security warnings [46]), privacy labels and transparency notices (to help users understand and control their privacy choices according to their profiles, e.g. Apple's App Store privacy labels [7]), and cohort-based recommendations (for group-based personalization, e.g. Google's Federated Learning of Cohorts (FLoC) [45] for ad targeting) (C.f. Section C).

UPP Baselines and Evaluation. To predict the privacy expertise of users and groups, we ran the same evaluation as in Section 5.1. The results are shown in Table 3, where results exceed the chance levels (0.5 and 0.01 accuracy for privacy expertise prediction and privacy-aware gaze identification, respectively). We observe the following key results:

Cognitive and Perceptual Adaptation. The difference in gaze behaviour between experts and non-experts arises due to cognitive and perceptual adaptations that develop with experience and training [14, 73, 81]. Our results show that privacy experts exhibit more focused attention on privacy-relevant information and quicker identification of potential privacy risks compared to non-experts. In other words, privacy experts have rapid gaze shifts, suggesting automatic heuristic-based decision-making, while non-experts have longer dwell time, indicating uncertainty and high cognitive load in privacy assessment (H_0: *'The gaze duration distributions of the two groups (experts and non-experts) are equal'* tested by a Mann-Whitney U-test with $p\text{-}value = 0.03 < 0.05$ and effect size $r = 0.3$), leading to distinguishing features that ML models can identify with high accuracy (c.f. Table 3).

Distinct Profiles. We further evaluated the participants' profiles per attribute and found out that, following [101], K-means clustering yields the lowest silhouette score with 12 distinct gaze behaviour profiles (as opposed to 30 profiles when clustering images [101]).

Table 3: Accuracy of the UPP tasks

	Privacy Expertise Prediction	Privacy-aware Gaze Identification
DT	0.78	0.05
SVM	0.45	0.04
LR	0.46	0.02
RF	0.87	0.04
KNN	0.85	0.03

Genuine significance. Previous studies on gaze behaviour indicate that viewing different images, such as personal photos, elicits distinct emotional responses that are reflected in gaze patterns and are unique to each individual [1, 30]. In our experimental setup, we tried to simulate this setup by instructing participants to conceptualise the stimuli as their own phone gallery without incorporating actual personal photos. Unfortunately, the previously observed personalised effect could not be replicated in our study and setup (cf. Table 3). We believe that this is due to the lack of genuine emotional attachment and/or lack of personal significance and familiarity.

6 Privacy-Preserving Applications

Our results so far showed how gaze can be used to predict the privacy perception of an individual or a group of users. There are many potential applications for this approach (we already hinted at some of them). Here, we concentrate on one important task, and study this in more detail, namely, finding good and tailored epsilon values for DP.

6.1 Privacy-Utility-Cognition Trade-Off

To use our results from Section 5 with a DP-mechanism, we first need to map user-perceived privacy levels $l \in \{1, \dots, L\}$ with a function f to ε-values for Differential Privacy.[5]

Depending on the cognitive aspects behind the user-perceived privacy preference, which we discussed already in Section 3, we want to connect l with different privacy loss functions $g(l)$, e.g. such that the privacy loss depends linearly on l as in [69, 83].

To construct a suitable mapping f to DP-privacy levels ε, recall from Section 2 that a privacy-budget ε ensures that an adversary $\mathcal{A}$ can distinguish two data sets with advantage bounded by $\mathrm{adv}_{\mathcal{A}}(\varepsilon) \coloneqq \frac{e^{\varepsilon}-1}{e^{\varepsilon}+1}$. Hence, if we require the privacy loss (in terms of the success of a potential adversary) to have a certain behaviour g, we need $g(l) = \mathrm{adv}_{\mathcal{A}}(f(l))$. Observe that we can compute $f(l) = \log\left(\frac{1+g(l)}{1-g(l)}\right)$ explicitly. In particular, we find a mapping function f for every positive loss function g with values smaller than 1.

Since users do not select privacy levels purely based on objective risks or technical parameters, but rather through subjective interpretations shaped by cognitive biases (cf. Section D), we now want to discuss the typical choices of g in cognitive science and how they affect the accuracy of a model obfuscated with ε-DP noise if $\varepsilon = f(l)$.

[5] In our evaluation, we use $L = 7$ (cf. Section 3).

Note that for real-world applications, we can usually restrict the privacy budget to $\varepsilon_{min} \leq \varepsilon \leq \varepsilon_{max}$ where the minimum $\varepsilon_{min} \geq 0$ and maximum $\varepsilon_{max} < \infty$ are application-specific [94]. We therefore, use functions g depending on $\varepsilon_{min}, \varepsilon_{max}$.

It is important to note that users employ diverse rationales and cognitive strategies when forming their privacy perceptions, and these strategies are shaped by multiple interacting factors, such as the stimuli, background knowledge, and demographics (cf. Section 5). As a result, there is no universally correct mapping function, as these differences reflect valid subjective interpretations rooted in cognitive and contextual variability. Rather than assuming a single ground truth, such mappings should be empirically learned from user behaviour and validated through data-driven analysis. Hence, later in this section, we illustrate this point by providing a practical example based on our dataset.

6.1.1 Linear Mapping. A straightforward way to interpret perceived privacy is a function g linear in l [69, 83]. Namely, we choose for $l \in \{1, \ldots, L\}$:

$$g(l) = \mathrm{adv}_{\mathcal{A}}(\varepsilon_{min}) + \frac{l-1}{L-1}(\mathrm{adv}_{\mathcal{A}}(\varepsilon_{max}) - \mathrm{adv}_{\mathcal{A}}(\varepsilon_{min}))$$

This mapping assumes equal intervals between privacy levels. Linear mappings are widely used in statistics, cognitive sciences, and machine learning for their mathematical simplicity, speed, and interoperability, and serve as good approximation functions. Nonetheless, they may potentially lead to inaccurate interpretations [16]. This is because, although privacy levels are numerically treated as equidistant, the psychological or cognitive perception of the distance between those points may not be equal [65, 121]. For example, the difference between 'very safe' and 'safe' may be much smaller in the user's mind than the difference between 'neutral' and 'moderately private'. Similarly, the midpoint ('neutral') may be seen not as a numerical centre but rather as a safe choice. Hence, we no longer have a linear relation.

6.1.2 Exponential Mapping. Alternatively, an exponential function could be used when users interpret privacy levels as a continuum, with thresholds separating the different privacy levels. In other words, in such cases, small variations in the privacy parameter may be perceived as insignificant at higher levels (e.g. 'very safe' and 'moderately safe'), while similar changes near specific lower thresholds elicit disproportionately larger responses (e.g.'neutral' and 'moderately private'). These thresholds may vary due to the user's cognitive processes, where users with less familiarity with privacy concepts may rely on intuitive or categorical decision-making or cases where specific features of the images (e.g., presence of identifiable faces or objects) might consistently evoke higher or lower ratings. An exponential mapping addresses this pattern and emphasises stronger privacy guarantees at lower ratings and captures a cognitive tendency where users may undervalue small privacy differences in higher ratings, improving utility. We get:

$$g(l) = \mathrm{adv}_{\mathcal{A}}(\varepsilon_{min})^{(L-l)/(L-1)}\, \mathrm{adv}_{\mathcal{A}}(\varepsilon_{max})^{(l-1)/(L-1)}$$

6.1.3 Sequential Mapping. Another alternative is a sequential mapping, where the choice of the privacy level is reached in steps. This models the probability of the selected privacy level being in a particular category, given that it has surpassed the previous category. In some cases, decision-making unfolds in steps [3, 43, 70, 96], such as deciding whether something is private and then determining the degree of privacy, e.g. "Is the stimulus sensitive/ does it violate a norm/should the information be shared? If yes, how private are the stimuli?/ How severe is the violation?/ How much information should be disclosed?". In these cases, automatic and heuristic-driven decisions (e.g., 'private or not') may precede more deliberative evaluations of privacy levels [3, 43]. In other cases, the stimuli's complexity may drive such behaviour, e.g. images with progressively more sensitive content may encourage stepwise evaluation (e.g., abstract shapes → objects → faces). Therefore, prior works [99, 106, 110] typically model this behaviour using step functions, incorporating heuristic-based decision-making as the primary mechanism to mitigate ambiguity aversion and prevent cognitive overload during deliberative evaluations.

Hence, a piecewise mapping (i.e. a step function) can be used to accommodate the different perceived privacy ranges to handle that different tiers of privacy ratings have different baseline ε values and slopes and allow more granularity within n ranges. Especially when certain ranges require higher privacy guarantees (e.g. handling extreme outliers). E.g. a look-up table where $n = 3$ (e.g. 'private', 'neutral', 'safe'):

$$g(l) = \begin{cases} \mathrm{adv}_{\mathcal{A}}(\varepsilon_{min}), & \text{if } l = 1 \text{ or } 2 \\ \frac{1}{2}(\mathrm{adv}_{\mathcal{A}}(\varepsilon_{max}) + \mathrm{adv}_{\mathcal{A}}(\varepsilon_{min})), & \text{if } 2 < l < L-1 \\ \mathrm{adv}_{\mathcal{A}}(\varepsilon_{max}), & \text{if } l = L \text{ or } L-1 \end{cases}$$

6.1.4 Sigmoid-based Mapping. Furthermore, adjacent-category models [16] are common in statistics and item response theory. They are usually used when thinking of a natural cognitive process is not possible and decisions may involve iterative, contextual, uncertain, or unstructured processes. In these cases, users can make fragmented decisions or skip steps due to limited information. In other cases, similar stimuli with only slight changes in features (e.g., cropped vs. full image) require users to make nuanced decisions. Commonly in machine learning and statistics, sigmoid (logistic) functions are used to describe the probability of selecting one option over another under uncertainty. It naturally models probabilistic decisions that become more deterministic as the evidence or difference between choices increases.

This behaviour aligns well with empirical findings in psychology, economics, and cognitive science, where decision-making often follows a logistic-like pattern. That is, in stochastic cases, for highly private images where the perceived privacy is greater (i.e. lower ratings), the sigmoid function outputs a lower ε value, accentuating privacy protection. Conversely, for images considered safe, the function assigns a higher ε, allowing less noise and thus better utility of the data.

Hence, a sigmoid mapping function can be expressed as the smooth continuation of

$$g(l) = \mathrm{adv}_{\mathcal{A}}(\varepsilon_{max}) + \frac{\mathrm{adv}_{\mathcal{A}}(\varepsilon_{min}) - \mathrm{adv}_{\mathcal{A}}(\varepsilon_{max})}{1 + \left(\frac{L-1}{l-1} - 1\right)^{-k}}$$

where k controls the steepness of the curve (higher values make the transition sharper, e.g. $k = 1.5$ represents moderate steepness). As shown in Fig. 7, such smooth transitions ensure that ε values transition gradually instead of changing too sharply with diminishing

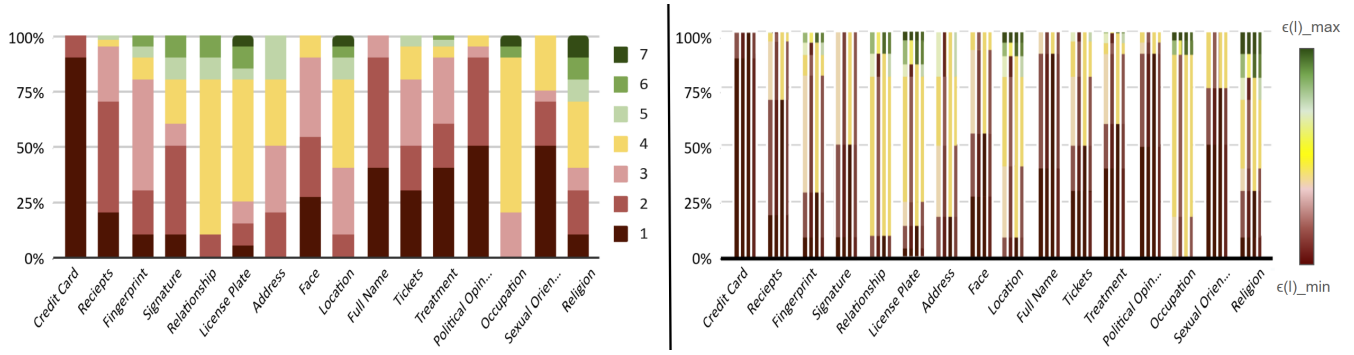


Figure 7: The left figure shows the distribution of the perceived privacy levels $l \in$ {1 (very private),...,7 (very safe)} selected by the participants per attribute. The right figure shows an example of mapping the perceived privacy level l to the corresponding $\varepsilon(l) = f(l)$ using the different mapping functions for $\varepsilon_{min} = 0.1, \varepsilon_{max} = 5$ and $k = 1.5$. The results are applied to the l distribution where, for each attribute, the mapping functions (linear, exponential, sequential, and sigmoid) are depicted from left to right.

effects at the extremes, allowing to smoothly handle uncertainty or borderline cases without abrupt changes in behaviour.

6.2 Empirical Evaluation

We now show how our results from Section 5 could be used in combination with mappings to epsilon values proposed in Section 6.1. To illustrate the applicability of our approach, we adopt a representative example from the *privacy level perception* task. This example serves to show how our method can be operationalised in practice. Importantly, the proposed framework is not limited to this specific task; it can be extended to other tasks by incorporating stimuli- or user-based information, thereby enhancing its generalizability across various privacy-sensitive applications.

We apply the state-of-the-art personalised differential privacy (PDP) mechanisms for privacy budgets ε determined with our mapping functions from Section 6.1. We employ the existing PDP methods without modification to their core mechanisms. The only alteration involves replacing the randomly generated privacy-level inputs (ε) commonly used in prior work with user-specified privacy preferences inferred from gaze. This allows for a more realistic and user-informed evaluation of PDP behaviour while preserving the original algorithmic structure. For our comparison, we use different common tasks from data analysis and machine learning tasks, which we want to describe briefly next.

Benchmarks. Our evaluation uses the same benchmarks as Jorgensen et al. [61] to evaluate PDP mechanisms for search analytics, e.g. count query for the number of documents that participants searched for, median and minimum queries for attention allocation (number of fixations) per stimulus. Once a ε-value is determined, either randomly by [61] or by applying our method on the user's gaze data, a trusted data analyst receives the data and the ε values, then adds suitable noise to query results and publishes the aggregate statistics. In addition to these classical data analysis benchmarks, we also evaluate PDP mechanisms for machine learning tasks, namely, search intention prediction, using the default models and parameters of [10, 11, 61] (cf. Section 3).

Results. As shown in Table 4, PDP approaches significantly improve over the static DP approaches (i.e. the worst-case privacy guarantee by uniformly applying the maximum privacy level required across all users, without adapting to individual preferences or contexts to address the most stringent privacy demands, the common approach in DP) and are significantly closer to the plain approaches (i.e. without noise addition). In addition, our proposed mapping functions yield better utility than the random benchmark (i.e. the random ε values generated by the existing benchmarks). To further analyse the effect of the different mapping functions, Fig. 7 shows the resulting ϵ values for different image attributes. The amount of added noise (according to the ε values) differs according to the mapping function (i.e. more noise is introduced to the private levels represented in red, whereas less noise is applied to those depicted in green), hence improving utility compared to static worst-case approaches. Our results indicate that:

- **The linear mapping** is particularly appealing in scenarios where exact precision is not critical, and approximate representations are sufficient for the intended application, e.g. when ratings are equally distributed, unlike our dataset, where the number of classes is skewed toward the private ratings. In former cases, linear functions offer a simpler and more computationally efficient mapping. However, the trade-off lies in the potential loss of granularity or accuracy.
- **The exponential mapping** is mostly suitable for applications where strong privacy is paramount and small changes in high ratings imply steep privacy needs. This is true for clearly sensitive attributes, i.e. attributes with an l range ≤ 2, such as credit cards. This could also be seen in the higher performance for weighting algorithms in Table 4 where less private data contributes more (given the skewed data distribution as shown in Fig. 7) with less noise to the final learning outcome.
- **The sequential mapping** demonstrates superior performance for attributes with deliberative privacy sensitivity ($2 < range(l) < 5$), such as political opinions, where participants' gaze shifts between the privacy levels before selection, suggesting an initial classification of the attribute as private, followed by a secondary assessment of its degree of privacy.

Table 4: Evaluation of the PDP benchmarks. The table shows results for data analysis and machine learning tasks where (i) plain is the non-private computation and, hence, the best utility, (ii) static is the worst-case privacy loss and the most commonly used in standard DP protocols, (iii) random is a random distribution of ϵ values that are commonly used in PDP protocols, and (iv) our four proposed mapping functions.

		Plain	Static	Random	Linear	Exponential	Sequential	Sigmoid
Analysis	Count [61]	100	76	84	87	90	86	87
	Median [61]	16	12	19	14	18	19	19
	Min [61]	5	8	7	7	4	7	6
Learning	Linear regression [61]	0.57	0.31	0.43	0.46	0.50	0.49	0.47
	Weighting [11]	0.68	0.38	0.52	0.53	0.57	0.56	0.55
	Weighting [10]	0.49	0.32	0.43	0.44	0.46	0.45	0.42
	Sampling [11]	0.68	0.38	0.55	0.58	0.60	0.61	0.63
	Sampling [10]	0.49	0.32	0.43	0.43	0.44	0.44	0.46

- **The sigmoid mapping** is most effective for attributes where participants either make arbitrary choices (can also be seen in minimal rating time) or are uncertain (can be seen in extended rating time), resulting in attributes with $range(l) > 5$, e.g., license plates that some participants did not pay attention to (i.e., no fixations) or fingerprints with long fixations. This could also be seen in the better performance of the sigmoid mapping for sampling algorithms in Table 4 where neutral samples are sampled more, balancing the data distribution and the added noise, hence improving utility.

User expectation alignment. To assess whether the predictions and protections provided by Gaze3P align with user expectations, we conducted a follow-up user validation study with a subset of $N = 20$ participants. While our quantitative evaluations above demonstrate the technical validity of Gaze3P (e.g., accuracy), the user study examines the users' subjective acceptance of Gaze3P.

For the validation study, each participant was shown the same set of images they had rated previously during the initial dataset collection (the search task), along with their user-specified privacy levels. For each image, we generated a reconstructed version using the standard reconstruction attack pipeline from MLDoctor [82] to attack the PDP learning models, simulating an adversarial attempt to recover visual data from differentially private representations. These reconstructions, shown in Fig. 8, reflect what an external observer might infer about the original image when Gaze3P's predicted privacy levels are used in PDP. Participants were asked three questions measuring (i) utility, (ii) privacy, and (iii) cognition: (i) if the model's search target prediction was correct. Responses were binary ('yes' or 'no'). (ii) whether the reconstructed image matched their privacy expectations, given the privacy level they previously selected. Responses were collected on a 5-point Likert scale, ranging from 'Not at all' to 'Perfectly aligned', and (iii) if they prefer a certain mapping function, given this privacy-utility tradeoff. Responses were again collected on a 5-point Likert scale per mapping function. We repeated this procedure for each stimulus, leading to each participant being shown 250 samples (the same 50 search stimuli that were previously shown during data collection times our 5 benchmark models).

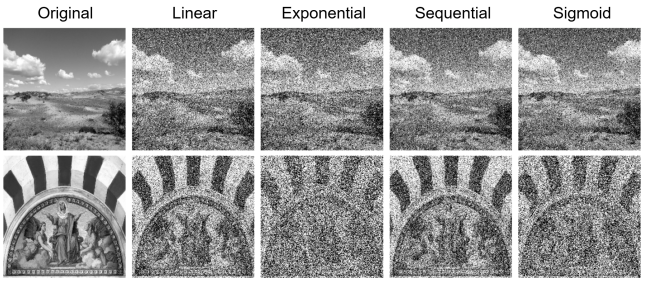


Figure 8: Random samples of the reconstructed stimuli (first row $l = 7$, second row $l = 1$.

Results show that the subsample shown to participants was representative, and utility evaluation was close to the numbers reported in Table 4 with 0.5 (i.e. 50%), 0.6, 0.5, 0.7, and 0.5 correct predictions on average for the learning models in the same order. In addition, with an average rating of 3.8/5, the participants' privacy expectations were met. Finally, overall, participants preferred the sequential mapping in 68% of the cases. They preferred the exponential mapping in 85% of the private images (rated with $l \geq 6$) and the sequential mapping in 52% of the images rated as 'neutral', which supports our quantitative results. However, the sigmoid mapping was only selected in 7% of the cases, showing no clear pattern. Therefore, given the absence of a universal solution, we recommend selecting the appropriate mapping function based on the specific application and the characteristics of its stimuli.

7 Discussion

The introduction of Gaze3P marks a pioneering step in privacy perception, advancing beyond traditional questionnaires by leveraging implicit, dynamic signals derived from gaze behaviour, underscoring that users' visual attention can implicitly reveal privacy sensitivities—offering a real-time, non-intrusive window into subjective privacy evaluation. More concretely, our dataset showed that eye gaze reliably reflects user perceptions across privacy-related tasks **(RQ1)**. Our ML approach further revealed user- and stimulus-based insights that could be used across a variety of applications **(RQ2)**:

We demonstrated the predictive power of gaze for privacy perception, the behavioural and contextual influences of gaze (e.g. fixation patterns, visual attention, demographic context, and expertise) in shaping privacy perceptions, the potential of gaze-based inference and enhancement of privacy-related attribute recognition and user profiling. By modelling perceived privacy alongside formal guarantees, we improved data utility while (cognitively) aligning with individual privacy expectations **(RQ3)**. Although a single universal mapping between perception and privacy budgets is unlikely to exist due to inter-individual variability, we propose a set of adaptable mapping functions that can be selected contextually based on application needs and behavioural insights.

7.1 Limitations

We acknowledge an age-related bias in our data, as the participant pool was not age-controlled. Despite a publicly announced call, recruitment likely skewed towards university students and staff due to convenience and accessibility. We also collected supplementary standard demographics, including education level, occupation, and field of study. However, our analysis revealed limited variability across these dimensions, as the majority of participants were students enrolled in STEM (Science, Technology, Engineering, Mathematics) disciplines, with only a few exceptions (4 participants). Consequently, while the current dataset may not support strong generalization across diverse demographic strata, this metadata, as well as our data collection software, is included in [8]. We hope that this allows others to extend analyses or augmentations in future work, particularly with a more demographically diverse sample.

Moreover, participants were explicitly instructed to treat the images as if they were their own (e.g., from a personal phone gallery), in line with Sattar et al. [114]. They were also informed that the study focused on privacy, in accordance with the ethical guidelines and prior research practices [101, 118]. While we acknowledge the limitations of this setup and the possible biases it may induce, such constraints are typical in early-stage studies within emerging research domains. Nonetheless, this work serves as a foundational step, enabling future research to adopt more comprehensive methodologies and engage more diverse participant groups to improve generalizability and depth of analysis.

Similarly, another limitation of the present study is the introduction of contextual bias through priming participants on the topic of privacy. This framing was necessary to align participants with the experimental objectives; however, it likely influenced cognitive processing and visual attention during the tasks. As a result, our predictions are not solely inferred from gaze behaviour in a neutral context but rather from gaze patterns shaped by a privacy-salient environment. This contextualization restricts the generalizability of the findings, as gaze allocation strategies may differ when privacy is not explicitly emphasised. Consequently, the reported predictive performance should be interpreted as applying to privacy-aware scenarios, not to all interaction contexts. Future research should address this by employing between-subject designs with primed and non-primed conditions or by modelling contextual factors explicitly, to disentangle intrinsic gaze-based indicators of privacy perception from those induced by experimental framing.

7.2 Future Work

We showed that perceived privacy preferences can be inferred from gaze behaviour alone. While not the main focus of the present analysis, we also observed that contextual data, such as demographics, have the potential to enrich predictions. Future work can build on this foundation by incorporating such auxiliary information to enhance both accuracy and personalisation. Similarly, additional metadata on individual user traits, such as trust propensity, risk perception, and prior exposure to privacy threats, can be collected to enable a deeper understanding of how such latent factors implicitly influence privacy-related behaviours [36, 50, 108].

Additionally, our proposed dataset includes rich information that we encourage the community to develop upon. This can include (i) additional tasks, such as multi-attribute and multi-user correlation analyses (i.e. each group of stimuli was shown to various participants, with each stimulus containing several attributes), (ii) tailored models to capture more patterns and enhance our baseline performance, (iii) other DP and privacy-preserving protocols (i.e. adding noise to different privacy units, the fundamental entity whose privacy is protected, user-, label-, feature- or pixel-level DP) that would benefit from the user (cognitive) privacy perceptions without explicit interaction, and (iv) fine-grained eye-tracking analyses to gain a deeper understanding of gaze behaviour in privacy-sensitive contexts. While the present study demonstrates that learned gaze features can serve as effective predictors of perceived privacy, more granular analyses - e.g. gaze entropy, micro-saccade dynamics, and scanpath structure - may reveal subtle cognitive and affective processes underlying privacy perception and could help disentangle the interplay between bottom-up and top-down visual attention in private settings, improving the interpretability of gaze-based privacy models.

8 Conclusion

Given its inherently subjective nature, which varies substantially across individuals, we presented the first large-scale dataset for studying user-perceived privacy. The dataset encompasses a diverse range of participants, demographic profiles, and visual stimuli. Using this novel dataset, we demonstrated that eye gaze can serve as a rich source of information on user-perceived privacy across multiple privacy-related tasks. Gaze behaviour, by providing implicit and dynamic feedback, offers a powerful and promising avenue for enhancing user interaction and overall system usability. Moreover, by modelling users' perceived privacy and applying our findings to PDP protocols – complementing the underlying mathematical and technical privacy guarantees – we were able to improve data utility while better aligning with users' expectations of privacy. As such, our work bridges the gap between technical and usable privacy by aligning theoretical privacy models with user perceptions.

Acknowledgments

We thank Manpa Barman and Yanhong Xu for their assistance in recruiting participants. This project was funded by the Ministry of Science, Research and the Arts Baden-Württemberg in the Artificial Intelligence Software Academy (AISA) and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - 411720488 and 548713845.

References

[1] Yaseem Abdrabou, Mariam Hassib, Shuqin Hu, Ken Pfeuffer, Mohamed Khamis, Andreas Bulling, and Florian Alt. 2024. EyeSeeIdentity: Exploring Natural Gaze Behavior for Implicit User Identification during Photo Viewing. (2024).

[2] Krishna Acharya, Franziska Boenisch, Rakshit Naidu, and Juba Ziani. 2024. Personalized differential privacy for ridge regression. *arXiv preprint arXiv:2401.17127* (2024).

[3] Zahra Aivazpour, Rohit Valecha, and Raghav H Rao. 2017. Unpacking privacy paradox: a dual process theory approach. (2017).

[4] Mohammad Alaggan, Sébastien Gambs, and Anne-Marie Kermarrec. 2015. Heterogeneous differential privacy. *arXiv preprint arXiv:1504.06998* (2015).

[5] Bonnie Brinton Anderson, Jeffrey L Jenkins, Anthony Vance, C Brock Kirwan, and David Eargle. 2016. Your memory is working against you: How eye tracking and memory explain habituation to security warnings. *Decision Support Systems* 92 (2016), 3–13.

[6] Ion Androutsopoulos, Georgios Paliouras, and Eirinaios Michelakis. 2004. *Learning to filter unsolicited commercial e-mail.* Vol. 2004. " DEMOKRITOS", National Center for Scientific Research.

[7] Apple. 2025. Privacy Labels. https://www.apple.com/privacy/labels/ Accessed: 2025-02-18.

[8] Anonymous Author(s). 2025. Our Implementation Data to Gaze3P. To be published upon acceptance.

[9] Ardion Beldad, Menno De Jong, and Michaël Steehouder. 2011. I trust not therefore it must be risky: Determinants of the perceived risks of disclosing personal data for e-government transactions. *Computers in Human Behavior* 27, 6 (2011), 2233–2242.

[10] Franziska Boenisch, Christopher Mühl, Adam Dziedzic, Roy Rinberg, and Nicolas Papernot. 2023. Have it your way: Individualized Privacy Assignment for DP-SGD. In *Advances in Neural Information Processing Systems*, A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (Eds.), Vol. 36. Curran Associates, Inc., 19073–19103. https://proceedings.neurips.cc/paper_files/paper/2023/file/3cbf627fa24fb6cb576e04e689b9428b-Paper-Conference.pdf

[11] Franziska Boenisch, Christopher Mühl, Roy Rinberg, Jannis Ihrig, and Adam Dziedzic. 2022. Individualized PATE: Differentially private machine learning with individual privacy guarantees. *arXiv preprint arXiv:2202.10517* (2022).

[12] 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.

[13] 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.

[14] Stephanie Brams, Gal Ziv, Oron Levin, Jochim Spitz, Johan Wagemans, A Mark Williams, and Werner F Helsen. 2019. The relationship between gaze behavior, expertise, and performance: A systematic review. *Psychological bulletin* 145, 10 (2019), 980.

[15] Andreas Bulling and Daniel Roggen. 2011. Recognition of Visual Memory Recall Processes Using Eye Movement Analysis. In *ACM UbiComp*. 455–464. https://doi.org/10.1145/2030112.2030172

[16] Paul-Christian Bürkner and Matti Vuorre. 2019. Ordinal regression models in psychology: A tutorial. *Advances in Methods and Practices in Psychological Science* 2, 1 (2019), 77–101.

[17] 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.

[18] Benjamin T Carter and Steven G Luke. 2020. Best practices in eye tracking research. *International Journal of Psychophysiology* 155 (2020), 49–62.

[19] Yimin Chen, Tao Li, Rui Zhang, Yanchao Zhang, and Terri Hedgpeth. 2018. EyeTell: Video-Assisted Touchscreen Keystroke Inference from Eye Movements. In *2018 IEEE Symposium on Security and Privacy (SP)*. 144–160. https://doi.org/10.1109/SP.2018.00010

[20] Bryan H Choi. 2014. A Prospect Theory of Privacy. *Idaho L. Rev.* 51 (2014), 623.

[21] Mary J Culnan and Pamela K Armstrong. 1999. Information privacy concerns, procedural fairness, and impersonal trust: An empirical investigation. *Organization science* 10, 1 (1999), 104–115.

[22] Rachel Cummings, Damien Desfontaines, David Evans, Roxana Geambasu, Yangsibo Huang, Matthew Jagielski, Peter Kairouz, Gautam Kamath, Sewoong Oh, Olga Ohrimenko, et al. 2023. Advancing differential privacy: Where we are now and future directions for real-world deployment. *arXiv preprint arXiv:2304.06929* (2023).

[23] Rachel Cummings and David Durfee. 2020. Individual sensitivity preprocessing for data privacy. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM, 528–547.

[24] Rachel Cummings, Katrina Ligett, Aaron Roth, Zhiwei Steven Wu, and Juba Ziani. 2015. Accuracy for sale: Aggregating data with a variance constraint. In *Proceedings of the 2015 conference on innovations in theoretical computer science*. 317–324.

[25] Brendan David-John, Kevin Butler, and Eakta Jain. 2022. For your eyes only: Privacy-preserving eye-tracking datasets. In *ACM ETRA*. 1–6.

[26] Brendan David-John, Diane Hosfelt, Kevin Butler, and Eakta Jain. 2021. A privacy-preserving approach to streaming eye-tracking data. *IEEE TVCG* 27, 5 (2021), 2555–2565.

[27] Brendan David-John, Candace Peacock, Ting Zhang, T Scott Murdison, Hrvoje Benko, and Tanya R Jonker. 2021. Towards gaze-based prediction of the intent to interact in virtual reality. In *ACM ETRA*. 1–7.

[28] Tamara Dinev and Paul Hart. 2006. An extended privacy calculus model for e-commerce transactions. *Information systems research* 17, 1 (2006), 61–80.

[29] D Dittrich and E Kenneally. 2012. *The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research.* Technical Report. U.S. Department of Homeland Security. https://doi.org/paper/2012_menlo_report_actual_formatted

[30] Essohanam Djeki, Jules Dégila, and Muhtar Hanif Alhassan. 2024. Reimagining Authentication: A User-Centric Two-Factor Authentication with Personalized Image Verification. In *2024 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETSIS)*. IEEE, 281–285.

[31] Cynthia Dwork. 2006. Differential Privacy. In *Automata, Languages and Programming*, Michele Bugliesi, Bart Preneel, Vladimiro Sassone, and Ingo Wegener (Eds.). Springer Berlin Heidelberg.

[32] Cynthia Dwork, Aaron Roth, et al. 2014. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science* 9, 3–4 (2014), 211–407.

[33] Alessandro D'Amelio, Sabrina Patania, Sathya Bursic, Vittorio Cuculo, and Giuseppe Boccignone. 2023. Using gaze for behavioural biometrics. *Sensors* 23, 3 (2023), 1262.

[34] Hamid Ebadi, David Sands, and Gerardo Schneider. 2015. Differential privacy: Now it's getting personal. *Acm Sigplan Notices* 50, 1 (2015), 69–81.

[35] Maria K Eckstein, Belén Guerra-Carrillo, Alison T Miller Singley, and Silvia A Bunge. 2017. Beyond eye gaze: What else can eyetracking reveal about cognition and cognitive development? *Developmental cognitive neuroscience* 25 (2017), 69–91.

[36] Yusra Elbitar, Michael Schilling, Trung Tin Nguyen, Michael Backes, and Sven Bugiel. 2021. Explanation beats context: The effect of timing & rationales on users' runtime permission decisions. In *30th USENIX Security Symposium (USENIX Security 21)*. 785–802.

[37] 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

[38] Mayar Elfares, Pascal Reisert, Zhiming Hu, Ralf Küsters, and Andreas Bulling. 2023. PrivatEyes: Appearance-based Gaze Estimation Using Federated Secure Multi-Party Computation. Preprint. A copy has been attached to this proposal..

[39] Mayar Elfares, Pascal Reisert, Ralf Küsters, and Andreas Bulling. 2025. QualitEye: Public and Privacy-preserving Gaze Data Quality Verification. *arXiv preprint arXiv:2506.05908* (2025).

[40] Myrthe Faber, Robert Bixler, and Sidney K D'Mello. 2018. An automated behavioral measure of mind wandering during computerized reading. *Behavior Research Methods* 50, 1 (2018), 134–150.

[41] Alireza Fallah, Ali Makhdoumi, Asuman Ozdaglar, et al. 2022. Bridging central and local differential privacy in data acquisition mechanisms. *Advances in Neural Information Processing Systems* 35 (2022), 21628–21639.

[42] Filippo Galli, Sayan Biswas, Kangsoo Jung, Tommaso Cucinotta, and Catuscia Palamidessi. 2022. Group privacy for personalized federated learning. https://doi.org/10.48550/ARXIV.2206.03396

[43] Mohamad Gharib. 2024. Towards a Heuristic Model for Usable Privacy.. In *RCIS Workshops*.

[44] Arpita Ghosh and Aaron Roth. 2011. Selling privacy at auction. In *Proceedings of the 12th ACM conference on Electronic commerce*. 199–208.

[45] Google. 2021. Federated Learning of Cohorts (FLoC) Whitepaper. https://raw.githubusercontent.com/google/ads-privacy/master/proposals/FLoC/FLOC-Whitepaper-Google.pdf Accessed: 2025-02-18.

[46] Google. 2025. Google Safe Browsing. https://safebrowsing.google.com/ Accessed: 2025-02-18.

[47] 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

[48] Michele Guerra, Roberto Milanese, Madalina G Ciobanu, Michele Deodato, and Fausto Fasano. 2023. Seeing is Believing: Assessing and Enhancing Android Privacy Indicators Through Eye-Tracking Analysis. In *International Conference on Information Systems Security and Privacy*. Springer, 127–158.

[49] Michele Guerra, Roberto Milanese, Michele Deodato, Vittorio Perozzi, Fausto Fasano, et al. 2024. Visual Attention and Privacy Indicators in Android: Insights from Eye Tracking.. In *ICISSP*. 320–329.

[50] Michele Guerra, Simone Scalabrino, Fausto Fasano, and Rocco Oliveto. 2023. An empirical study on the effectiveness of privacy indicators. *IEEE Transactions on Software Engineering* 49, 10 (2023), 4610–4623.

[51] Eddie Harmon-Jones and Cindy Harmon-Jones. 2012. Cognitive dissonance theory. *Handbook of motivation science* 71 (2012).

[52] Glenn W Harrison and J Todd Swarthout. 2019. Eye-tracking and economic theories of choice under risk. *Journal of the Economic Science Association* 5, 1 (2019), 26–37.

[53] Roy S Hessels, Antje Nuthmann, Marcus Nyström, Richard Andersson, Diederick C Niehorster, and Ignace TC Hooge. 2025. The fundamentals of eye tracking part 1: The link between theory and research question. *Behavior Research Methods* 57, 1 (2025), 1–18.

[54] Ignace TC Hooge, Antje Nuthmann, Marcus Nyström, Diederick C Niehorster, Gijs A Holleman, Richard Andersson, and Roy S Hessels. 2025. The fundamentals of eye tracking part 2: From research question to operationalization. *Behavior Research Methods* 57, 2 (2025), 73.

[55] Roberto Hoyle, Luke Stark, Qatrunnada Ismail, David Crandall, Apu Kapadia, and Denise Anthony. 2020. Privacy norms and preferences for photos posted online. *ACM Transactions on Computer-Human Interaction (TOCHI)* 27, 4 (2020), 1–27.

[56] Justin Hsu, Marco Gaboardi, Andreas Haeberlen, Sanjeev Khanna, Arjun Narayan, Benjamin C Pierce, and Aaron Roth. 2014. Differential privacy: An economic method for choosing epsilon. In *2014 IEEE 27th Computer Security Foundations Symposium*. IEEE, 398–410.

[57] Vincent C Hu, D Richard Kuhn, David F Ferraiolo, and Jeffrey Voas. 2015. Attribute-based access control. *Computer* 48, 2 (2015), 85–88.

[58] 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.

[59] Athina Ioannou, Iis Tussyadiah, Graham Miller, Shujun Li, and Mario Weick. 2021. Privacy nudges for disclosure of personal information: A systematic literature review and meta-analysis. *PloS one* 16, 8 (2021), e0256822.

[60] Carl Anderson Johnson. 1974. Privacy as personal control. *Man-environment interactions: evaluations and applications: part 2* (1974), 83–100.

[61] Zach Jorgensen, Ting Yu, and Graham Cormode. 2015. Conservative or liberal? Personalized differential privacy. In *2015 IEEE 31St international conference on data engineering*. IEEE, 1023–1034.

[62] Daniel Kahneman and Amos Tversky. 2013. Prospect theory: An analysis of decision under risk. In *Handbook of the fundamentals of financial decision making: Part I*. World Scientific, 99–127.

[63] Peter Kairouz, H Brendan McMahan, Brendan Avent, Bellet, et al. 2021. Advances and open problems in federated learning. *Foundations and Trends® in Machine Learning* 14, 1–2 (2021), 1–210.

[64] Christina Katsini, Yasmeen Abdrabou, George E. Raptis, Mohamed Khamis, and Florian Alt. 2020. The Role of Eye Gaze in Security and Privacy Applications: Survey and Future HCI Research Directions. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–21. https://doi.org/10.1145/3313831.3376840

[65] Murat Kezer, Tobias Dienlin, and Lemi Baruh. 2022. Getting the privacy calculus right: Analyzing the relations between privacy concerns, expected benefits, and self-disclosure using response surface analysis. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace* 16, 4 (2022).

[66] Wonjae Kim, Bokyung Son, and Ildoo Kim. 2021. Vilt: Vision-and-language transformer without convolution or region supervision. In *International conference on machine learning*. PMLR, 5583–5594.

[67] Bart P Knijnenburg, Reza Ghaiumy Anaraky, Daricia Wilkinson, Moses Namara, Yangyang He, David Cherry, and Erin Ash. 2022. User-Tailored Privacy.

[68] Leanne K Knobloch. 2008. Uncertainty reduction theory. *Engaging theories in interpersonal communication* (2008), 133–144.

[69] Nitin Kohli and Paul Laskowski. 2018. Epsilon voting: Mechanism design for parameter selection in differential privacy. In *2018 IEEE Symposium on Privacy-Aware Computing (PAC)*. IEEE, 19–30.

[70] Spyros Kokolakis. 2017. Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers & security* 64 (2017), 122–134.

[71] 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.

[72] Matthias Kümmerer, Matthias Bethge, and Thomas S. A. Wallis. 2022. DeepGaze III: Modeling free-viewing human scanpaths with deep learning. *Journal of Vision* 22, 5 (04 2022), 7–7. https://doi.org/10.1167/jov.22.5.7 arXiv:https://arvojournals.org/arvo/content_public/journal/jov/938587/i1534-7362-22-5-7_1650885429.74577.pdf

[73] Benjamin Law, M Stella Atkins, Arthur E Kirkpatrick, and Alan J Lomax. 2004. Eye gaze patterns differentiate novice and experts in a virtual laparoscopic surgery training environment. In *Proceedings of the 2004 symposium on Eye tracking research & applications*. 41–48.

[74] Jaewoo Lee and Chris Clifton. 2011. How much is enough? choosing ε for differential privacy. In *Information Security: 14th International Conference, ISC 2011, Xi'an, China, October 26-29, 2011. Proceedings 14*. Springer, 325–340.

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

[76] Meng Li, Liehuang Zhu, Zijian Zhang, and Rixin Xu. 2017. Achieving differential privacy of trajectory data publishing in participatory sensing. *Information Sciences* 400 (2017), 1–13.

[77] Ninghui Li, Tiancheng Li, and Suresh Venkatasubramanian. 2006. t-closeness: Privacy beyond k-anonymity and l-diversity. In *2007 IEEE 23rd international conference on data engineering*. IEEE, 106–115.

[78] Jonathan Liebers and Stefan Schneegass. 2020. Gaze-based authentication in virtual reality. In *ACM Symposium on Eye Tracking Research and Applications*. 1–2.

[79] 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.

[80] Fan Liu, Zhiyong Cheng, Huilin Chen, Yinwei Wei, Liqiang Nie, and Mohan Kankanhalli. 2022. Privacy-preserving synthetic data generation for recommendation systems. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1379–1389.

[81] Yan Liu, Pei-Yun Hsueh, Jennifer Lai, Mirweis Sangin, Marc-Antoine Nussli, and Pierre Dillenbourg. 2009. Who is the expert? Analyzing gaze data to predict expertise level in collaborative applications. In *2009 IEEE international conference on Multimedia and Expo*. IEEE, 898–901.

[82] Yugeng Liu, Rui Wen, xinlei.he, Ahmed Salem, Zhikun Zhang, Michael Backes, Emiliano De Cristofaro, Mario Fritz, and Yang Zhang. 2022. ML-Doctor: Holistic Risk Assessment of Inference Attacks Against Machine Learning Models. (8 2022). https://doi.org/10.60882/cispa.24614037.v1

[83] Mingjie Lu and Zhenhua Liu. 2023. Improving Accuracy of Interactive Queries in Personalized Differential Privacy. In *International Conference on Frontiers in Cyber Security*. Springer, 141–159.

[84] Shiqing Luo, Anh Nguyen, Chen Song, Feng Lin, Wenyao Xu, and Zhisheng Yan. 2020. OcuLock: Exploring Human Visual System for Authentication in Virtual Reality Head-mounted Display. In *2020 Network and Distributed System Security Symposium (NDSS)*. https://doi.org/10.14722/ndss.2020.24079

[85] Ashwin Machanavajjhala, Daniel Kifer, Johannes Gehrke, and Muthuramakrishnan Venkitasubramaniam. 2007. l-diversity: Privacy beyond k-anonymity. *Acm transactions on knowledge discovery from data (tkdd)* 1, 1 (2007), 3–es.

[86] Kelly D Martin and Patrick E Murphy. 2017. The role of data privacy in marketing. *Journal of the Academy of Marketing Science* 45 (2017), 135–155.

[87] Philipp K Masur. 2018. *Situational privacy and self-disclosure: Communication processes in online environments*. Springer.

[88] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. 2017. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*. PMLR, 1273–1282.

[89] Yannic Meier and Nicole C Krämer. 2024. The privacy calculus revisited: an empirical investigation of online privacy decisions on between-and within-person levels. *Communication Research* 51, 2 (2024), 178–202.

[90] Meta. 2024. Providing Context on Sensitive or Misleading Content. https://transparency.meta.com/en-gb/enforcement/taking-action/context-on-sensitive-misleading-content/ Accessed: 2025-02-18.

[91] Sounak Mondal, Zhibo Yang, Seoyoung Ahn, Dimitris Samaras, Gregory Zelinsky, and Minh Hoai. 2023. Gazeformer: Scalable, effective and fast prediction of goal-directed human attention. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 1441–1450.

[92] Milica Milosavljevic Mormann and Cary Frydman. 2016. The role of salience and attention in choice under risk: An experimental investigation. *SSRN Electronic Journal* 10 (2016).

[93] Gilbert Mushure. 2017. Measuring Perception: The limitations of the questionnaire methodology for gathering research data. (09 2017). https://doi.org/10.13140/RG.2.2.24689.17766

[94] Milad Nasr, Shuang Songi, Abhradeep Thakurta, Nicolas Papernot, and Nicholas Carlin. 2021. Adversary instantiation: Lower bounds for differentially private machine learning. In *2021 IEEE Symposium on security and privacy (SP)*. IEEE, 866–882.

[95] Diederick C Niehorster, Marcus Nyström, Roy S Hessels, Richard Andersson, Jeroen S Benjamins, Dan Witzner Hansen, and Ignace TC Hooge. 2025. The fundamentals of eye tracking part 4: Tools for conducting an eye tracking study. *Behavior Research Methods* 57, 1 (2025), 46.

[96] Helen Nissenbaum. 2004. Privacy as contextual integrity. *Wash. L. Rev.* 79 (2004), 119.

[97] Helen Nissenbaum. 2009. Privacy in context: Technology, policy, and the integrity of social life. In *Privacy in Context*. Stanford University Press.

[98] Ben Niu, Yahong Chen, Boyang Wang, Zhibo Wang, Fenghua Li, and Jin Cao. 2021. AdaPDP: Adaptive personalized differential privacy. In *IEEE INFOCOM 2021-IEEE conference on computer communications*. IEEE, 1–10.

[99] Georg Northoff and Felix Bermpohl. 2004. Cortical midline structures and the self. *Trends in cognitive sciences* 8, 3 (2004), 102–107.

[100] Marcus Nyström, Ignace TC Hooge, Roy S Hessels, Richard Andersson, Dan Witzner Hansen, Roger Johansson, and Diederick C Niehorster. 2025. The fundamentals of eye tracking part 3: How to choose an eye tracker. *Behavior Research Methods* 57, 2 (2025), 67.
[101] Tribhuvanesh Orekondy, Bernt Schiele, and Mario Fritz. 2017. Towards a visual privacy advisor: Understanding and predicting privacy risks in images. In *Proceedings of the IEEE international conference on computer vision*. 3686–3695.
[102] Sina Ostendorf, Silke M Müller, and Matthias Brand. 2020. Neglecting long-term risks: self-disclosure on social media and its relation to individual decision-making tendencies and problematic social-networks-use. *Frontiers in Psychology* 11 (2020), 543388.
[103] Yong Jin Park, Scott W Campbell, and Nojin Kwak. 2012. Affect, cognition and reward: Predictors of privacy protection online. *Computers in Human Behavior* 28, 3 (2012), 1019–1027.
[104] John H Parmelee and Nataliya Roman. 2020. Insta-echoes: Selective exposure and selective avoidance on Instagram. *Telematics and Informatics* 52 (2020), 101432.
[105] Sandra Petronio, Jeffrey T Child, and Robert D Hall. 2021. Communication privacy management theory: Significance for interpersonal communication. In *Engaging theories in interpersonal communication*. Routledge, 314–327.
[106] Elizabeth A Phelps and Joseph E LeDoux. 2005. Contributions of the amygdala to emotion processing: from animal models to human behavior. *Neuron* 48, 2 (2005), 175–187.
[107] Eric Plutzer. 2019. Privacy, sensitive questions, and informed consent: Their impacts on total survey error, and the future of survey research. *Public Opinion Quarterly* 83, S1 (2019), 169–184.
[108] Sarah Prange, Pascal Knierim, Gabriel Knoll, Felix Dietz, Alexander De Luca, and Florian Alt. 2024. {"I} do (not) need that {Feature!"}–Understanding {Users'} Awareness and Control of Privacy Permissions on Android Smartphones. In *Twentieth Symposium on Usable Privacy and Security (SOUPS 2024)*. 453–472.
[109] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*. PmLR, 8748–8763.
[110] Edmund T Rolls. 2004. The functions of the orbitofrontal cortex. *Brain and cognition* 55, 1 (2004), 11–29.
[111] Alexander K Saeri, Claudette Ogilvie, Stephen T La Macchia, Joanne R Smith, and Winnifred R Louis. 2014. Predicting Facebook users' online privacy protection: Risk, trust, norm focus theory, and the theory of planned behavior. *The Journal of social psychology* 154, 4 (2014), 352–369.
[112] Pierangela Samarati and Latanya Sweeney. 1998. Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression. (1998).
[113] 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.
[114] Hosnieh Sattar, Mario Fritz, and Andreas Bulling. 2020. Deep gaze pooling: Inferring and visually decoding search intents from human gaze fixations. *Neurocomputing* 387 (2020), 369–382. https://doi.org/10.1016/j.neucom.2020.01.028
[115] Herbert A Simon. 1990. Bounded rationality. *Utility and probability* (1990), 15–18.
[116] Malte Sönnichsen, Mayar Elfares, Yao Wang, Ralf Küsters, Alina Roitberg, and Andreas Bulling. 2025. AttentionLeak: What Does Human Attention Reveal About Information Visualisation? *IACR* (2025).
[117] 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
[118] Julian Steil, Marion Koelle, Wilko Heuten, Susanne Boll, and Andreas Bulling. 2019. PrivacEye: Privacy-Preserving Head-Mounted Eye Tracking Using Egocentric Scene Image and Eye Movement Features. In *ACM ETRA*. 1–10. https://doi.org/10.1145/3314111.3319913
[119] Filip-Mihai Toma, Cosmin-Octavian Cepoi, Matei Nicolae Kubinschi, and Makoto Miyakoshi. 2023. Gazing through the bubble: an experimental investigation into financial risk-taking using eye-tracking. *Financial Innovation* 9, 1 (2023), 28.
[120] Roel Vertegaal et al. 2003. Attentive user interfaces. *Commun. ACM* 46, 3 (2003), 30–33.
[121] Ari Ezra Waldman. 2020. Cognitive biases, dark patterns, and the 'privacy paradox'. *Current opinion in psychology* 31 (2020), 105–109.
[122] Hanqiu Wang, Zihao Zhan, Haoqi Shan, Siqi Dai, Maximilian Panoff, and Shuo Wang. 2024. GAZEploit: Remote Keystroke Inference Attack by Gaze Estimation from Avatar Views in VR/MR Devices. In *Proceedings of the 2024 on ACM SIGSAC Conference on Computer and Communications Security* (Salt Lake City, UT, USA) *(CCS '24)*. Association for Computing Machinery, New York, NY, USA, 1731–1745. https://doi.org/10.1145/3658644.3690285
[123] Yang Wang, Huichuan Xia, and Yun Huang. 2016. Examining American and Chinese internet users' contextual privacy preferences of behavioral advertising. In *Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing*. 539–552.
[124] Mengxiao Zhang, Fernando Beltran, and Jiamou Liu. 2020. Selling data at an auction under privacy constraints. In *Conference on Uncertainty in Artificial Intelligence*. PMLR, 669–678.

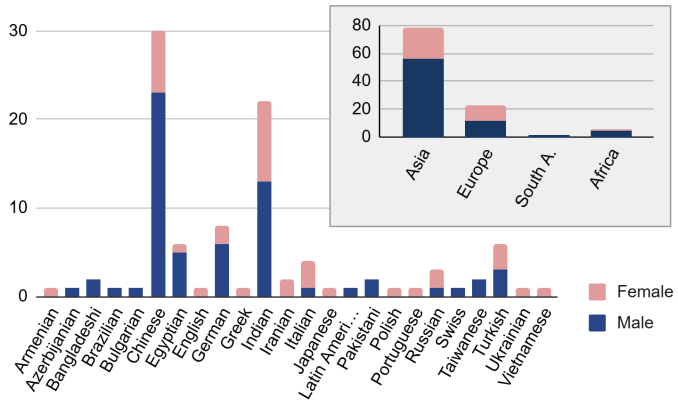


Figure 9: Demographics' distribution of the number of male and female participants per nationality and continent

Appendix

A Eye Tracking Data

Here, we present a detailed breakdown of the process of data collection with an eye tracker (EyeLink 1000), starting from participant recruitment to creating a public dataset:

A.1 Participant Recruitment

As shown in Figure 9, we aimed to recruit a diverse set of participants with different demographics, e.g. age, gender, nationalities, and AI/Security expertise. The call for participation was sent out on different channels, e.g. online platforms, university participant pools, and social media, with clear information about the study, duration, and incentives (e.g., monetary compensation or credits). Informed consent was obtained following ethical guidelines of the author's institution while ensuring that participants can withdraw at any time. We excluded participants with specific conditions like eye disorders that may affect tracking accuracy (e.g., nystagmus, extremely poor vision).

A.2 Eye Tracker Configuration

We used the EyeLink-1000 eye tracker, the current state-of-the-art in terms of precision and accuracy for video-based eye tracking. We created an in-lab setup as shown in Figure 10.

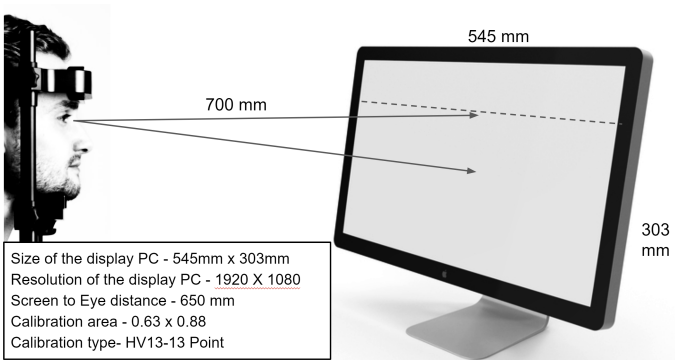


Figure 10: The eye-tracker setup

A.3 Experiment Design

A sample trial is conducted as shown in Figure 11.

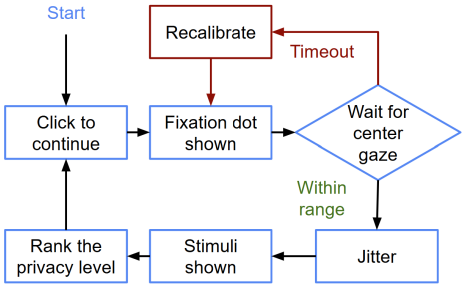


Figure 11: A sample trial flow-chart

The stimuli were gathered from the validation and test sets of the VISPR dataset [101]. Every block contained writing-based and human stimuli to avoid bias, with a distribution of private attributes as shown in Table 5.

Table 5: Stimuli distribution over blocks

Block	Attribute	Category
1	Fingerprint, Receipts, Occupation, Sexual Orientation, Political Opinion	Personal Description, Documents, Employment, Personal Life, Personal Life
2	Signature, Tickets, Medical Treatment, Personal Occasion, Home address	Personal Information, Documents, Health, Personal Life, Whereabouts
3	Face Complete, Credit Card, Medical History, Email Content, Religion	Personal Description, Documents, Health, Internet, Personal Life
4	Full Name, Mail, License Plate Complete, Personal Relationship, Visited Location	Personal Information, Documents, Automobile, Personal Life, Whereabouts

A.4 Data Collection

To accurately track participants' gaze and record eye movement data during the experiment, we first conduct a 13-point calibration to ensure the eye tracker is accurately mapping gaze coordinates to the screen, followed by a validation check to confirm that gaze accuracy is within acceptable error limits (e.g., <0.5°). We further recalibrate if drift or errors occur during the session. The stimuli are then presented, and data streams are recorded.

Essential methodological safeguards were implemented to maintain participant engagement and ensure high data quality throughout the experiment. These safeguards served to minimise bias, reduce participant fatigue, and confirm that responses reflected genuine attention and comprehension.

Randomised attention checks were embedded at multiple points during the study. These checks consisted of stimuli with clearly identifiable attributes—some explicitly private (e.g., credit cards) and others non-private (e.g., randomly generated colour patches). Participants were asked to classify or respond to these items. Correct responses indicated attentiveness, while incorrect answers were flagged for potential disengagement or misunderstanding.

To mitigate cognitive fatigue, structured breaks were introduced at predefined intervals (i.e. between tasks), allowing participants to rest and maintain focus. Participants were also allowed to stop the experiment at any point, if needed.

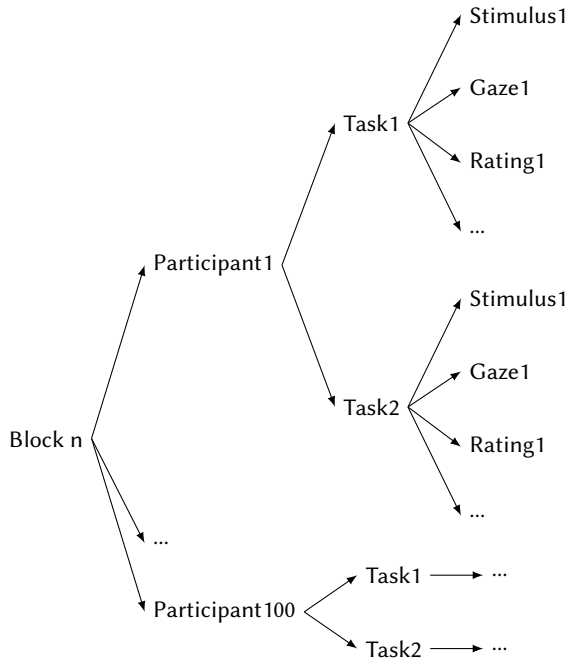
To address order and learning effects, both the sequence of experimental tasks and the presentation order of stimuli were fully randomised for each participant. This ensured that performance patterns could not be attributed to predictable task progression, practice effects, or fatigue tied to task order.

All these procedures align with established best practices in behavioural and user research [18, 53, 54, 95, 100], ensuring internal validity while safeguarding the participant experience.

A.5 Data Processing

We cleaned and processed raw eye-tracking data for further analysis. The raw data is extracted from the EyeLink Data Viewer as .edf files, including the timestamped gaze coordinates (X, Y), fixation and saccade metrics (e.g., duration, amplitude, velocity), and pupil size. We then used the stimulus metadata to map gaze coordinates to specific ROIs, categorised gaze events into relevant regions for analysis, and merged eye-tracking data with task-specific inputs (e.g., participant ratings and mouse clicks).

We structured the processed data into a usable dataset for analysis by organizing data into rows and columns (.csv format), including information such as participant demographics, stimuli details, and task condition. Finally, we formatted the data for statistical or machine learning tools (e.g., Python) in the following structure:



B ML Implementation Details

Given a dataset of inputs (e.g. gaze data) annotated with labels (e.g. user-selected privacy levels), we define the training set as:

$$\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$$

where: $\mathbf{x}_i \in \mathbb{R}^d$ represents the gaze data of the i^{th} sample (or a feature vector extracted from the stimuli and the gaze when both are inputs, as defined by each task), $y_i \in \mathcal{Y} = \{1, 2, \ldots, K\}$ is the corresponding label (e.g. user-provided privacy level or attribute name or privacy expertise, as defined by each task).

The training phase: In a supervised learning setting like classification, the training phase uses labelled data—pairs of input features (x_i) and their corresponding ground truth labels (y_i). During training, the model learns a mapping from inputs to labels by minimising a loss function that measures the discrepancy between predicted and true labels.

The goal is to learn a function $f_\theta : \mathbb{R}^d \to \mathcal{Y}$, parameterized by θ, that minimizes the classification loss:

$$\mathcal{L}(\theta) = \frac{1}{N} \sum_{i=1}^{N} \ell(f_\theta(\mathbf{x}_i), y_i)$$

where $\ell(\cdot, \cdot)$ is the loss function.

The inference phase: During the inference (testing) phase, the model receives only input features from new, unseen samples and generates predicted labels based on the patterns it learned during training. No true labels are available during inference; predictions are made autonomously using the trained model parameters.

More technically, a new unseen gaze sample $\mathbf{x}^* \in \mathbb{R}^d$ is processed through the trained model to produce a predicted label:

$$\hat{y} = f_{\theta^*}(\mathbf{x}^*)$$

We then compare the predictions $\hat{y}_i$ with the true labels y_i only to quantify the models' performance.

Train-test split: We employ cross-validation to optimise generalisation performance and prevent overfitting. In cross-validation, the data is split into k folds; the model is trained on $k-1$ folds and tested on the remaining one, repeating this process k times so that each fold serves as a test set once. The final performance is averaged across all folds, providing a more robust and unbiased estimate than a single train-test split. Here, we used an 80-20 train-test split.

ML models: In our implementation, we mostly employed classical machine learning algorithms available within the *Scikit-learn* framework, using their default configurations. *Decision Trees* were implemented using 'DecisionTreeClassifier'. *Support Vector Machines (SVMs)* were used via 'SVC' with an 'RBF kernel'. *Logistic Regression* was applied using 'LogisticRegression' with L2 regularisation. For *Random Forests*, we utilised 'RandomForestClassifier', an ensemble of decision trees trained on bootstrapped subsets of data with feature bagging. *K-Nearest Neighbors (KNN)* was implemented using 'KNeighborsClassifier', which assigns class labels based on the majority vote among the k most similar training samples in feature space. Further information and default hyperparameters can be found in the Scikit-Learn documentation: scikit-learn.org/stable/supervised_learning.html. Finally, for other advanced models, we used the original implementations provided by the referenced papers.

C Potential Real-World Applications

In this section, we discuss the potential SPP and UPP applications in more details:

SPP applications. Stimuli-based perceived privacy results can be used in applications such as:

- **Setting hand-picked parameters in privacy-preserving protocols**: Similar to our DP application in Section 6, federated learning (FL) [63, 88] can also benefit from our approach. For example, the *model update perturbation* step involves adding random noise to the model updates (usually gradients or weights) before sending them to the central server. This includes noise scale σ and a clipping norm c parameters. These are typically hand-picked through empirical tuning or heuristics to achieve a privacy-utility tradeoff. Hence, Gaze3P can be used to implicitly set these parameters and personalise them to the users' expectations. The same applies to K-anonymity [112], L-diversity [85], T-closeness [77], to set the k, L, and T values.
- **Privacy auctions:** Privacy auctions [44, 124] are mechanisms where users "sell" their private data or privacy loss in exchange for compensation. These auctions aim to determine how much privacy loss users are willing to tolerate, and at what cost, allowing systems to personalise privacy levels across individuals based on their preferences. Users specify a subjective cost or price that they associate with a unit of privacy loss, typically through surveys. Hence, Gaze3P offers a more robust and implicit feedback with the need for users to understand and evaluate their data.
- **Synthetic data generation:** Synthetic data generation [80] refers to the creation of artificial data that mimics the statistical

properties of real user data, often used to preserve privacy while enabling data analysis or model training. Here, again Gaze3P can be used to match the user privacy expectations with respect to the data, e.g. data fidelity preferences (acceptable levels of distortion or allowed deviation from real data distributions).

- **Access control models:**Access control models regulate who can access what data under which conditions. Two key types are: (i) Attribute-Based Access Control (ABAC) [57], which dynamically grants or denies access based on user, resource, and environmental attributes (e.g., role, location, time). The stimuli-specific insights of Gaze3P, can be used to automatically infer these attributes (e.g. *the private attribute recognition task*) and set the corresponding privacy parameters (e.g. access) without explicitly defining and listing all possible roles (e.g. *the privacy expertise* or the *user profiling* tasks). (ii) Human-in-the-loop privacy controls, which empower users to manage privacy interactively in real-time systems. In this case, Gaze3Pcan be integrated with, for example, Instagram's 'Restrict' feature [104] or e-mail spam filters to allow users to flag content without explicitly selecting buttons and going through lists.

UPP applications. User-specific insights can also be integrated into several real-world applications such as:

- **Privacy nudges:** Privacy nudges [59] are subtle interventions designed to guide users toward making more privacy-conscious decisions without restricting their freedom of choice. They are commonly used in social media platforms, e.g. Facebook's contextual privacy warnings [90] and Chrome browser-based security warnings [46] These nudges typically provide a uniform textual content—such as generic privacy explanations—to all users, regardless of individual differences in privacy literacy. As a result, they risk producing mismatched comprehension: users with limited privacy knowledge may struggle to understand the information (under-comprehension), while more knowledgeable users may find it redundant or oversimplified (over-comprehension), ultimately reducing the effectiveness of the intervention. Hence, they can be personalised based on a user's expertise or profile, implicitly via Gaze3P.
- **Privacy labels and transparency notices:** Similar to privacy nudges, Gaze3P can be used in privacy labels and transparency notices, e.g. Apple's App Store privacy labels [7], to help users understand and control their privacy choices according to their profiles to avoid mismatched comprehension.
- **Cohort-based recommendations:** Cohort-based recommendations, e.g. Google's Federated Learning of Cohorts (FLoC) [45], group users into segments (or cohorts) based on shared characteristics, such as behaviour, preferences, or demographics, and generate recommendations tailored to each group. Instead of personalising for individuals, the system provides suggestions optimised for the typical member of a cohort, balancing personalisation and privacy by avoiding the need for fine-grained individual profiling. Here, again, Gaze3P can be used to profile users according to their gaze behaviour such as privacy preferences.

D Cognitive Theories and Privacy

Here, we give more details about some cognitive theories that further support our mapping functions.

The Exponential Mapping. captures a perception that escalates rapidly with small increases in privacy risk or sensitivity. The related theories include:
(i) The prospect theory [20, 62] where people weigh potential losses and gains and can provide a direct assessment of privacy levels based on their perception of risks and benefits. Individuals may perceive increasing risk with diminishing marginal tolerance—well captured by an exponential curve. The exponential shape models the non-linear, often risk-averse valuation of privacy losses.
(ii) The risk-reward trade-off theory [103] where individuals balance risks and rewards in a unified decision, leading to a cumulative rating of privacy. This model implies that perception accumulates as users weigh these aspects, with increasingly steep aversion to risk—supporting an exponential model for mapping ratings to privacy budgets.
(iii) The communication privacy management (CPM) theory [105] frames privacy as the control of boundary permeability based on accumulated context and sensitivity. The control intensifies sharply as users assess perceived violations, again suggesting an exponential increase in perceived privacy sensitivity.
(iv) The theory of planned behaviour [111] where attitudes, norms, and perceived control influence a single privacy decision, often resulting in a direct rating. The interaction of these variables can collectively lead to a compounded privacy concern that builds up non-linearly, fitting an exponential growth in privacy valuation.

The Sequential Mapping. corresponds to decision-making that unfolds in discrete steps. The following theories underpin this logic:
(i) Contextual integrity [96] sequentially considers factors like actors, attributes, and transmission principles. For example, a user might first decide if a context violates norms, then determine the severity of the violation, resulting in a layered decision path consistent with step-wise or rule-based mappings.
(ii) The privacy paradox [70] suggests that users might decide in one step whether to share information and then, based on cognitive dissonance, adjust how much information they disclose or rate its privacy (i.e. a post-hoc justification). This aligns with a sequential structure where decisions are refined over time.
(iii) The dual-process theory [3] assumes that automatic, heuristic-driven decisions (e.g., 'private or not') may precede more deliberative evaluations of privacy levels.
(iv) The heuristic-systematic model [43] suggests that a heuristic (i.e., quick judgment) may guide the first decision, followed by a deeper systematic analysis to refine privacy preferences.

The Sigmoid Mapping. reflects bounded sensitivity at both extremes: users are easily decide on very safe or very private data, but become highly sensitive in an intermediate uncertainty zone. The following theories justify this mapping:
(i) The bounded rationality theory [115] suggests that decisions are made with limited information, often leading to "good enough" rather than systematically cumulative or sequential outcomes. Users might skip steps or make fragmented decisions. The sigmoid captures this minimal sensitivity at low and high certainty, with steep

reactivity in ambiguous cases.
(ii) The uncertainty reduction theory [68] suggests that decisions aim to reduce uncertainty and may involve multiple rounds of information gathering and refinement.
(iii) The cognitive dissonance theory [51] suggests that users adjust decisions retroactively to reduce dissonance. This retroactive calibration results in smooth but non-linear adjustments over time—reflected in the sigmoid's gentle asymptotes and steep central slope.

E Adverserial Perspective on Differential Privacy

We want to briefly motivate the definition of the adversarial advantage for the differential private mechanism we use in Section 2 and Section 6.

We use the following security game for a DP-mechanism M, an adversary $\mathcal{A}$ and a challenger C.

(1) The adversary $\mathcal{A}$ chooses two valid adjacent inputs sets D_0, D_1 for M and sends them to the challenger.
(2) The challenger samples a bit b. It runs $M(D_b)$ a random number of times and stores the outputs in a set S.
(3) Upon receiving S, $\mathcal{A}$ outputs a bit b'.

The adversary wins the security game if $b' = b$. The advantage of $\mathcal{A}$ is defined as $\mathrm{adv}_{\mathcal{A}} \coloneqq 2p - 1$, where p is the maximal probability that $\mathcal{A}$ wins for *any* S. Note that the adversary in this game is exceptionally strong since it only needs to win the game for one specific output set S (no matter how unlikely S itself is). The setup is nevertheless relevant, since in real-world use cases, unlikely outputs might nevertheless occur, and even then, the privacy of the input data should be preserved.

In a slight abuse of notation, we also denote by $\mathrm{adv}_{\mathcal{A}}$ the maximal advantage achieved by a ppt. adversary $\mathcal{A}$. We want to determine an upper bound on the advantage. Given the limited information available the most successful adversary $\mathcal{A}$ outputs $b' = 0$ if $\Pr(b = 0|S) \coloneqq \Pr(D_0|M(D_0) \in S) \geq \Pr(D_1|M(D_1) \in S) \eqqcolon \Pr(b = 1|S)$ and $b' = 1$ otherwise. Let w.l.o.g. $\Pr(b = 0|S) \geq \Pr(b = 1|S)$. Hence, this adversary has an advantage $\mathrm{adv}_{\mathcal{A}} = 2\Pr(b = 0|S) - 1$. If the mechanism satisfies ε-differential privacy, Eq. (1) implies $\Pr(S|b = 0) \leq e^{\varepsilon} \Pr(S|b = 1)$. But $\Pr(S|b = i) = \frac{\Pr(S, b=i)}{\Pr(b=i)} = 2\Pr(S, b = i) = 2\Pr(S)\Pr(b = i|S)$ for $i = 0, 1$. Thus $\Pr(S|b = 0) \leq e^{\varepsilon} \Pr(S|b = 1) \Rightarrow \Pr(b = 0|S) \leq e^{\varepsilon} \Pr(b = 1|S) = e^{\varepsilon}(1 - \Pr(b = 0|S)) \Rightarrow \Pr(b = 0|S) \leq \frac{e^{\varepsilon}}{1+e^{\varepsilon}}$. We conclude that $\mathrm{adv}_{\mathcal{A}} = 2\Pr(0|S) - 1 \leq \frac{e^{\varepsilon}}{1+e^{\varepsilon}} - 1 = \frac{e^{\varepsilon}-1}{1+e^{\varepsilon}}$.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009