Information-Theoretic Fairness with A Bounded Statistical Parity Constraint

Amirreza Zamani, Abolfazl Changizi, Ragnar Thobaben, Mikael Skoglund
Division of Information Science and Engineering, KTH Royal Institute of Technology
Email: amizam@kth.se, changizi@kth.se, ragnart@kth.se, skoglund@kth.se

***Abstract*—In this paper, we study an information-theoretic problem of designing a fair representation that attains bounded statistical (demographic) parity. More specifically, an agent uses some useful data X to solve a task T. Since both X and T are correlated with some sensitive attribute or secret S, the agent designs a representation Y that satisfies a bounded statistical parity and/or privacy leakage constraint, that is, such that $I(Y;S) \leq \epsilon$. Here, we relax the perfect demographic (statistical) parity and consider a bounded-parity constraint. In this work, we design the representation Y that maximizes the mutual information $I(Y;T)$ about the task while satisfying a bounded compression (or encoding rate) constraint, that is, ensuring that $I(Y;X) \leq r$. Simultaneously, Y satisfies the bounded statistical parity constraint $I(Y;S) \leq \epsilon$.**

To design Y, we use extended versions of the Functional Representation Lemma and the Strong Functional Representation Lemma which are based on randomization techniques and study the tightness of the obtained bounds in special cases. The main idea to derive the lower bounds is to use randomization over useful data X or sensitive data S. Considering perfect demographic parity, i.e., $\epsilon = 0$, we improve the existing results (lower bounds) by using a tighter version of the Strong Functional Representation Lemma and propose new upper bounds. We then propose upper and lower bounds for the main problem and show that allowing non-zero leakage can improve the attained utility. Finally, we study the bounds and compare them in a numerical example.

The problem studied in this paper can also be interpreted as one of code design with bounded leakage and bounded rate privacy considering the sensitive attribute as a secret.

I. Introduction

In this paper, we consider the scenario illustrated in Figure 1, where we want to use some observable useful data X to draw inferences or make some decision about a certain task T. Here, we assume that both the useful data and the task are arbitrarily correlated with some sensitive attribute or secret S. For example, the data X could represent a person's resume, the task T could determine whether this person should be employed in a position or not, and the sensitive attribute S could correspond to the person's gender or disabilities.

As outlined in [1], [2], it is important to ensure that decisions are not unfair and/or that inferences do not violate privacy leakage constraints. To this end, we can design a *private* or *fair* representation Y of the data, that is, representations that contain the maximum possible information about the task T while leaking no more than an acceptable threshold about the sensitive attribute or secret S [1]–[18]. As discussed in [1], [2], [18], privacy and fairness problems are very similar, and their intersections can be utilized in broad areas such as classification algorithms. More specifically, in [1], the design of representations Y that contain no information about the sensitive attribute or secret S, i.e., $I(Y;S) = 0$, has been considered. In the *information-theoretic privacy* literature, the independence of Y and S is known as *perfect privacy* or *perfect secrecy* [10], [19]. In the *algorithmic fairness* literature, the independence of the algorithm's output Y and the sensitive attribute S is known as *perfect demographic (or statistical) parity* [2], [4]–[6].

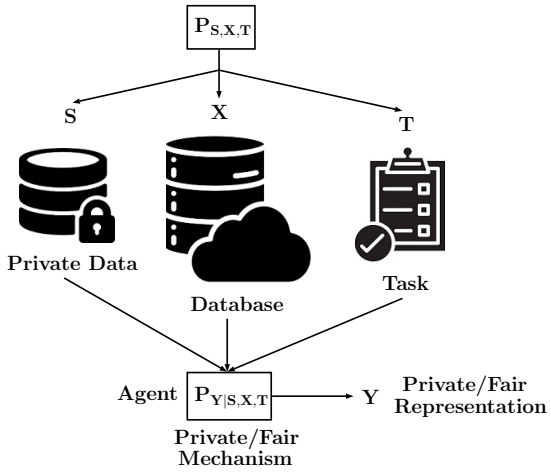

Fig. 1. Data representation with a bounded statistical (demographic) parity or privacy leakage constraint. The goal is to design a representation Y of the data X that is useful for the task T, is compressed, and leaks within a controlled threshold of the sensitive attribute or secret S.

As outlined in [9], [11], [13], [18], [20], there are many cases where perfect privacy (secrecy) or perfect demographic (statistical) parity is not achievable and we need to relax the restriction. Specifically, there are applications where perfect privacy (secrecy) leads to zero utility while relaxing it and allowing a small leakage can result in a significant gain [11]; see, for example, Example 1. Therefore, in this paper, we relax the perfect demographic parity constraint in [1], and consider a bounded statistical parity condition. This relaxation unlocks new design strategies, which lead to improved utilities and are not applicable under the perfect demographic parity or perfect privacy constraint (e.g., see Remark 6).

Under the outlined requirements on the fairness and/or privacy of the representations, we consider a problem that solves a trade-off between utility, compression or encoding rate and statistical parity: To ensure that the representation Y has as much impact as possible, we maximize the information it contains about the task T. Moreover, we impose a minimum level of compression r to the data representation, that is, $I(Y;X) \leq r$. Finally, the representation Y should satisfy the bounded statistical parity constraint $I(S;Y) \leq \epsilon$. This optimization problem is described in (3).

ISBN 978-3-903176-73-7 © 2025 IFIP

A. Prior work on fairness mechanism design

The concept of *fair representations* was introduced by Zemel et al. [6], marking a significant advancement in the field of algorithmic fairness through the expressive capabilities of deep learning. Subsequent research has been predominantly characterized by adversarial [5], [6], [21] and variational [2], [22]–[24] methodologies. The theoretical exploration of the trade-offs between utility and fairness is further elaborated in [4].

In [2], the authors study an information-theoretic fairness problem. They introduce the CFB (Conditional Fairness Bottleneck), which quantifies the trades-off between the utility, fairness, and *compression* of the representations in terms of mutual information. This additional criterion has since been used in subsequent studies [7], [24], [25]. [7], [24], [25]. More specifically, in [2], the representation design is based on a variational approach. In contrast with [2], in [1], the authors propose a simple and constructive theoretical approach to design fair representations with perfect demographic parity and/or private representations with perfect privacy. They provide upper and lower bounds on a trade-off between the utility, fairness, and compression rate of the representation. To attain the lower bounds, they use randomization techniques and extended versions of *Functional Representation Lemma* (FRL) and *Strong Functional Representation Lemma* (SFRL) derived in [9]. Furthermore, they have shown that the lower bounds can be tight under specific assumptions. In [18], a binary classification problem subject to both differential privacy and fairness constraints is studied.

B. Prior work on privacy mechanism design

In [10], the authors study the trade-off between privacy and utility in terms of mutual information. They show that the optimal privacy mechanism under perfect privacy can be obtained as the solution to a linear problem. In [11], [12], this is generalized by relaxing the perfect privacy assumption and allowing some small bounded privacy leakage. In both [11] and [12], point-wise (per-letter) measures have been used in the leakage constraint. Furthermore, the privacy design is based on the information geometry concept which allows them to study the problem geometrically. The notion of the *Privacy Funnel* is introduced in [14], where the trade-off between privacy and utility considers the log-loss as the privacy and distortion metrics.

In [17], the authors studied the fundamental limits of the privacy and utility trade-off under an estimation-theoretic approach. In [16], [26], the trade-off between privacy and utility is studied considering the equivocation as the privacy measure and the expected distortion as a utility. In [13], the authors introduce the notion of *secrecy by design* and apply it to privacy mechanism design and lossless compression problems. They find bounds on the trade-off between perfect privacy and utility using the *Functional Representation Lemma* (FRL). In [9], these results are generalized to an arbitrary privacy leakage. To attain privacy mechanism, extensions of the FRL and SFRL have been used. In [27], a compression design problem under the secrecy constraints is studied and bounds on the average length of the encoded message are derived. In [28], fundamental limits of private data disclosure are studied, where the goal is to minimize leakage under utility constraints with non-specific tasks. This result is generalized in [29]. The concept of lift is studied in [30] which represents the likelihood ratio between the posterior and prior beliefs concerning sensitive features within a data set. A privacy mechanism design in a cache-aided network has been studied in [31].

C. Contributions

In this paper, we propose a simple and constructive theoretical approach to design fair representations with bounded statistical parity and/or private representations with bounded privacy leakage. We study the problem of finding a trade-off between utility, fairness and/or privacy and compression, which is outlined earlier.

The design of the mechanism yielding the representations is based on the extensions of the FRL and the SFRL derived in [9]. As outlined in [9], the main idea for extending FRL [13] and SFRL [32] is to use a randomization technique, which is also employed in [33]. Furthermore, it is shown that both lemmas are constructive and simple, which helps in obtaining mechanism designs that can be optimal in special cases. We follow a similar approach as used in [9] and [13] to obtain simple mechanisms for fair and private representations. We emphasize that to extend the SFRL, in this paper, we use a tighter version of it which is introduced in [34]. Finally, we compare the obtained designs in different scenarios and show that they are optimal in special cases.
In more details, we can summarize our contribution as follows

- In Section II, we provide a background on the important lemmas such as the FRL, SFRL, and their extended versions. Furthermore, we provide an overview of the previous fairness problem studied in [1].
- In Section III, we introduce the fair/private representation design problem with a bounded statistical (demographic) parity. As outlined earlier, we relax the perfect demographic constraint in [1] and consider a bounded condition.
- In Section IV, we first improve the previous results in [1] by using a tighter version of the SFRL [34] and also provide new upper bounds. Next, we study the problem with a bounded constraint. We provide upper and lower bounds on the trade-off and analyze their tightness in special cases. We show that relaxing either the zero-leakage or the perfect statistical parity constraint leads to new design strategies that were previously inapplicable. Finally, in a numerical example, we compare the obtained bounds.

  The paper is concluded in Section V.

II. Preliminaries and Background

In this section, we first recall the FRL [13, Lemma 1], a tighter version of the SFRL [34, Theorem 16] and the extended version FRL [9]. We then extend the new version of SFRL using [9].

Lemma 1. *(FRL [13, Lemma 1]): For any pair of RVs (X, Y), there exists a RV U supported on $\mathcal{U}$ such that X and U are independent, i.e., we have $I(U;X) = 0$, Y is a deterministic function of (U, X), i.e., we have $H(Y|U, X) = 0$, and $|\mathcal{U}| \leq |\mathcal{X}|(|\mathcal{Y}| - 1) + 1$.*

Lemma 2. *(A Tighter SFRL [34, Theorem 16]): For any pair of RVs (X,Y), there exists a RV U supported on $\mathcal{U}$ such that X and U are independent, i.e., we have $I(U;X)=0$, Y is a deterministic function of (U,X), i.e., we have $H(Y|U,X)=0$, and $I(X;U|Y)$ can be upper bounded as follows $I(X;U|Y)\leq \log(I(X;Y)+5.51)+1.06$.*

Lemma 3. *(Extended Functional Representation Lemma (EFRL)): For a pair of RVs (X,Y) and any $0\leq\epsilon<I(X;Y)$, there exists a RV U defined on $\mathcal{U}$ such that the leakage between X and U is equal to ϵ, i.e., we have $I(U;X)=\epsilon$, Y is a deterministic function of (U,X), i.e., we have $H(Y|U,X)=0$, and $|\mathcal{U}|\leq[|\mathcal{X}|(|\mathcal{Y}|-1)+1]\,[|\mathcal{X}|+1]$.*

Next, we improve the ESFRL in [9] using Lemma 2 instead of [32, Theorem 1].

Lemma 4. *For a pair of RVs (X,Y) and any $0\leq\epsilon<I(X;Y)$, there exists a RV U defined on $\mathcal{U}$ such that the leakage between X and U is equal to ϵ, i.e., we have $I(U;X)=\epsilon$, Y is a deterministic function of (U,X), i.e., we have $H(Y|U,X)=0$, and $I(X;U|Y)$ can be upper bounded as follows $I(X;U|Y)\leq\alpha H(X|Y)+(1-\alpha)\left[\log(I(X;Y)+5.51)+1.06\right]$, where $\alpha=\frac{\epsilon}{H(X)}$.*

Proof. The proof follows similar arguments as [9, Lemma 4]. The only difference is to use Lemma 2 instead of the SFRL in [32]. ■

The main idea of constructing U, which satisfies the conditions in Lemmas 3 and 4, is to add a randomized response to the output of Lemma 1 and 2, respectively. The output refers to a random variable that is produced by the FRL or the tighter version of SFRL and satisfies the corresponding constraints. This idea has been used to find fair and private representations of a dataset in [1]. In more detail, in [1], the following problems are studied:

$$g^r(P_{S,X,T})=\sup_{\substack{P_{Y|S,X,T}:I(Y;S)=0,\\ I(X;Y)\leq r}} I(Y;T), \tag{1}$$

$$g_c^r(P_{S,X,T})=\sup_{\substack{P_{Y|S,X,T}:I(Y;S)=0,\\ I(X;Y|S,T)\leq r}} I(Y;T|S), \tag{2}$$

where X, S, T, and Y denote the dataset, sensitive attribute, target, and the designed representation, respectively. To find fair representations, we used randomization techniques similar to those in Lemma 3 and Lemma 4, which lead to lower bounds on (1) and (2). Furthermore, this randomization technique has also been applied in the design of differentially private mechanisms [35]–[38]. A similar randomization technique has been used in [39] to address private compression design problems.

III. Problem Formulation

In this section, we introduce the problem of designing a compressed representation of data with a bounded statistical (demographic) parity and/or privacy leakage constraint. Let the sensitive data S, shared data (useful data) X, and task T be discrete random variables (RVs) defined on alphabets $\mathcal{S}$ of finite cardinality $|\mathcal{S}|$, $\mathcal{X}$ of finite or countably infinite cardinality $|\mathcal{X}|$, and $\mathcal{T}$ of finite or countably infinite cardinality $|\mathcal{T}|$, respectively. The marginal probability distributions of S, X, and T are denoted by $P_S\in\mathbb{R}^{|\mathcal{S}|}$, $P_X\in\mathbb{R}^{|\mathcal{X}|}$, and $P_T\in\mathbb{R}^{|\mathcal{T}|}$, respectively. Furthermore, $P_{S,X,T}\in\mathbb{R}^{|\mathcal{S}|\times|\mathcal{X}|\times|\mathcal{T}|}$ denotes the joint distribution of S, X and T.

Let the discrete RV $Y\in\mathcal{Y}$ denote the fair representation. The goal is to design a mapping $P_{Y|S,X,T}\in\mathbb{R}^{|\mathcal{Y}|\times|\mathcal{S}|\times|\mathcal{X}|\times|\mathcal{T}|}$ that maximizes the information it keeps about the task T, while maintaining a minimum level of compression r and allowing a certain level of leakage ϵ. As we outlined earlier, in this work, we relax the fairness constraint in [1], and we allow a bounded leakage, i.e., $I(Y;S)\leq\epsilon$. Hence, the fair/private representation design problem can be stated as follows:

$$g_\epsilon^r(P_{S,X,T})=\sup_{\substack{P_{Y|S,X,T}:I(Y;S)\leq\epsilon,\\ I(X;Y)\leq r}} I(Y;T). \tag{3}$$

The condition $I(U;S)\leq\epsilon$ corresponds to the bounded statistical parity or privacy leakage constraint, and $I(X;Y)\leq r$ represents the compression rate constraint.

Remark 1. Similar to [1], we consider the case where $r<H(X)$. Otherwise, this leads to the privacy-utility trade-off problem studied in [9], [13]. Using a similar argument, we assume that $\epsilon<H(S)$, otherwise, considering X as a private attribute, we get the problem in [9], [13].

Remark 2. Similar to [1] and in contrast with [3], we assume that X, S and T are arbitrarily correlated and we ignored the Markov chain $(S,T)-X-Y$. As we mentioned in [1], an example where the Markov chain naturally holds is to let both S and T be deterministic functions of X.

Remark 3. For $\epsilon=0$, (3) leads to designing a fair representation of data with perfect demographic parity studied in [1], where upper and lower bounds have been obtained. We improve the results in [1] using a tighter SFRL in [34], i.e., Lemma 2, as presented in Theorem 1.

Remark 4. Although in this paper we assume that X, S and T are discrete RVs, all results can be extended to continuous RVs since Lemma 2 also holds for continuous RVs. For more detail, see [34, Th. 16.].

As we outlined earlier, there are many cases where zero demographic parity/privacy leakage leads to zero utility. Next, we provide an example where this is the case for perfect demographic parity, i.e., $I(U;S)=0$.

Example 1. *Let the Markov chain $(S,T)-X-Y$ hold and the leakage matrix $P_{S|X}\in\mathbb{R}^{|\mathcal{S}|\times|\mathcal{X}|}$ be invertible where the columns of $P_{S|X}$ are conditional distributions $P_{S|X=x}\in\mathbb{R}^{|\mathcal{S}|}$. In this case, using the Markov chain $S-X-Y$, we have*

$$P_{X|Y=y}=P_{S|X}^{-1}P_{S|Y=y}\overset{(a)}{=}P_{S|X}^{-1}P_S\overset{(b)}{=}P_X,$$

where (a) follows by the independence of S and Y, (b) holds since by using the Markov chain $S-X-Y$ we have $P_{S|X}P_X=P_S$. Hence, in this case, $I(S;Y)=0$ leads to $I(Y;X)=0$. Furthermore, using the Markov chain $T-X-Y$, we have

$$P_{T|Y=y}=P_{T|X}P_{X|Y=y}\overset{(a)}{=}P_{T|X}P_X=P_T,$$

where (a) follows by $I(X;Y)=0$. Thus, we conclude that $I(Y;S)=0$ leads to $I(T;Y)=0$, i.e., zero utility in (3)*.*

IV. Main Results

In this section, we first improve the bounds derived in [1] using a tighter version of SFRL in [34], i.e., Lemma 2. We then provide lower and upper bounds for the trade-off in (3), under the bounded demographic parity (i.e., non-zero leakage) and a certain encoding rate. To this end, we utilize the extended version of FRL and SFRL, i.e., Lemmas 3 and 4.

Before presenting the theorems, we first write the utility $I(Y,T)$ in $g_\epsilon^r(P_{S,X,T})$ as follows

$$\begin{aligned} I(Y;T) &= I(X,S,T;Y) - I(X,S;Y|T), \\ &= I(X,S;Y) + H(T|X,S) - H(T|Y,X,S) \\ &\quad - I(X,S;Y|T). \end{aligned} \tag{4}$$

Equation (4) helps us to find upper and lower bounds on (3). In the next result, compared to [1, Theorem 1], we provide new upper bounds and improve the lower bounds L_3^r and L_2 in [1, Theorem 1] using Lemma 2.

Theorem 1. *For every compression level $0 \le r \le H(X|S)$ and RVs $(S,X,T) \sim P_{S,X,T}$, we have that*

$$\begin{aligned} \max\{L_1^r, L_2, L_3^r\} &\le g_0^r(P_{S,X,T}) \\ &\le \min\{H(T|X)+r, H(T|S), U_0\}, \end{aligned}$$

where

$$\begin{aligned} L_1^r &= H(T|X,S) + r - H(X,S|T), \\ L_2 &= H(T|X,S) - \left(\log(I(X,S;T)+5.51)+1.06\right), \\ L_3^r &= H(T|X,S) + r - \alpha H(X,S|T) - 1.06 \\ &\quad -\log((1-\alpha)I(X,S;T)+\alpha\min\{H(T),H(X,S)\}+5.51), \\ U_0 &= H(T)+ \\ &\quad \sum_{t\in\mathcal{T}} \int_0^1 \mathbb{P}_S\{P_{T|S}(t|S)\ge m\}\log(\mathbb{P}_S\{P_{T|S}(t|S)\ge m\})dm \end{aligned}$$

and $\alpha = \frac{r}{H(X|S)}$. Moreover, for $H(X|S) \le r < H(X)$ we have that

$$\max\{L'^r_1, L_2\} \le g^r(P_{S,X,T}) \le \tag{5}$$

$$\min\{H(T|X)+r, H(T|S), U_0\}, \tag{6}$$

where $L'^r_1 = H(T|S) - H(S|T) = H(T) - H(S)$.

Proof. To obtain L_3^r and L_2, we use arguments similar to those in [1, Th. 1]. The only difference is that we use Lemma 2 instead of SFRL. To obtain the new upper bound $H(T|X)+r$, by removing the constraint $I(U;S)=0$, we obtain

$$g_0^r(P_{S,X,T}) \le \sup_{I(X;Y)\le r} I(Y;T) \overset{(a)}{\le} H(T|X)+r,$$

where (a) follows by [9, Lemma 6]. Finally, to obtain the upper bound U_0, by removing the rate constraint and using [9, Th. 4] with $Y \leftarrow T$ and $X \leftarrow S$, we have

$$g_0^r(P_{S,X,T}) \le \sup_{I(S;Y)=0} I(Y;T) \le U_0.$$

■

Corollary 1. *As argued in [1], considering the regime $H(X|S) \le r \le H(X)$, when S is a deterministic function of T, i.e., $S=f(T)$, L_1^r is tight achieving the upper bound $H(T|S)$. Furthermore, considering the regime $0 \le r \le H(X|S)$, when S is a deterministic function of X, and X a deterministic function of T, i.e., $S=f(X)$ and $X=g(T)$, L_1^r is tight achieving the new upper bound $H(T|X)+r$, which is presented in this paper. As argued in [9], there are many cases where U_0 is a tighter bound than $H(T|S)$, i.e., $U_0 \le H(T|S)$. For more detail, see [9, Example 3].*

Next, we present the results using the bounded statistical parity constraint, that is, $\epsilon \ge 0$.

Theorem 2. *For every compression level $r \ge 0$, statistical parity bound $\epsilon \ge 0$, and RVs $(S,X,T) \sim P_{S,X,T}$, the trade-off denoted as $g_\epsilon^r(P_{S,X,T})$, can be upper bounded by*

$$g_\epsilon^r(P_{S,X,T}) \le \min\left\{H(T|S)+\epsilon, H(T|X)+r\right\}. \tag{7}$$

Moreover, the following lower bounds can be attained depending on the regime of parameters:

1) $0 \le r \le H(X|S)+\epsilon$*:*

$$\max\{L_0, L_1^{r,\epsilon}, L_2^{r,\epsilon}\} \le g_\epsilon^r(P_{S,X,T}), \tag{8}$$

where

$$\begin{aligned} L_0 &= H(T|X,S) - \left(\log(I(X,S;T)+5.51)+1.06\right), \\ L_1^{r,\epsilon} &= H(T|X,S) + r - H(X,S|T), \\ L_2^{r,\epsilon} &= H(T|X,S) + r - \alpha H(X,S|T) \\ &\quad - 1.06 - \log((1-\alpha)I(X,S;T) \\ &\quad + \alpha\min\{H(T),H(X,S)\}+5.51), \end{aligned}$$

and $\alpha = \frac{r}{H(X|S)+\epsilon}$.

2) $H(X|S)+\epsilon \le r < H(X)$

$$\max\{L_0, L_1'^{r,\epsilon}\} \le g_\epsilon^r(P_{S,X,T}), \tag{9}$$

where $L_1'^{r,\epsilon} = H(T|S) - H(S|T) + \epsilon$.

3) $r \le \epsilon \le H(S|X)+r$*:*

$$\max\{L_0, L_3^{r,\epsilon}, L_4^{r,\epsilon}\} \le g_\epsilon^r(P_{S,X,T}), \tag{10}$$

where

$$\begin{aligned} L_3^{r,\epsilon} &= H(T|X,S) + \epsilon - H(X,S|T), \\ L_4^{r,\epsilon} &= H(T|X,S) + \epsilon - \tilde{\alpha} H(X,S|T) \\ &\quad - 1.06 - \log((1-\tilde{\alpha})I(X,S;T) \\ &\quad + \tilde{\alpha}\min\{H(T),H(X,S)\}+5.51), \end{aligned}$$

and $\tilde{\alpha} = \frac{\epsilon}{H(S|X)+r}$.

4) $H(S|X)+r \le \epsilon < H(S)$

$$\max\{L_0, L_3'^{r,\epsilon}\} \le g_\epsilon^r(P_{S,X,T}), \tag{11}$$

where $L_3'^{r,\epsilon} = H(T|X) - H(X|T) + r$.

Sketch of proof: Here, the main idea of the proof is provided and the complete proof is provided in Appendix A. To derive the upper bounds, we first remove either the parity or rate constraints, and then we use [9, Lemma 6]. The derivation of L_0 is similar to the proof of [1, Theorem 1] since it does not depend on r and ϵ. In other words, the lower bounds in [1, Theorem 1] can be also used for $g_\epsilon^r(P_{S,X,T})$, since we have $g_0^r(P_{S,X,T}) \le g_\epsilon^r(P_{S,X,T})$. The only difference is that

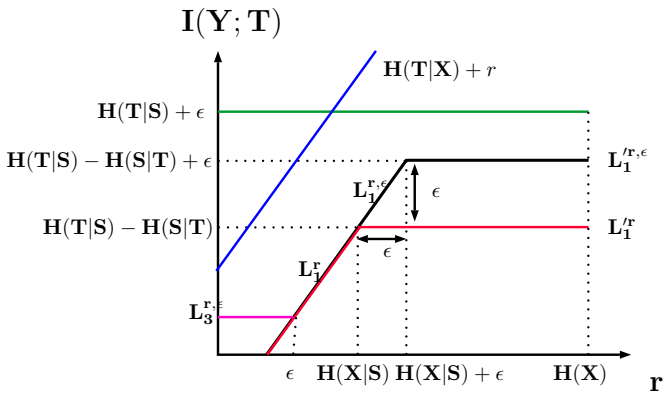

Fig. 2. Comparison of the bounds in Theorems 1 and 2.

in this paper we use Lemma 2 instead of the SFRL. To obtain $L_1^{r,\epsilon}$, we first construct U using the FRL (Lemma 1) so that $I(U;S) = \epsilon$ and $H(X|U,S) = 0$. Since such U does not necessarily satisfy the rate constraint, we randomize over U to ensure that $I(U';X) \leq r$ and $I(U';S) \leq \epsilon$. Finally, we produce Y' using Lemma 1 such that $I(Y';S,X,U') = 0$ and $H(T|Y',U',X,S) = 0$. Then we let $Y = (Y',U')$ and we show that it achieves $L_1^{r,\epsilon}$. To achieve $L_2^{r,\epsilon}$, we use same U' as before, but to produce Y' we use Lemma 2 conditionally with $X \leftarrow (S,X)|U'$ and $Y \leftarrow T$. To attain $L_3^{r,\epsilon}$, we first build U using Lemma 3 ensuring $I(U;X) = r$ and then randomize it so that U' satisfies $I(U';S) \leq \epsilon$. Finally, we build Y' similar as in achieving $L_1^{r,\epsilon}$. The achievability of $L_4^{r,\epsilon}$ is similar and the only difference is to build Y' by using Lemma 2 conditionally. To achieve $L_1'^{r,\epsilon}$ and $L_3'^{r,\epsilon}$ we use $L_1'^{r,\epsilon} = L_1^{r=H(X|S)+\epsilon,\epsilon}$ and $L_3'^{r,\epsilon} = L_1^{r,\epsilon=H(S|X)+r}$. The main reason for producing Y' in the lower bounds is to make the term $H(T|X,S,Y)$ in (4) zero, thereby improving the utility. Furthermore, the main reason to produce U in $L_1^{r,\epsilon}$ is that the term $I(X,S;Y)$ in (4) achieves its maximum which equals to $\epsilon + H(X|S)$, however, it does not necessarily satisfy the rate constraint. Hence, we randomized over U to produce U'. A similar reason holds for achieving $L_3^{r,\epsilon}$. ■

Remark 5. As shown in Fig. 2, comparing Theorems 1 and 2, we can see that relaxing the perfect demographic parity constraint leads to improved utility. For instance, $L_1^{r,\epsilon}$ improves upon L_1^r as it is attained over a larger regime. Furthermore, $L_2^{r,\epsilon}$ can improve L_3^r, as the corresponding α is smaller compared to Theorem 1. The latter follows since by ignoring the $\log(\cdot)$ terms (since they have smaller values compared to the other terms), $\alpha H(X,S|T)$ is smaller in $L_2^{r,\epsilon}$.

Remark 6. *As we outlined earlier, to achieve $L_1^{r,\epsilon}$ and $L_2^{r,\epsilon}$, we first build U using the EFRL and ESFRL, i.e., Lemmas 3 and 4, ensuring that $I(U;S) = \epsilon$. We then randomize over U such that U' satisfies $I(U';X) \leq r$. On the other hand, to attain $L_3^{r,\epsilon}$ and $L_4^{r,\epsilon}$, we first build U using Lemmas 3 and 4 ensuring $I(U;X) = r$ and then randomize it so that U' satisfies $I(U';S) \leq \epsilon$. The novelty in achieving $L_3^{r,\epsilon}$ and $L_4^{r,\epsilon}$ follows by using the relaxed statistical parity constraint $I(Y;S) \leq \epsilon$. In other words, since we are allowed to leak information about S, we can construct U to attain ϵ for the parity constraint and then update it to satisfy the rate constraint. Finally, if we enforce perfect demographic parity, we cannot use such a technique.*

In the following, we illustrate some of the bounds of Theorem 2 using an example inspired by the Experiment section provided in [27].

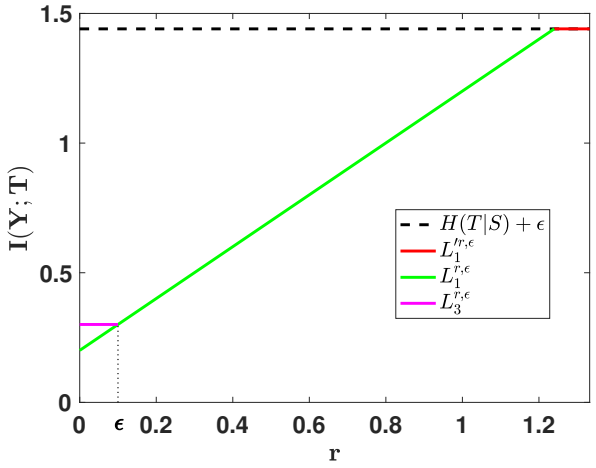

Fig. 3. Comparison of the bounds in Example 2.

Example 2. Considering [27, Experiment], let X denote the useful data defined as a "quantized histogram." Similar to [27], we find the quantized histogram by converting the images into black and white and then quantizing the histogram of any image using the following intervals: Interval 1 $= [0, 0.1)$, Interval 2 $= [0.1, 0.15)$, Interval 3 $= [0.15, 0.2)$, Interval 4 $= [0.2, 0.25)$, Interval 5 $= [0.25, 0.3)$, Interval 6 $= [0.3, 0.35)$ and Interval 7 $= [0.35, 1]$. Therefore, each image in the database is labeled with a number between 1 and 7, indicating the interval to which its histogram belongs. Let the target T be the label of each image that is a number from 0 to 9 indicating the digit which the image represents, i.e., $T \in \{0, 1, ..., 9\}$. Furthermore, let the private attribute S be a ternary RV, i.e., $S \in \{0, 1, 2\}$ and be defiend as $S = \begin{cases} 2 \text{ if } T \in \{4,5,7,9\}, \\ 1 \text{ if } T \in \{0,2,3,6,8\}, \\ 0 \text{ if } T \in \{1\} \end{cases}$. To find the joint distribution of (S,X,T), we use [27, Eq. (94)] and [27, Table. 1], substituting Z with T, Y with X and X with S. Here, we choose $\epsilon = 0.1$. In Fig. 2, the lower bounds $L_1^{r,\epsilon}$, $L_1'^{r,\epsilon}$, $L_3^{r,\epsilon}$, and the upper bound $H(T|S) + \epsilon$ are shown. Clearly, in this example S is a deterministic function of T, therefore; the lower bound $L_1'^{r,\epsilon}$, meets the upper bound $H(T|S) + \epsilon$. This can be verified in Fig. 2. Furthermore, for $r \leq \epsilon$, we can see that $L_3^{r,\epsilon}$ improves $L_1^{r,\epsilon}$.

Next, we provide an example inspired by the noisy typewriter in [40], where the lower bound $L_2^{r,\epsilon}$ improves $L_1^{r,\epsilon}$.

Example 3. Considering the noisy typewriter example in [40], let T correspond to the input alphabet and X be the output alphabet, where X is a noisy version of T. Here, let T be a uniform RV over $\{1, \ldots, 1000\}$. A typewriter produces the output by choosing the same input symbol with probability $\frac{1}{3}$ and any other character uniformly, i.e., with probability $\frac{2}{3\times 999}$ for each of the remaining characters. In

other words, we have

$$P_{X|T=t} = \begin{cases} X = t \text{ w.p. } \frac{1}{3}, \\ X = i \text{ w.p. } \frac{2}{3*999}, \ \forall i \in \{1, ..., 1000\} \backslash t, \end{cases}.$$

Furthermore, let the private attribute S be a ternary RV, i.e., $S \in \{0, 1, 2\}$, and define it as

$$S = \begin{cases} 2, & \text{if } \text{res}_{10}(T) \in \{2, \dots, 9\}, \\ 1, & \text{if } \text{res}_{10}(T) \in \{1\}, \\ 0, & \text{if } \text{res}_{10}(T) \in \{0\}, \end{cases}$$

where $\text{res}_{10}(T)$ denotes the residue of T modulo 10. Clearly, in this example, S is a deterministic function of T. Here, we choose $\epsilon = 0.1$. As shown in Fig. 4, for the regime $r \leq 2.52$, $L_2^{r,\epsilon}$ improves upon $L_1^{r,\epsilon}$. The lower bound $L_3^{r,\epsilon}$ becomes negative and is therefore omitted. Furthermore, for $r \leq 1.08$, the upper bound $H(T|X) + r$ is tighter than the upper bound $H(T|S) + \epsilon$. Finally, similar to the previous example, the lower bound $L_1'^{r,\epsilon}$ is tight since $S = f(T)$.

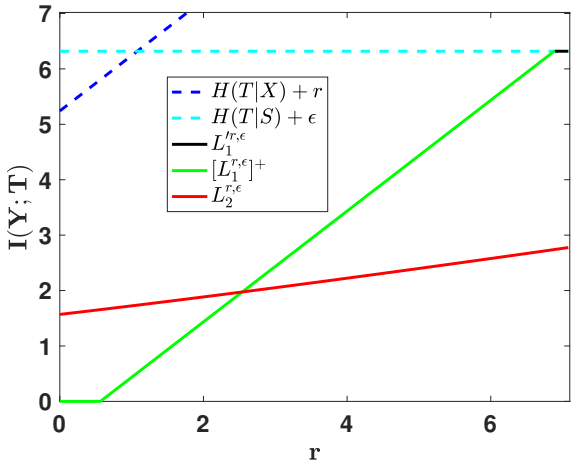

Fig. 4. Comparison of the bounds in Example 3. Here, $[L_1^{r,\epsilon}]^+$ corresponds to $\max\{0, L_1^{r,\epsilon}\}$.

V. Conclusion

In this paper, we studied the design of fair/private representations of a database under a compression rate constraint. We argued that, since perfect privacy or perfect demographic parity is not always feasible, we extended the fairness/privacy constraint to allow a bounded statistical parity/privacy leakage constraint. More precisely, we have considered a trade-off between utility $I(Y;T)$, the bounded statistical parity/privacy leakage constraint $I(Y;S) \leq \epsilon$, and the bounded compression rate condition $I(Y;X) \leq r$ where the goal is to maximize the utility under the two constraints.

To find the representations, we used randomization techniques on top of the Functional Representation Lemma and the Strong Functional Representation Lemma, which led to the lower bounds on the trade-off. This randomization approach has also been used to extend the FRL and SFRL. A benefit of using the FRL, SFRL, and their extended versions is that they are constructive and have low complexity. Therefore, the combination of this randomization technique with the lemmas leads to simple representation designs. We have shown that when the compression rate r is greater than a threshold the proposed designs can be optimal. Furthermore, we have shown that using the bounded statistical parity constraint, when the compression rate is less than ϵ, the new bounds can improve the previous bounds. Finally, we studied the bounds in a numerical example and showed that when the private data S is a deterministic function of the target T, the lower bounds meet the upper bound.

VI. Acknowledgment

This work is supported in part by the Swedish Innovation Agency through the SweWIN center.

Appendix A

Proof of Theorem 2: The upper bound is achieved by removing one constraint from the trade-off at a time while keeping the other. To this end, we first omit the rate constraint. Therefore, we have

$$g_\epsilon^r(P_{S,X,T}) \leq \sup_{P_{Y|S,X,T}: I(Y;S) \leq \epsilon} I(Y;T) \overset{(a)}{\leq} H(T|S) + \epsilon, \quad (12)$$

where (a) holds as

$$\begin{aligned} I(Y;T) &= H(T|S) + I(S;Y) - H(T|Y;S) - I(S;Y|T) \\ &\leq H(T|S) + \epsilon. \end{aligned}$$

Next, by removing the bounded statistical parity and/or privacy leakage constraint, we can write

$$g_\epsilon^r(P_{S,X,T}) \leq \sup_{P_{Y|S,X,T}: I(X;Y) \leq r} I(Y;T) \overset{(a)}{\leq} H(T|X) + r, \quad (13)$$

where (a) holds as

$$\begin{aligned} I(Y;T) &= H(T|X) + I(X;Y) - H(T|Y;X) - I(X;Y|T) \\ &\leq H(T|X) + r. \end{aligned}$$

Finally, the upper bound (7) follows from the combination of (12) and (13).

To obtain L_0, we first note that it holds in all regimes. Using Lemma 2 considering $U \leftarrow Y$, $X \leftarrow (X, S)$ and $Y \leftarrow T$, we have

$$\begin{aligned} I(T;Y) &= H(T|X,S) - I(X,S;Y|T) \\ &\geq H(T|X,S) - (\log(I(X,S;T) + 5.51) + 1.06) \\ &= L_0. \end{aligned}$$

For the lower bound $L_1^{r,\epsilon}$, we use Lemma 3 with $U \leftarrow U$, $X \leftarrow S$, and $Y \leftarrow X$ such that

$$I(S;U) = \epsilon, \quad (14)$$

$$H(X|U,S) = 0. \quad (15)$$

Moreover, we have

$$I(U;X) \leq H(X|S) + \epsilon, \quad (16)$$

where (16) holds since

$$\begin{aligned} I(U;X) &\overset{(a)}{=} I(U;S) + H(X|S) - H(X|U,S) - I(S;U|X) \\ &\overset{(b)}{\leq} \epsilon + H(X|S). \end{aligned} \quad (17)$$

where (a) is valid for any correlated X, S, and U and (b) is due to (14) and (15). We now employ the randomization technique to construct U' from U.

Let $U' = \begin{cases} U, & \text{w.p. } \alpha \\ c, & \text{w.p. } 1-\alpha \end{cases}$, where c is a constant which does not belong to $\mathcal{X} \cup \mathcal{S} \cup \mathcal{U}$ and $\alpha = \frac{r}{H(X|S)+\epsilon}$. Note that we have $0 \leq \alpha \leq 1$ in the given regime since $0 \leq r \leq H(X|S)+\epsilon$. Furthermore, assume Y' be the output of Lemma 1 with $X \leftarrow (S, X, U')$ and $Y \leftarrow T$. Hence, we can write

$$I(Y'; S, X, U') = 0, \tag{18}$$

$$H(T|Y', S, X, U') = 0. \tag{19}$$

Now, it is sufficient to let $Y = (U', Y')$. We first verify that the constraints of $I(S;Y) \leq \epsilon$ and $I(X;Y) \leq r$ are satisfied. To this end, $I(S;Y) = I(S;Y',U') \overset{(a)}{=} I(S;U') = \alpha I(S;U) = \alpha\epsilon \leq \epsilon$, where (a) follows by (18). Moreover, we have $I(X;Y) = I(X;Y',U') \overset{(a)}{=} I(X;U') = \alpha I(X;U) \overset{(b)}{\leq} \alpha(H(X|S)+\epsilon) = r$, where (a) follows from (18) and (b) follows from (16). Next, it follows from (4) that

$$\begin{aligned} I(T;Y) &= I(T;Y',U') \\ &\overset{(a)}{=} H(T|X,S)+I(Y',U';X,S)-I(X,S;Y',U'|T) \\ &\overset{(b)}{=} H(T|X,S) + I(U';X,S) - I(X,S;Y',U'|T) \\ &\geq H(T|X,S) + I(U';X,S) - H(X,S|T) \\ &\overset{(c)}{=} H(T|X,S) + r - H(X,S|T) \end{aligned}$$

where we applied (19) and (18) in (a) and (b), respectively. Furthermore, (c) follows as

$$\begin{aligned} I(U';X,S) &= I(U';S) + I(U';X|S) \\ &= \alpha(I(U;S) + I(U;X|S)) \\ &\overset{(a)}{=} \alpha(\epsilon + H(X|S)) = r, \end{aligned} \tag{20}$$

where we used (15) in (a). To derive $L_2^{r,\epsilon}$, we use the same U' as in deriving $L_1^{r,\epsilon}$, but to produce Y', we use Lemma 2 with $X \leftarrow (S,X)|U'$ and $Y \leftarrow T$. In this case, in addition to (19), the following bound for Y' holds

$$\begin{aligned} &I(Y';S,X|T,U') \\ &\quad\leq \log(I(X,S;T|U') + 5.51) + 1.06 \\ &\quad= \log((1-\alpha)I(X,S;T) + \alpha I(X,S;T|U) + 5.51) + 1.06 \\ &\quad\leq \log((1-\alpha)I(X,S;T) + \alpha \min\{H(X,S), H(T)\} + 5.51) \\ &\qquad+1.06. \end{aligned} \tag{21}$$

Furthermore, $I(X,S;Y',U'|T)$ can be bounded as

$$\begin{aligned} &I(X,S;Y',U'|T) \\ &\quad= I(X,S;U'|T) + I(X,S;Y'|T,U') \\ &\quad= \alpha I(X,S;U|T) + I(X,S;Y'|T,U') \\ &\quad\leq \alpha H(X,S|T) + \log((1-\alpha)I(X,S;T) \\ &\qquad+\alpha \min\{H(X,S), H(T)\} + 5.51) + 1.06, \end{aligned} \tag{22}$$

where (22) follows from (21). Now, using (4), (19), (20), and (22), we have $I(Y;T) \geq L_2^{r,\epsilon}$. To obtain $L_1'^{r,\epsilon}$ in the region $H(X|S)+\epsilon \leq r < H(X)$, note that

$$\begin{aligned} g_\epsilon^r(P_{S,X,T}) &\geq g_\epsilon^{H(X|S)+\epsilon}(P_{S,X,T}) \\ &\geq L_1^{H(X|S)+\epsilon,\epsilon} \\ &= H(T|X,S) + H(X|S) + \epsilon - H(X,S|T) \\ &= H(T,X|S) - H(S,X|T) + \epsilon \\ &= H(T|S) - H(S|T) + \epsilon. \end{aligned}$$

To derive the lower bound $L_3^{r,\epsilon}$, we use Lemma 3 with $U \leftarrow U$, $X \leftarrow X$ and $Y \leftarrow S$ such that

$$I(U;X) = r, \tag{23}$$

$$H(S|U,X) = 0. \tag{24}$$

Moreover, we have

$$I(U;S) \leq H(S|X) + r, \tag{25}$$

where (25) is valid since

$$\begin{aligned} I(U;S) &\overset{(a)}{=} I(U;X) + H(S|X) - H(S|U,X) - I(X;U|S) \\ &\overset{(b)}{\leq} r + H(S|X). \end{aligned} \tag{26}$$

where (a) holds for any correlated X, S, and U and (b) is due to (23) and (24). We now employ the randomization technique to construct U' from U. Let $U' = \begin{cases} U, & \text{w.p. } \tilde{\alpha} \\ \tilde{c}, & \text{w.p. } 1-\tilde{\alpha} \end{cases}$, where $\tilde{c}$ is a constant which does not belong to $\mathcal{X} \cup \mathcal{S} \cup \mathcal{U}$ and $\tilde{\alpha} = \frac{\epsilon}{H(X|S)+r}$. It is worth mentioning that we have $0 \leq \tilde{\alpha} \leq 1$ in the given regime since $r \leq \epsilon \leq H(S|X)+r$. Furthermore, assume Y' is constructed as in deriving $L_1^{r,\epsilon}$. Therefore, (18) and (19) hold. Now, it is sufficient to let $Y = (U', Y')$. First, we show the validity of the constraints $I(S;Y) \leq \epsilon$ and $I(X;Y) \leq r$. To this end, $I(S;Y) = I(S;Y',U') \overset{(a)}{=} I(S;U') = \tilde{\alpha} I(S;U) \overset{(b)}{\leq} \tilde{\alpha}(H(S|X)+r) \leq \epsilon$, where (a) and (b) follow from (18) and (26), respectively. Moreover, we have $I(X;Y) = I(X;Y',U') \overset{(a)}{=} I(X;U') = \tilde{\alpha} I(X;U) \overset{(b)}{=} \tilde{\alpha} r \leq r$, where (a) follows from (18) and (b) follows from (23). Next, it follows from (4) that

$$\begin{aligned} I(T;Y) &= I(T;Y',U') \\ &\overset{(a)}{=} H(T|X,S)+I(Y',U';X,S)-I(X,S;Y',U'|T) \\ &\overset{(b)}{=} H(T|X,S) + I(U';X,S) - I(X,S;Y',U'|T) \\ &\geq H(T|X,S) + I(U';X,S) - H(X,S|T) \\ &\overset{(c)}{=} H(T|X,S) + \epsilon - H(X,S|T) \end{aligned}$$

where we applied (19) and (18) in (a) and (b), respectively. Moreover, (c) follows as

$$\begin{aligned} I(U';X,S) &= I(U';X) + I(U';S|X) = \tilde{\alpha} I(U;X) \\ &+ \tilde{\alpha} I(U;S|X) = \tilde{\alpha}(r + H(S|X)) = \epsilon, \end{aligned} \tag{27}$$

where we used (24). To derive $L_4^{r,\epsilon}$, we use the same U' as in deriving $L_3^{r,\epsilon}$, but to produce Y', we use Lemma 2 with $X \leftarrow (S,X)|U'$ and $Y \leftarrow T$. In this case, in addition to (19), the following bound for Y' holds

$$\begin{aligned} &I(Y';S,X|T,U') \leq \log(I(X,S;T|U') + 5.51) + 1.06 \\ &\quad= \log((1-\tilde{\alpha})I(X,S;T) + \tilde{\alpha} I(X,S;T|U) + 5.51) + 1.06 \\ &\quad\leq \log((1-\tilde{\alpha})I(X,S;T) + \tilde{\alpha} \min\{H(X,S), H(T)\} + 5.51) \\ &\qquad+1.06. \end{aligned} \tag{28}$$

Furthermore, $I(X,S;Y',U'|T)$ can be bounded as

$$\begin{aligned} I(X,S;Y',U'|T) &= I(X,S;U'|T) + I(X,S;Y'|T,U') \\ &= \tilde{\alpha} I(X,S;U|T) + I(X,S;Y'|T,U') \\ &\leq \tilde{\alpha} H(X,S|T) + \log((1-\tilde{\alpha})I(X,S;T) \\ &\quad+\tilde{\alpha} \min\{H(X,S), H(T)\} + 5.51) + 1.06, \end{aligned} \tag{29}$$

where (29) follows from (28). Now, using (4), (19), (20), and (29), we have $g_\epsilon^r \geq L_4^{r,\epsilon}$. Finally, to obtain $L'^{r,\epsilon}_3$ in the region $H(S|X)+r \leq \epsilon < H(S)$, note that

$$g_\epsilon^r(P_{S,X,T}) \geq g^r_{H(S|X)+r}(P_{S,X,T}) \geq L_3^{r,H(S|X)+r}$$
$$= H(T|X,S) + H(S|X) + \epsilon - H(X,S|T) = H(T,S|X)$$
$$- H(X,S|T) + r = H(T|X) - H(X|T) + r.$$

REFERENCES

[1] A. Zamani, B. Rodríguez-Gálvez, and M. Skoglund, "On information theoretic fairness: Compressed representations with perfect demographic parity," in *2024 IEEE Information Theory Workshop (ITW)*, 2024, pp. 25–30.

[2] B. Rodríguez-Gálvez, R. Thobaben, and M. Skoglund, "A variational approach to privacy and fairness," in *2021 IEEE Information Theory Workshop (ITW)*. IEEE, 2021, pp. 1–6.

[3] S. Sreekumar and D. Gündüz, "Optimal privacy-utility trade-off under a rate constraint," in *2019 IEEE International Symposium on Information Theory*, July 2019, pp. 2159–2163.

[4] H. Zhao and G. J. Gordon, "Inherent tradeoffs in learning fair representations," *Journal of Machine Learning Research*, vol. 23, no. 57, pp. 1–26, 2022.

[5] H. Zhao, A. Coston, T. Adel, and G. J. Gordon, "Conditional learning of fair representations," in *International Conference on Learning Representations*, 2019.

[6] R. Zemel, Y. Wu, K. Swersky, T. Pitassi, and C. Dwork, "Learning fair representations," in *International conference on machine learning*. PMLR, 2013, pp. 325–333.

[7] A. Gronowski, W. Paul, F. Alajaji, B. Gharesifard, and P. Burlina, "Classification utility, fairness, and compactness via tunable information bottleneck and rényi measures," *IEEE Transactions on Information Forensics and Security*, 2023.

[8] M. Hardt, E. Price, and N. Srebro, "Equality of opportunity in supervised learning," *Advances in neural information processing systems*, vol. 29, 2016.

[9] A. Zamani, T. J. Oechtering, and M. Skoglund, "On the privacy-utility trade-off with and without direct access to the private data," *IEEE Transactions on Information Theory*, vol. 70, no. 3, pp. 2177–2200, 2024.

[10] B. Rassouli and D. Gündüz, "On perfect privacy," *IEEE Journal on Selected Areas in Information Theory*, vol. 2, no. 1, pp. 177–191, 2021.

[11] A. Zamani, T. J. Oechtering, and M. Skoglund, "A design framework for strongly χ^2-private data disclosure," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 2312–2325, 2021.

[12] A. Zamani, T. J. Oechtering, and M. Skoglund, "Data disclosure with non-zero leakage and non-invertible leakage matrix," *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 165–179, 2022.

[13] Y. Y. Shkel, R. S. Blum, and H. V. Poor, "Secrecy by design with applications to privacy and compression," *IEEE Transactions on Information Theory*, vol. 67, no. 2, pp. 824–843, 2021.

[14] A. Makhdoumi, S. Salamatian, N. Fawaz, and M. Médard, "From the information bottleneck to the privacy funnel," in *2014 IEEE Information Theory Workshop*, 2014, pp. 501–505.

[15] H. Yamamoto, "A rate-distortion problem for a communication system with a secondary decoder to be hindered," *IEEE Transactions on Information Theory*, vol. 34, no. 4, pp. 835–842, 1988.

[16] L. Sankar, S. R. Rajagopalan, and H. V. Poor, "Utility-privacy tradeoffs in databases: An information-theoretic approach," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 6, pp. 838–852, 2013.

[17] H. Wang, L. Vo, F. P. Calmon, M. Médard, K. R. Duffy, and M. Varia, "Privacy with estimation guarantees," *IEEE Transactions on Information Theory*, vol. 65, no. 12, pp. 8025–8042, Dec 2019.

[18] H. Ghoukasian and S. Asoodeh, "Differentially private fair binary classifications," in *2024 IEEE International Symposium on Information Theory (ISIT)*, 2024, pp. 611–616.

[19] Y. Y. Shkel and H. V. Poor, "A compression perspective on secrecy measures," *IEEE Journal on Selected Areas in Information Theory*, vol. 2, no. 1, pp. 163–176, 2021.

[20] C. E. Shannon, "Communication theory of secrecy systems," *The Bell system technical journal*, vol. 28, no. 4, pp. 656–715, 1949.

[21] H. Edwards and A. Storkey, "Censoring representations with an adversary," in *International Conference on Learning Representations (ICLR)*, 2016, pp. 1–14.

[22] E. Creager, D. Madras, J.-H. Jacobsen, M. Weis, K. Swersky, T. Pitassi, and R. Zemel, "Flexibly fair representation learning by disentanglement," in *International conference on machine learning*. PMLR, 2019, pp. 1436–1445.

[23] C. Louizos, K. Swersky, Y. Li, M. Welling, and R. Zemel, "The variational fair autoencoder," *International Conference on Learning Representations (ICLR)*, 2016.

[24] U. Gupta, A. M. Ferber, B. Dilkina, and G. Ver Steeg, "Controllable guarantees for fair outcomes via contrastive information estimation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 9, 2021, pp. 7610–7619.

[25] J. M. de Freitas and B. C. Geiger, "Funck: Information funnels and bottlenecks for invariant representation learning," *arXiv preprint arXiv:2211.01446*, 2022.

[26] H. Yamamoto, "A source coding problem for sources with additional outputs to keep secret from the receiver or wiretappers (corresp.)," *IEEE Transactions on Information Theory*, vol. 29, no. 6, pp. 918–923, 1983.

[27] A. Zamani and M. Skoglund, "Variable-length coding with zero and non-zero privacy leakage," *Entropy*, vol. 27, no. 2, p. 124, 2025.

[28] T.-Y. Liu and I.-H. Wang, "Privacy-utility tradeoff with nonspecific tasks: Robust privatization and minimum leakage," in *2020 IEEE Information Theory Workshop (ITW)*, 2021, pp. 1–5.

[29] A. Zamani, T. J. Oechtering, and M. Skoglund, "Multi-user privacy mechanism design with non-zero leakage," in *2023 IEEE Information Theory Workshop (ITW)*. IEEE, 2023, pp. 401–405.

[30] M. A. Zarrabian, N. Ding, and P. Sadeghi, "On the lift, related privacy measures, and applications to privacy–utility trade-offs," *Entropy*, vol. 25, no. 4, p. 679, 2023.

[31] A. Zamani, T. J. Oechtering, D. Gündüz, and M. Skoglund, "Cache-aided private variable-length coding with zero and non-zero leakage," in *2023 21st International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, 2023, pp. 247–254.

[32] C. T. Li and A. El Gamal, "Strong functional representation lemma and applications to coding theorems," *IEEE Transactions on Information Theory*, vol. 64, no. 11, pp. 6967–6978, 2018.

[33] S. L. Warner, "Randomized response: A survey technique for eliminating evasive answer bias," *Journal of the American Statistical Association*, vol. 60, no. 309, pp. 63–69, 1965.

[34] C. T. Li, "Discrete layered entropy, conditional compression and a tighter strong functional representation lemma," 2025. [Online]. Available: https://arxiv.org/abs/2501.13736

[35] Ú. Erlingsson, V. Pihur, and A. Korolova, "Rappor: Randomized aggregatable privacy-preserving ordinal response," in *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*, 2014, pp. 1054–1067.

[36] S. Asoodeh, J. Liao, F. P. Calmon, O. Kosut, and L. Sankar, "Three variants of differential privacy: Lossless conversion and applications," *IEEE Journal on Selected Areas in Information Theory*, vol. 2, no. 1, pp. 208–222, 2021.

[37] W. Alghamdi, J. F. Gomez, S. Asoodeh, F. Calmon, O. Kosut, and L. Sankar, "The saddle-point method in differential privacy," in *International Conference on Machine Learning*. PMLR, 2023, pp. 508–528.

[38] C. Dwork, G. N. Rothblum, and S. Vadhan, "Boosting and differential privacy," in *2010 IEEE 51st annual symposium on foundations of computer science*. IEEE, 2010, pp. 51–60.

[39] A. Zamani, T. J. Oechtering, and M. Skoglund, "Private variable-length coding with non-zero leakage," in *2023 IEEE International Workshop on Information Forensics and Security (WIFS)*. IEEE, 2023, pp. 1–6.

[40] T. M. Cover, *Elements of information theory*. John Wiley & Sons, 1999.