NeurIPS24:
FairQueue: Rethinking Prompt Learning for Fair Text-to-Image Generation

Christopher T. H. Teo; Milad Abdollahzadeh; Xinda Ma; Ngai-Man Cheung*

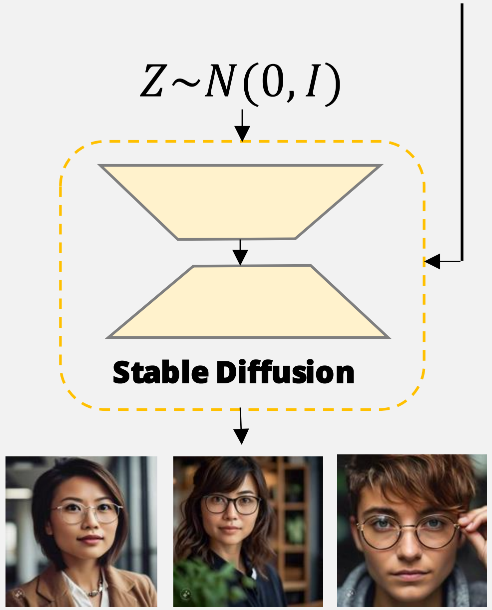
Preliminaries

Fair Text-to-Image Generation

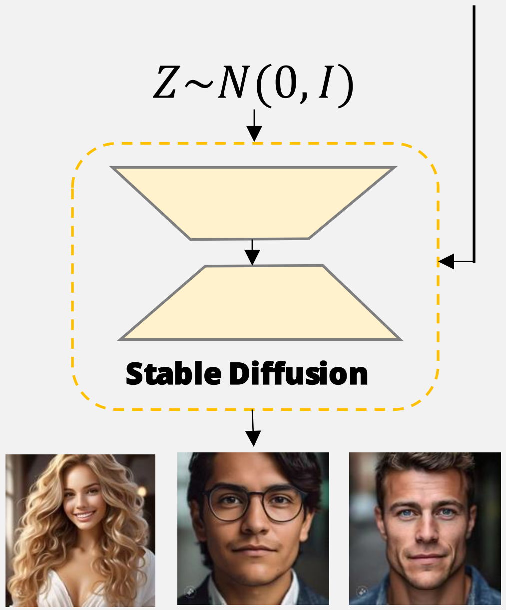
Hard-Prompt Approach[1]

Generate samples with the base prompt ($\boldsymbol{T}$) concatenate with the target sensitive attribute specific inclusive hard prompt ($\boldsymbol{F}$)

Concat ("A headshot of a person" , "with Eyeglasses")



Concat ("A headshot of a person" , "without Eyeglasses")



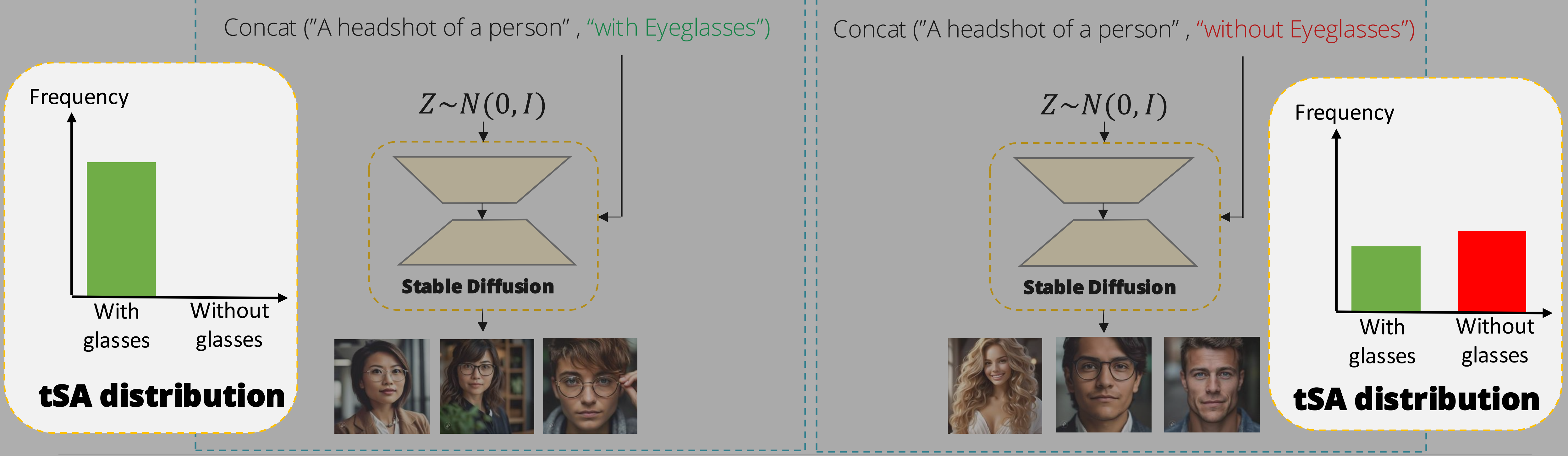
1. Hritik Bansal et al. "How well can text-to-image generative models understand ethical natural language interventions?" In: arXiv preprint arXiv:2210.15230 (2022).
2. Cheng Zhang et al. "ITI-GEN: Inclusive Text-to-Image Generation". In: Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023, pp. 3969–3980.

Preliminaries

Fair Text-to-Image Generation

Problem with Hard-Prompt Approach[2]

- Many target sensitive attribute (tSA) are linguistically ambiguous – have misleading or deceptive language – resulting in poor performance.
- However, there exist some tSA that generate both high-quality and class specific samples, which we coin to have Minimal Linguistic Ambiguity (MLA)



1. Hritik Bansal et al. "How well can text-to-image generative models understand ethical natural language interventions?" In: arXiv preprint arXiv:2210.15230 (2022).
2. Cheng Zhang et al. "ITI-GEN: Inclusive Text-to-Image Generation". In: Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023, pp. 3969–3980.

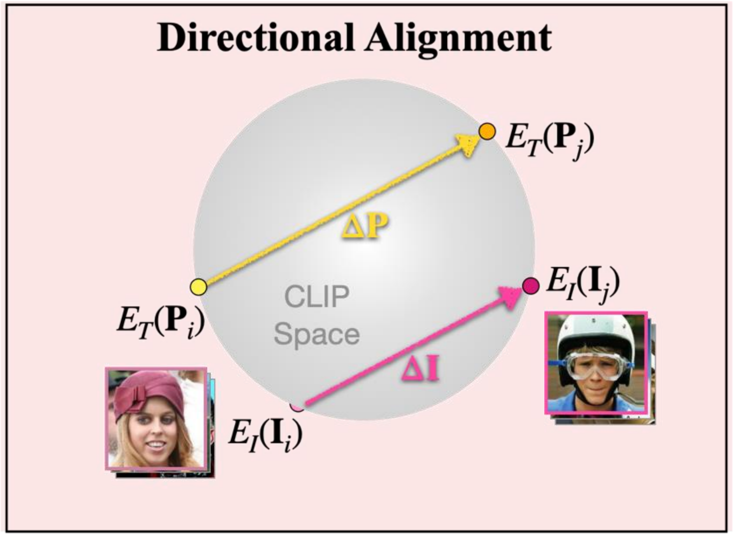
Preliminaries

Fair Text-to-Image Generation - Prompt Learning Approach (ITI-Gen[1])

Setup

- We assume the availability of a **small reference dataset** with samples pertaining to the target sensitive attribute e.g., Smiling.
- Given a base prompt ($\boldsymbol{T}$) and some learnable tokens ($\boldsymbol{S}$), ITI-Gen first forms a new inclusive prompt $\boldsymbol{P_j} = \{\boldsymbol{T}; \boldsymbol{S^j}\}$, where j are different categories for a target sensitive attribute.

Training Objectives



Then during training, ITI-Gen tries to align the directional vector of the trainable inclusive prompt embedding ΔP with the direction vector of a small reference dataset pertaining to the target Sensitive attribute, ΔI

$$\min_{\boldsymbol{S}^i, \boldsymbol{S}^j} \mathcal{L}_{dir} = 1 - \frac{(\Delta \boldsymbol{I}_{(i,j)} \cdot \Delta \boldsymbol{P}_{(i,j)})}{(|\Delta \boldsymbol{I}_{(i,j)}||\Delta \boldsymbol{P}_{(i,j)}|)}$$

1. Cheng Zhang et al. “ITI-GEN: Inclusive Text-to-Image Generation”. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023, pp. 3969–3980.

Motivation

Analyzing the Performance of ITI-Gen[1]

Setup

- *As a baseline,* We first identify target sensitive attributes with minimum linguistic ambiguity (MLA) i.e., can be accurately generated with the hard prompt approach.
    - E.g., Target sensitive attribute={*High Cheekbones, Smiling*} attained a 98% accuracy of generating the target attribute
- Then utilizing the same 500 noise input sampled from $Z_i \sim N(0, I)$ we generate samples with:
    1. Base Prompt $\boldsymbol{T}$ = "A headshot of a person"
    2. Hard Prompt $\boldsymbol{F}$ = "A headshot of a person <*Sensitive attribute*>"
    3. ITI-Gen Prompt $\boldsymbol{P}$

Performance Metrics

1. Fairness Discrepancy Metric (FD ↓) – measures fairness
2. Fréchet Inception Distance (FID ↓) and Text-Alignment (TA ↑) – measures quality
3. DreamSIM (DS ↓) – Measures the preservation of non-target sensitive attributes

1. Cheng Zhang et al. "*ITI-GEN: Inclusive Text-to-Image Generation*". In: Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023, pp. 3969–3980.

Motivation

Analyzing the Performance of ITI-Gen[1]

Results:

	FD(↓)	TA(↑)	FID(↓)	DS(↓)
BP	—	—	—	—
HP (Smiling)	**$8.4e^{-3}$**	<u>0.672</u>	**79.82**	<u>0.323</u>
ITI-GEN (Smiling)	0.124	0.605	88.63	0.557
HP (High Cheekbones)	**$3.68e^{-3}$**	<u>0.672</u>	<u>80.32</u>	<u>0.332</u>
ITI-GEN (High Cheekbones)	0.318	0.595	86.42	0.538

1. Cheng Zhang et al. "ITI-GEN: Inclusive Text-to-Image Generation". In: Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023, pp. 3969–3980.

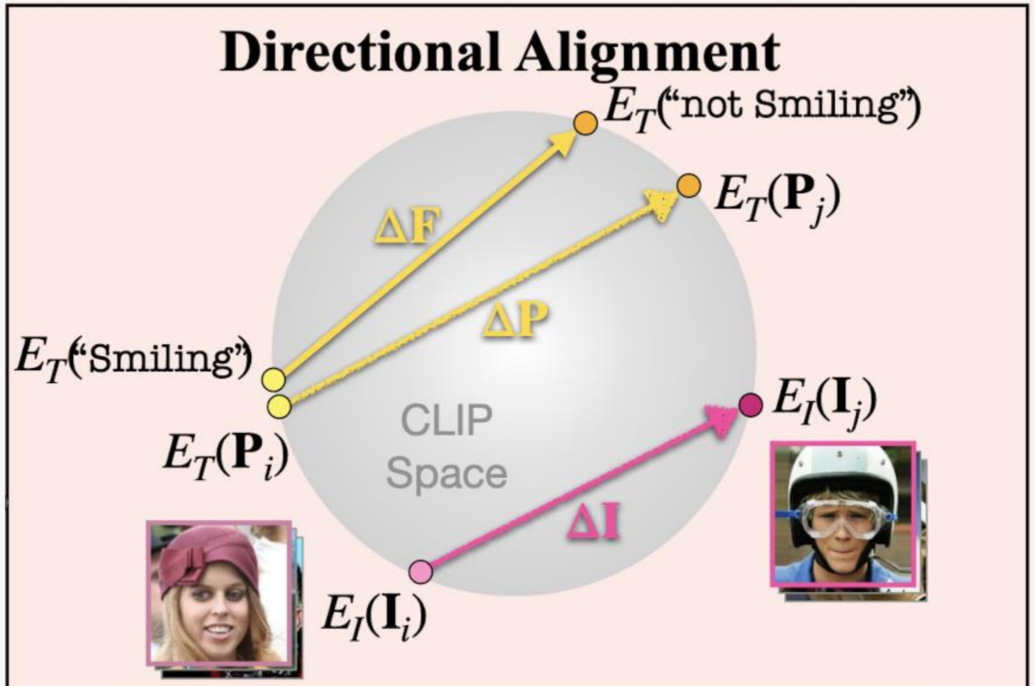
Analysis

Language Model: Analyzing Directional Loss as an Objective Function

Closer Inspection on Directional loss:

- Near perfect alignment between ITI-Gen prompt and reference image, $\mathcal{L}_{dir}(\Delta \boldsymbol{I}, \Delta \boldsymbol{P}) \approx \mathbf{0}$
- Misalignment between hard prompt and reference image, $\mathcal{L}_{dir}(\Delta \boldsymbol{I}, \Delta \boldsymbol{F}) \not\approx \mathbf{0}$

Indicating that Directional loss objective may be sub-optimal resulting in distorted tokens



tSA	$\mathcal{L}_{dir}(\Delta I, \Delta P)$	$\mathcal{L}_{dir}(\Delta I, \Delta F)$
Male	0.06	0.29
Young	0.04	0.47
Smiling	0.06	0.38
High Cheekbones	0.05	0.48

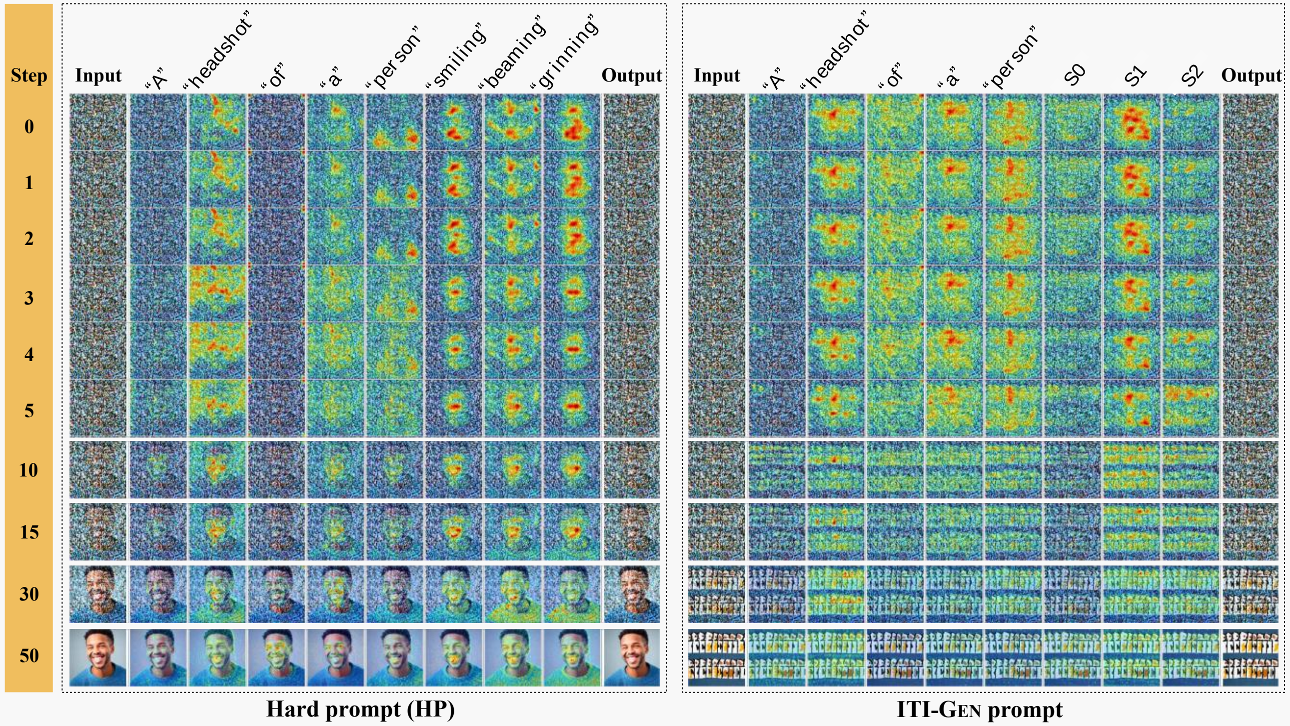
Analysis

Generation Model: Analyzing the Impact of Distorted tokens

Approach

We analyze the cross-attention maps (M) with DAAM[1] for:

- Each respective tokens e.g., "A" , "Headshot" , "of" ….
- At each diffusion steps , $t \in [0,50]$



1. Tang, Raphael, et al. "What the daam: Interpreting stable diffusion using cross attention." *arXiv preprint arXiv:2210.04885* (2022).

Analysis

Generation Model: Analyzing the Impact of Distorted tokens

Prompt Switching Analysis

We dissect the diffusion process into two steps in order to isolate the problem:

- I2H: Begin diffusion with the ITI-Gen prompt, then at time step n switch to Hard-prompt
- H2I: Begin diffusion with the Hard-prompt , then at time step n switch to ITI-Gen prompt

$$\text{I2H} = \begin{cases} DM(Z_t, \boldsymbol{P}, t, s) & t \in [0, n-1] \\ DM(Z_t, \boldsymbol{F}, t, s) & t \in [n, l] \end{cases}$$

$$\text{H2I} = \begin{cases} DM(Z_t, \boldsymbol{F}, t, s) & t \in [0, n-1] \\ DM(Z_t, \boldsymbol{P}, t, s) & t \in [n, l] \end{cases}$$

Quantitative Metric

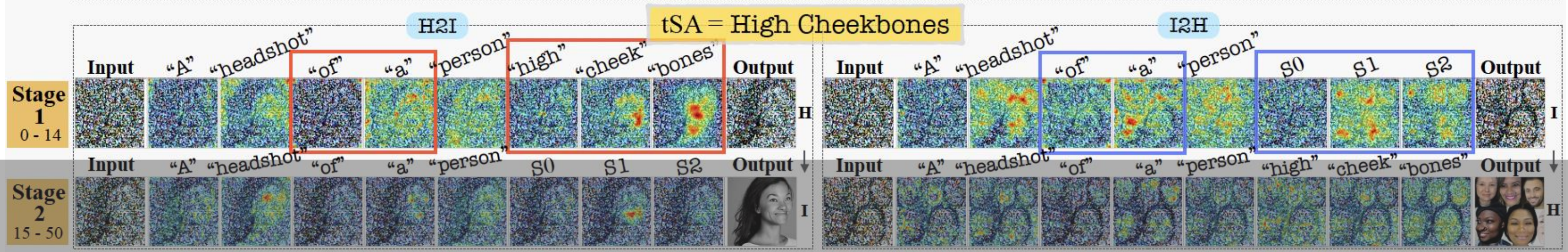
1. **Expected Attention Amplitude –** The mean attention per token, per sample i.e., $\mathrm{E}_{\{x,y\}}\{M[.]\}$
2. **Central moment –** Determines how "scattered" each attention maps is i.e.,

$$\mu(\boldsymbol{K}) = \sum_{x,y} \{[(x-\bar{x})^2 + (y-\bar{y})^2]\tilde{\boldsymbol{M}}[\boldsymbol{K}]_{(x,y)}\}$$

Analysis

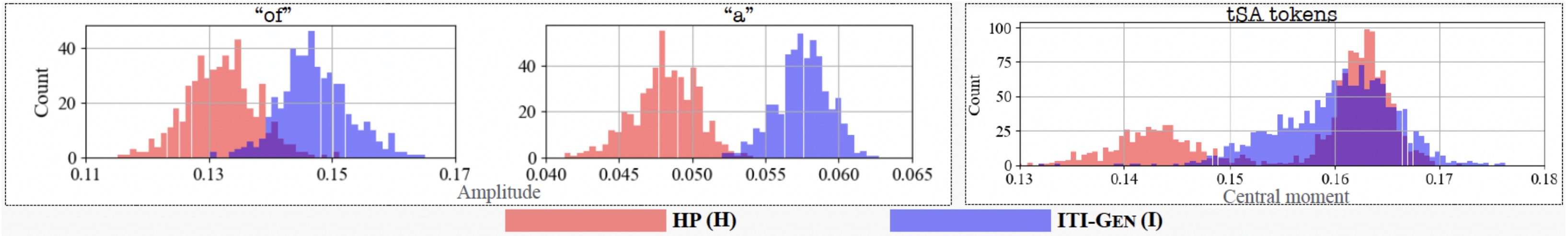
Generation Model: Analyzing the Impact of Distorted tokens

Results:



1. **I2H:** ITI-Gen tokens (S_i) affect the early stage of the diffusion process, resulting in distorted global structure
2. **H2I:** If the global structure is synthesized correctly, ITI-Gen tokens (S_i) work well to generated the target sensitive attribute

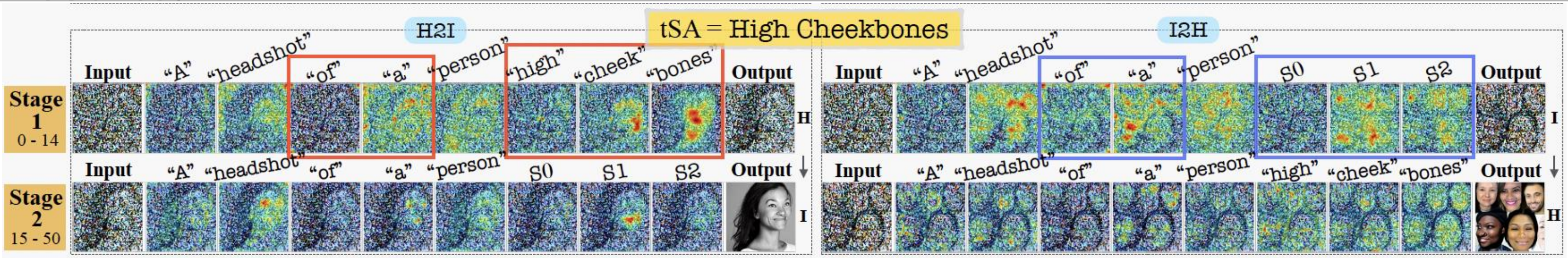
Quantitative Results (Stage 1)



Analysis

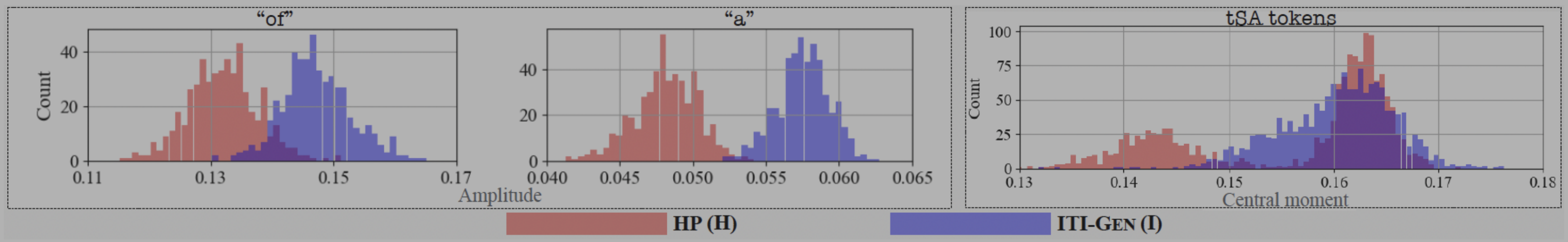
Generation Model: Analyzing the Impact of Distorted tokens

Results:



1. **I2H:** ITI-Gen tokens (S_i) affect the early stage of the diffusion process, resulting in distorted global structure
2. **H2I:** If the global structure is synthesized correctly, ITI-Gen tokens (S_i) work well to generated the target sensitive attribute

Quantitative Results (Stage 1)



Proposed Solution

FairQueue[1]

Prompt Queuing

We consider that:

- The base prompt ($\boldsymbol{T}$) generates good quality (global structure) but is not target sensitive attribute aware.
- Meanwhile, the ITI-Gen prompt can enforce the target sensitive attribute if a good global structure has already been formed.

and propose prompt queuing which "queues" the two prompts during diffusion.

$$\text{Prompt Queuing} = \begin{cases} DM(Z_t, \boldsymbol{T}, t, s) & t \in [0, n-1] \\ DM(Z_t, \boldsymbol{P}, t, s) & t \in [n, l] \end{cases}$$

Attention Amplification

Considering that prompt queueing reduces the number of diffusion steps that the generated samples is exposed to $\boldsymbol{P}$, we emphasizes the influence of the target sensitive attribute specific tokens:

$$\boldsymbol{c} * \boldsymbol{M}[S_i] \, , \boldsymbol{c} > \mathbf{1}$$

1. Christopher T. H. Teo, Milad Abdollahzadeh, Xinda Ma, and Ngai-Man Cheung. "FairQueue: Rethinking Prompt Learning for Fair Text-to-Image Generation". 37th Conference on Neural Information Processing Systems (NeurIPS), 2024

Experiments

Analyzing the Performance of FairQueue[1]

Setup

- We utilizing the same 500 noise input sampled from $Z_i \sim N(0, I)$ and generate samples with *FairQueue*[1] and compare them with the existing SOTA *ITI-Gen*
- Selected Target Sensitive Attributes (tSA) include:
  - **Single tSA (CelebA)** ∊ {Gender, Young, Smiling , … , Gray Hair}
  - **Multi tSA (CelebA)** ∊ {Gender x Young, …, Gender x Young x Eyeglasses x Smiling}
  - **Multi tSA (FairFace & Fair Benchmark)** ∊ {Gender x Age , Gender x Skin tone}
    - Age – 9 Categories
    - Skin tones - 5 Categories

1. Christopher T. H. Teo, Milad Abdollahzadeh, Xinda Ma, and Ngai-Man Cheung. "FairQueue: Rethinking Prompt Learning for Fair Text-to-Image Generation". 37th Conference on Neural Information Processing Systems (NeurIPS), 2024

Experiments

Analyzing the Performance of FairQueue[1]

Results:

FairQueue

1. Preserves the fairness (FD) from ITI-Gen
2. While improving quality (TA) and (FID)
3. Additionally, it helps to preserve the semantics of non target sensitive attribute (DS)

tSA		FD (↓)	TA (↑)	FID (↓)	DS (↓)
Single tSA (CelebA)					
Gender	ITI-Gen	$6.41e^{-3} \pm 4.2e^{-3}$	$0.655 \pm 1.2e^{-2}$	78.9 ± 1.3	$0.337 \pm 1.4e^{-2}$
	Ours	$6.41e^{-3} \pm 3.8e^{-3}$	$0.676 \pm 5.2e^{-3}$	78.3 ± 1.5	$0.308 \pm 1.2e^{-2}$
Young	ITI-Gen	$13.1e^{-3} \pm 8.1e^{-3}$	$0.653 \pm 9.4e^{-3}$	82.9 ± 1.4	$0.552 \pm 3.2e^{-2}$
	Ours	$15.5e^{-3} \pm 3.8e^{-3}$	$0.678 \pm 8.1e^{-3}$	75.3 ± 2.1	$0.370 \pm 2.7e^{-2}$
Smiling	ITI-Gen	$124e^{-3} \pm 9.2e^{-3}$	$0.605 \pm 1.2e^{-2}$	88.6 ± 0.9	$0.557 \pm 2.2e^{-2}$
	Ours	$69.0e^{-3} \pm 4.2e^{-3}$	$0.674 \pm 1.7e^{-2}$	80.0 ± 1.3	$0.284 \pm 1.0e^{-2}$
High Cheekbones	ITI-Gen	$318e^{-3} \pm 12.0e^{-3}$	$0.595 \pm 1.2e^{-3}$	86.40 ± 2.1	$0.538 \pm 1.6e^{-2}$
	Ours	$4.92e^{-3} \pm 3.6e^{-3}$	$0.685 \pm 7.2e^{-3}$	79.7 ± 2.4	$0.330 \pm 2.2e^{-2}$
Pale Skin	ITI-Gen	$1.41e^{-3} \pm 1.2e^{-3}$	$0.646 \pm 1.8e^{-2}$	101.3 ± 4.6	$0.525 \pm 2.8e^{-2}$
	Ours	$1.41e^{-3} \pm 1.2e^{-3}$	$0.666 \pm 1.9e^{-2}$	97.0 ± 3.2	$0.408 \pm 3.0e^{-2}$
Eyeglasses	ITI-Gen	$14.1e^{-3} \pm 2.6e^{-3}$	$0.654 \pm 3.3e^{-3}$	83.5 ± 1.4	$0.486 \pm 1.4e^{-2}$
	Ours	$25.4e^{-3} \pm 1.9e^{-3}$	$0.670 \pm 6.1e^{-3}$	79.4 ± 2.3	$0.391 \pm 1.6e^{-2}$
Mustache	ITI-Gen	$26.2e^{-3} \pm 1.8e^{-3}$	$0.670 \pm 4.2e^{-3}$	85.0 ± 3.3	$0.452 \pm 1.9e^{-3}$
	Ours	$22.6e^{-3} \pm 1.2e^{-3}$	$0.680 \pm 5.3e^{-3}$	80.2 ± 3.0	$0.345 \pm 3.1e^{-3}$
Chubby	ITI-Gen	$112e^{-3} \pm 8.8e^{-3}$	$0.647 \pm 2.2e^{-3}$	79.2 ± 1.5	$0.551 \pm 3.6e^{-3}$
	Ours	$119e^{-3} \pm 7.2e^{-3}$	$0.675 \pm 2.3e^{-3}$	78.3 ± 1.4	$0.387 \pm 3.0e^{-3}$
Gray Hair	ITI-Gen	$286e^{-3} \pm 6.8e^{-3}$	$0.640 \pm 4.3e^{-3}$	87.3 ± 2.1	$0.533 \pm 2.9e^{-3}$
	Ours	$266e^{-3} \pm 7.1e^{-3}$	$0.669 \pm 3.7e^{-3}$	82.2 ± 2.3	$0.417 \pm 3.1e^{-3}$
Multi tSA (CelebA)					
Gender × Young	ITI-Gen	$39.1e^{-3} \pm 1.2e^{-3}$	$0.668 \pm 7.1e^{-3}$	72.6 ± 3.1	$0.458 \pm 7.8e^{-3}$
	Ours	$12.4e^{-3} \pm 2.3e^{-3}$	$0.686 \pm 5.7e^{-3}$	71.7 ± 2.5	$0.373 \pm 4.4e^{-3}$
Gender × Young × Eyeglasses	ITI-Gen	$257e^{-3} \pm 8.7e^{-3}$	$0.654 \pm 3.3e^{-3}$	65.2 ± 1.6	$0.475 \pm 1.1e^{-3}$
	Ours	$208e^{-3} \pm 7.3e^{-3}$	$0.671 \pm 4.1e^{-3}$	61.5 ± 2.7	$0.360 \pm 6.3e^{-3}$
Gender × Young × Eyeglasses × Smiling	ITI-Gen	$190e^{-3} \pm 1.7e^{-2}$	$0.643 \pm 7.7e^{-3}$	65.5 ± 2.7	$0.475 \pm 9.1e^{-3}$
	Ours	$168e^{-3} \pm 1.0e^{-2}$	$0.661 \pm 2.4e^{-3}$	60.8 ± 1.1	$0.379 \pm 9.7e^{-3}$
Multi tSA (Fairface & Fair Benchmark)					
Gender × Age	ITI-Gen	$142e^{-3} \pm 4.2e^{-3}$	$0.659 \pm 7.2e^{-3}$	58.24 ± 3.4	$0.445 \pm 1.2e^{-3}$
	Ours	$108e^{-3} \pm 4.3e^{-3}$	$0.672 \pm 1.1e^{-3}$	58.81 ± 3.3	$0.359 \pm 3.5e^{-3}$
Gender × Skin Tone	ITI-Gen	$166e^{-3} \pm 3.7e^{-3}$	$0.670 \pm 2.2e^{-3}$	59.56 ± 3.6	$0.463 \pm 7.7e^{-3}$
	Ours	$116e^{-3} \pm 4.4e^{-3}$	$0.686 \pm 2.3e^{-3}$	54.66 ± 2.7	$0.390 \pm 1.8e^{-3}$

1. Christopher T. H. Teo, Milad Abdollahzadeh, Xinda Ma, and Ngai-Man Cheung. "FairQueue: Rethinking Prompt Learning for Fair Text-to-Image Generation". 37th Conference on Neural Information Processing Systems (NeurIPS), 2024

Experiments

Analyzing the Performance of FairQueue[1]

Results:

FairQueue

1. Preserves the fairness (FD) from ITI-Gen
2. While improving quality (TA) and (FID)
3. Additionally, it helps to preserve the semantics of non target sensitive attribute (DS)

tSA		FD (↓)	TA (↑)	FID (↓)	DS (↓)
Single tSA (CelebA)					
Gender	ITI-GEN	**$6.41e^{-3} \pm 4.2e^{-3}$**	$0.655 \pm 1.2e^{-2}$	78.9 ± 1.3	$0.337 \pm 1.4e^{-2}$
	Ours	**$6.41e^{-3} \pm 3.8e^{-3}$**	**$0.676 \pm 5.2e^{-3}$**	**78.3 ± 1.5**	**$0.308 \pm 1.2e^{-2}$**
Young	ITI-GEN	**$13.1e^{-3} \pm 8.1e^{-3}$**	$0.653 \pm 9.4e^{-3}$	82.9 ± 1.4	$0.552 \pm 3.2e^{-2}$
	Ours	$15.5e^{-3} \pm 3.8e^{-3}$	**$0.678 \pm 8.1e^{-3}$**	**75.3 ± 2.1**	**$0.370 \pm 2.7e^{-2}$**
Smiling	ITI-GEN	$124e^{-3} \pm 9.2e^{-3}$	$0.605 \pm 1.2e^{-2}$	88.6 ± 0.9	$0.557 \pm 2.2e^{-2}$
	Ours	**$69.0e^{-3} \pm 4.2e^{-3}$**	**$0.674 \pm 1.7e^{-2}$**	**80.0 ± 1.3**	**$0.284 \pm 1.0e^{-2}$**
High Cheekbones	ITI-GEN	$318e^{-3} \pm 12.0e^{-3}$	$0.595 \pm 1.2e^{-3}$	86.40 ± 2.1	$0.538 \pm 1.6e^{-2}$
	Ours	**$4.92e^{-3} \pm 3.6e^{-3}$**	**$0.685 \pm 7.2e^{-3}$**	**79.7 ± 2.4**	**$0.330 \pm 2.2e^{-2}$**
Pale Skin	ITI-GEN	**$1.41e^{-3} \pm 1.2e^{-3}$**	$0.646 \pm 1.8e^{-2}$	101.3 ± 4.6	$0.525 \pm 2.8e^{-2}$
	Ours	**$1.41e^{-3} \pm 1.2e^{-3}$**	**$0.666 \pm 1.9e^{-2}$**	**97.0 ± 3.2**	**$0.408 \pm 3.0e^{-2}$**
Eyeglasses	ITI-GEN	**$14.1e^{-3} \pm 2.6e^{-3}$**	$0.654 \pm 3.3e^{-3}$	83.5 ± 1.4	$0.486 \pm 1.4e^{-2}$
	Ours	$25.4e^{-3} \pm 1.9e^{-3}$	**$0.670 \pm 6.1e^{-3}$**	**79.4 ± 2.3**	**$0.391 \pm 1.6e^{-2}$**
Mustache	ITI-GEN	$26.2e^{-3} \pm 1.8e^{-3}$	$0.670 \pm 4.2e^{-3}$	85.0 ± 3.3	$0.452 \pm 1.9e^{-3}$
	Ours	**$22.6e^{-3} \pm 1.2e^{-3}$**	**$0.680 \pm 5.3e^{-3}$**	**80.2 ± 3.0**	**$0.345 \pm 3.1e^{-3}$**
Chubby	ITI-GEN	**$112e^{-3} \pm 8.8e^{-3}$**	$0.647 \pm 2.2e^{-3}$	79.2 ± 1.5	$0.551 \pm 3.6e^{-3}$
	Ours	$119e^{-3} \pm 7.2e^{-3}$	**$0.675 \pm 2.3e^{-3}$**	**78.3 ± 1.4**	**$0.387 \pm 3.0e^{-3}$**
Gray Hair	ITI-GEN	$286e^{-3} \pm 6.8e^{-3}$	$0.640 \pm 4.3e^{-3}$	87.3 ± 2.1	$0.533 \pm 2.9e^{-3}$
	Ours	**$266e^{-3} \pm 7.1e^{-3}$**	**$0.669 \pm 3.7e^{-3}$**	**82.2 ± 2.3**	**$0.417 \pm 3.1e^{-3}$**
Multi tSA (CelebA)					
Gender × Young	ITI-GEN	$39.1e^{-3} \pm 1.2e^{-3}$	$0.668 \pm 7.1e^{-3}$	72.6 ± 3.1	$0.458 \pm 7.8e^{-3}$
	Ours	**$12.4e^{-3} \pm 2.3e^{-3}$**	$0.686 \pm 5.7e^{-3}$	**71.7 ± 2.5**	**$0.373 \pm 4.4e^{-3}$**
Gender × Young × Eyeglasses	ITI-GEN	$257e^{-3} \pm 8.7e^{-3}$	$0.654 \pm 3.3e^{-3}$	65.2 ± 1.6	$0.475 \pm 1.1e^{-3}$
	Ours	**$208e^{-3} \pm 7.3e^{-3}$**	**$0.671 \pm 4.1e^{-3}$**	**61.5 ± 2.7**	**$0.360 \pm 6.3e^{-3}$**
Gender × Young × Eyeglasses × Smiling	ITI-GEN	$190e^{-3} \pm 1.7e^{-2}$	$0.643 \pm 7.7e^{-3}$	65.5 ± 2.7	$0.475 \pm 9.1e^{-3}$
	Ours	**$168e^{-3} \pm 1.0e^{-2}$**	**$0.661 \pm 2.4e^{-3}$**	**60.8 ± 1.1**	**$0.379 \pm 9.7e^{-3}$**
Multi tSA (Fairface & Fair Benchmark)					
Gender × Age	ITI-GEN	$142e^{-3} \pm 4.2e^{-3}$	$0.659 \pm 7.2e^{-3}$	**58.24 ± 3.4**	$0.445 \pm 1.2e^{-3}$
	Ours	**$108e^{-3} \pm 4.3e^{-3}$**	**$0.672 \pm 1.1e^{-3}$**	58.81 ± 3.3	**$0.359 \pm 3.5e^{-3}$**
Gender × Skin Tone	ITI-GEN	$166e^{-3} \pm 3.7e^{-3}$	$0.670 \pm 2.2e^{-3}$	59.56 ± 3.6	$0.463 \pm 7.7e^{-3}$
	Ours	**$116e^{-3} \pm 4.4e^{-3}$**	**$0.686 \pm 2.3e^{-3}$**	**54.66 ± 2.7**	**$0.390 \pm 1.8e^{-3}$**

1. Christopher T. H. Teo, Milad Abdollahzadeh, Xinda Ma, and Ngai-Man Cheung. "FairQueue: Rethinking Prompt Learning for Fair Text-to-Image Generation". 37th Conference on Neural Information Processing Systems (NeurIPS), 2024

Experiments

Analyzing the Performance of FairQueue[1]

Results:

FairQueue

1. Preserves the fairness (FD) from ITI-Gen
2. While improving quality (TA) and (FID)
3. Additionally, it helps to preserve the semantics of non target sensitive attribute (DS)

tSA		FD (↓)	TA (↑)	FID (↓)	DS (↓)
Single tSA (CelebA)					
Gender	ITI-Gen	$6.41e^{-3} \pm 4.2e^{-3}$	$0.655 \pm 1.2e^{-2}$	78.9 ± 1.3	$0.337 \pm 1.4e^{-2}$
	Ours	$6.41e^{-3} \pm 3.8e^{-3}$	$0.676 \pm 5.2e^{-3}$	78.3 ± 1.5	$0.308 \pm 1.2e^{-2}$
Young	ITI-Gen	$13.1e^{-3} \pm 8.1e^{-3}$	$0.653 \pm 9.4e^{-3}$	82.9 ± 1.4	$0.552 \pm 3.2e^{-2}$
	Ours	$15.5e^{-3} \pm 3.8e^{-3}$	$0.678 \pm 8.1e^{-3}$	75.3 ± 2.1	$0.370 \pm 2.7e^{-2}$
Smiling	ITI-Gen	$124e^{-3} \pm 9.2e^{-3}$	$0.605 \pm 1.2e^{-2}$	88.6 ± 0.9	$0.557 \pm 2.2e^{-2}$
	Ours	$69.0e^{-3} \pm 4.2e^{-3}$	$0.674 \pm 1.7e^{-2}$	80.0 ± 1.3	$0.284 \pm 1.0e^{-2}$
High Cheekbones	ITI-Gen	$318e^{-3} \pm 12.0e^{-3}$	$0.595 \pm 1.2e^{-3}$	86.40 ± 2.1	$0.538 \pm 1.6e^{-2}$
	Ours	$4.92e^{-3} \pm 3.6e^{-3}$	$0.685 \pm 7.2e^{-3}$	79.7 ± 2.4	$0.330 \pm 2.2e^{-2}$
Pale Skin	ITI-Gen	$1.41e^{-3} \pm 1.2e^{-3}$	$0.646 \pm 1.8e^{-2}$	101.3 ± 4.6	$0.525 \pm 2.8e^{-2}$
	Ours	$1.41e^{-3} \pm 1.2e^{-3}$	$0.666 \pm 1.9e^{-2}$	97.0 ± 3.2	$0.408 \pm 3.0e^{-2}$
Eyeglasses	ITI-Gen	$14.1e^{-3} \pm 2.6e^{-3}$	$0.654 \pm 3.3e^{-3}$	83.5 ± 1.4	$0.486 \pm 1.4e^{-2}$
	Ours	$25.4e^{-3} \pm 1.9e^{-3}$	$0.670 \pm 6.1e^{-3}$	79.4 ± 2.3	$0.391 \pm 1.6e^{-2}$
Mustache	ITI-Gen	$26.2e^{-3} \pm 1.8e^{-3}$	$0.670 \pm 4.2e^{-3}$	85.0 ± 3.3	$0.452 \pm 1.9e^{-3}$
	Ours	$22.6e^{-3} \pm 1.2e^{-3}$	$0.680 \pm 5.3e^{-3}$	80.2 ± 3.0	$0.345 \pm 3.1e^{-3}$
Chubby	ITI-Gen	$112e^{-3} \pm 8.8e^{-3}$	$0.647 \pm 2.2e^{-3}$	79.2 ± 1.5	$0.551 \pm 3.6e^{-3}$
	Ours	$119e^{-3} \pm 7.2e^{-3}$	$0.675 \pm 2.3e^{-3}$	78.3 ± 1.4	$0.387 \pm 3.0e^{-3}$
Gray Hair	ITI-Gen	$286e^{-3} \pm 6.8e^{-3}$	$0.640 \pm 4.3e^{-3}$	87.3 ± 2.1	$0.533 \pm 2.9e^{-3}$
	Ours	$266e^{-3} \pm 7.1e^{-3}$	$0.669 \pm 3.7e^{-3}$	82.2 ± 2.3	$0.417 \pm 3.1e^{-3}$
Multi tSA (CelebA)					
Gender × Young	ITI-Gen	$39.1e^{-3} \pm 1.2e^{-3}$	$0.668 \pm 7.1e^{-3}$	72.6 ± 3.1	$0.458 \pm 7.8e^{-3}$
	Ours	$12.4e^{-3} \pm 2.3e^{-3}$	$0.686 \pm 5.7e^{-3}$	71.7 ± 2.5	$0.373 \pm 4.4e^{-3}$
Gender × Young × Eyeglasses	ITI-Gen	$257e^{-3} \pm 8.7e^{-3}$	$0.654 \pm 3.3e^{-3}$	65.2 ± 1.6	$0.475 \pm 1.1e^{-3}$
	Ours	$208e^{-3} \pm 7.3e^{-3}$	$0.671 \pm 4.1e^{-3}$	61.5 ± 2.7	$0.360 \pm 6.3e^{-3}$
Gender × Young × Eyeglasses × Smiling	ITI-Gen	$190e^{-3} \pm 1.7e^{-2}$	$0.643 \pm 7.7e^{-3}$	65.5 ± 2.7	$0.475 \pm 9.1e^{-3}$
	Ours	$168e^{-3} \pm 1.0e^{-2}$	$0.661 \pm 2.4e^{-3}$	60.8 ± 1.1	$0.379 \pm 9.7e^{-3}$
Multi tSA (Fairface & Fair Benchmark)					
Gender × Age	ITI-Gen	$142e^{-3} \pm 4.2e^{-3}$	$0.659 \pm 7.2e^{-3}$	58.24 ± 3.4	$0.445 \pm 1.2e^{-3}$
	Ours	$108e^{-3} \pm 4.3e^{-3}$	$0.672 \pm 1.1e^{-3}$	58.81 ± 3.3	$0.359 \pm 3.5e^{-3}$
Gender × Skin Tone	ITI-Gen	$166e^{-3} \pm 3.7e^{-3}$	$0.670 \pm 2.2e^{-3}$	59.56 ± 3.6	$0.463 \pm 7.7e^{-3}$
	Ours	$116e^{-3} \pm 4.4e^{-3}$	$0.686 \pm 2.3e^{-3}$	54.66 ± 2.7	$0.390 \pm 1.8e^{-3}$

1. Christopher T. H. Teo, Milad Abdollahzadeh, Xinda Ma, and Ngai-Man Cheung. "FairQueue: Rethinking Prompt Learning for Fair Text-to-Image Generation". 37th Conference on Neural Information Processing Systems (NeurIPS), 2024

Experiments

Analyzing the Performance of FairQueue[1]

Results:

1. Christopher T. H. Teo, Milad Abdollahzadeh, Xinda Ma, and Ngai-Man Cheung. "FairQueue: Rethinking Prompt Learning for Fair Text-to-Image Generation". 37th Conference on Neural Information Processing Systems (NeurIPS), 2024

Thank You

For more details:

Project Page