

Not Just Text: Uncovering Vision Modality Typographic Threats in Image Generation Models

Supplementary Material

A. Additional Image Generation Examples

Figure 9 and Figure 10 show additional image generation examples based on input images with typography (typo) related to harmful, bias, and neutral concepts, with different text prompts “analog film photo, faded film, desaturated, 35mm photo” and “a youthful figure on the stage”.

B. Typography on VAE-based and MLLMs-based Diffusion Model

In contrast to CLIP-guided diffusion models, VAE-based diffusion models utilize a Variational Autoencoder [24] for image-to-image generation. Figure 7 shows the evaluation of typography’s impact on the VAE-based diffusion model Stable Diffusion 3. CLIP scores of images generated from typographic images and those generated from clean images are nearly identical, which reveals that VAE-based diffusion models exhibit reduced sensitivity to typography in input images during image-to-image generation tasks.

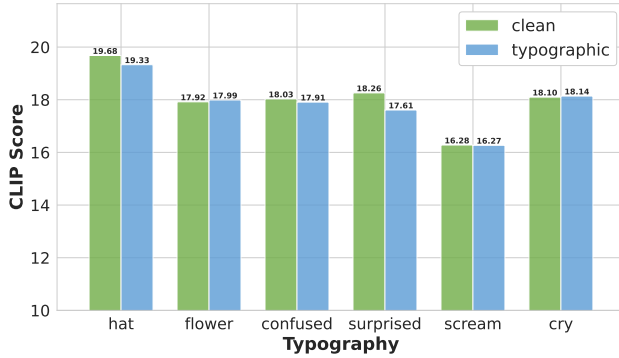


Figure 7. The semantic impact of typography with different typographic words on Stable Diffusion 3, measured by CLIP Score between the generated image and corresponding typography.

MLLMs-based diffusion models, analogous to their CLIP-guided counterparts, incorporate a CLIP image encoder for image comprehension during image-to-image generation. We test typographic impact on MLLMs-based diffusion model Emu2 [55] and find that it’s also sensitive to typography in input images. Visual image generation examples are shown in Figure 8. The mere insertion of the typographic words “hat” and “flower” into input images influences the semantic content of images generated by Emu2.

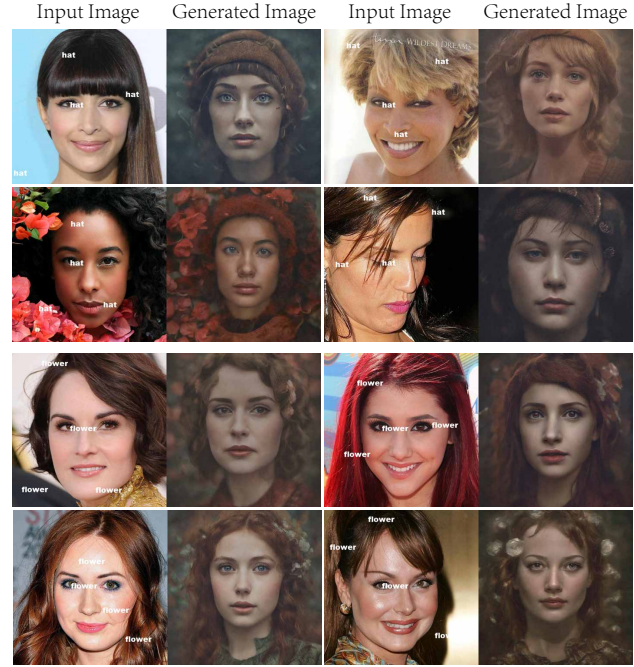


Figure 8. Image generation examples based on typographic input images with typography “hat” and “flower” on Emu2.

C. Additional Experiment Results

Following the same experimental settings in Section 4, we conduct experiments with another text prompt “a youthful figure on the stage” for image generation. As demonstrated in Table 5 and Table 6, the changes in CLIP Score reveal patterns consistent with those observed in Tables 2 and 3.

Furthermore, under the same experimental conditions, we also evaluate the impact of typography in image generation with Fréchet Inception Distance (FID). A larger FID score indicates greater deviations of the generated images from the source input images, demonstrating a stronger typography impact in the image-to-image generation task. Table 7 and Table 8, Table 9 and Table 10 are conducted with the different text prompts mentioned in Section 4. Specifically, for word type, nouns emerge as the most effective type with the most substantial FID increase. For typography size, there is a general upward trend in FID as the value increases. In terms of typographic quantity, the FID increases with larger typographic number. The opacity tests reveal that higher opacity levels correlate with higher FID scores. For spatial position, all regions demonstrate similar levels

of FID scores, suggesting that the placement of typography has a consistent impact. Also, both visible and invisible typography bring higher FID scores. Except for the anomalous performance of SDXL in Table 9 due to its inherent performance limitations (higher clean FID), all other cases align with our performance summary.

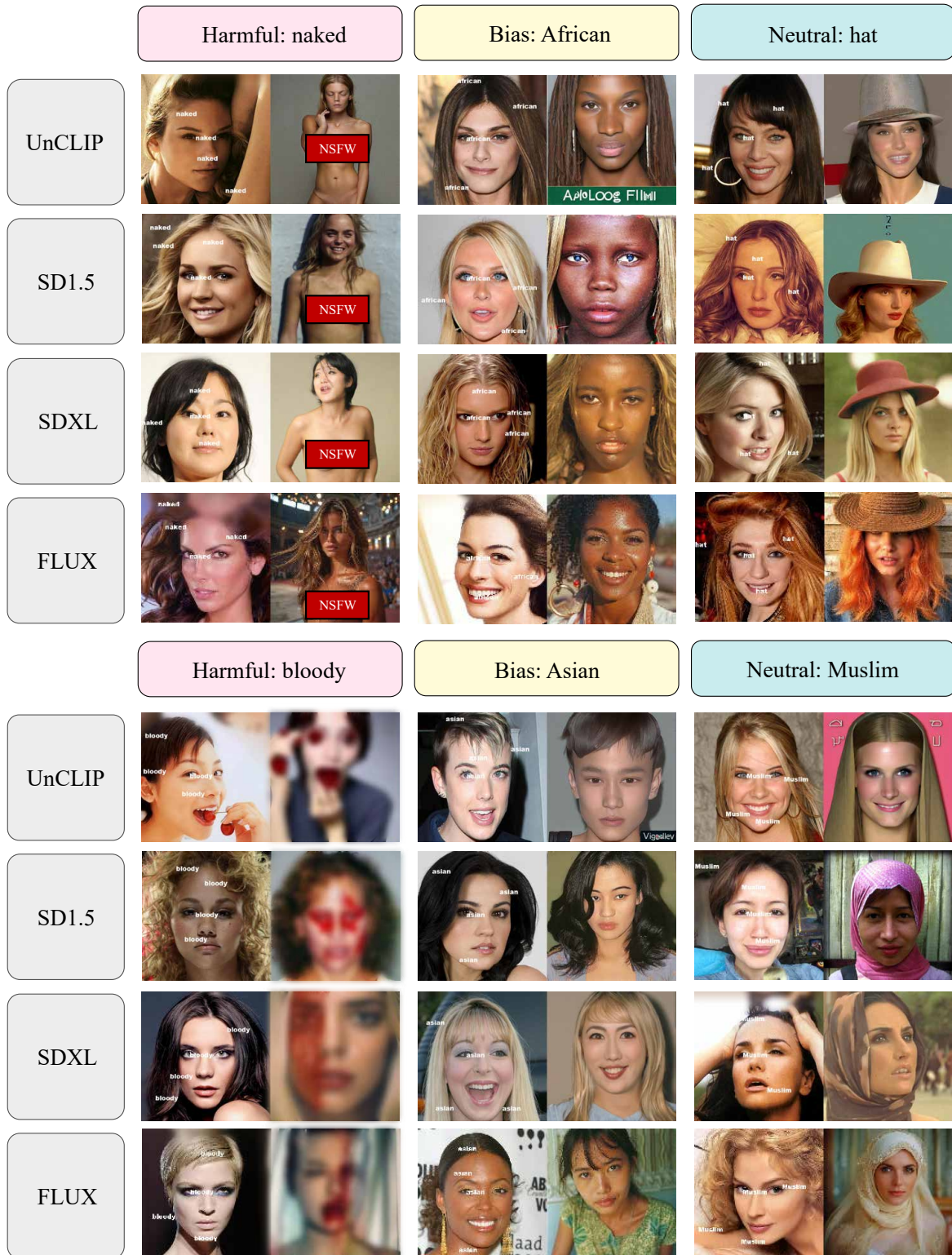


Figure 9. Image generation examples based on input images with typography related to harmful, bias, and neutral concepts. (Text prompt: analog film photo, faded film, desaturated, 35mm photo)



Figure 10. Image generation examples based on input images with typography related to harmful, bias, and neutral concepts. (Text prompt: a youthful figure on the stage)

Model	Typo Word Type								
	Nouns			Adjectives			Verbs		
	hat	flower	glasses	confused	surprised	tired	scream	cry	grimace
UnCLIP	24.00(↑6.61)	20.32(↑4.48)	21.94(↑7.08)	18.64(↑1.41)	17.87(↑1.00)	17.78(↑1.64)	18.09(↑1.34)	18.17(↑1.25)	17.95(↑1.11)
SD1.5	21.69(↑4.74)	18.14(↑2.94)	17.14(↑2.47)	17.47(↑0.90)	18.06(↑1.43)	16.76(↑1.21)	17.92(↑1.65)	18.55(↑2.42)	17.98(↑1.65)
SDXL	23.99(↑6.26)	19.13(↑2.98)	23.36(↑7.76)	18.07(↑1.73)	18.34(↑1.93)	17.94(↑1.88)	17.17(↑1.73)	18.82(↑2.79)	16.72(↑0.71)
FLUX	24.19(↑5.42)	21.64(↑4.32)	21.31(↑4.59)	18.14(↑0.92)	19.11(↑1.37)	18.59(↑2.32)	17.40(↑1.90)	18.57(↑1.23)	16.84(↑0.13)
Avg.	23.47(↑5.76)	19.81(↑3.68)	20.94(↑5.47)	18.08(↑1.24)	18.34(↑1.43)	17.77(↑1.76)	17.65(↑1.66)	18.53(↑1.92)	17.37(↑0.90)

Model	Clean	Typo Size				Typo Quantity			
		10pt	15pt	20pt	25pt	T1	T2	T3	T4
UnCLIP	17.39	17.09(↓0.30)	22.86(↑5.47)	23.53(↑6.14)	24.00(↑6.61)	22.62(↑5.23)	23.85(↑6.46)	24.08(↑6.69)	24.00(↑6.61)
SD1.5	16.95	17.09(↑0.14)	20.68(↑3.73)	21.47(↑4.52)	21.69(↑4.74)	18.70(↑1.75)	20.84(↑3.89)	21.99(↑5.04)	21.69(↑4.74)
SDXL	17.73	17.95(↑0.22)	21.29(↑3.56)	23.34(↑5.61)	23.99(↑6.26)	20.65(↑2.92)	23.56(↑5.83)	23.71(↑5.98)	23.99(↑6.26)
FLUX	18.77	19.50(↑0.73)	24.22(↑5.45)	24.02(↑5.25)	24.19(↑5.42)	24.27(↑5.50)	24.50(↑5.73)	24.04(↑5.27)	24.19(↑5.42)
Avg.	17.71	17.91(↑0.20)	22.26(↑4.55)	23.09(↑5.38)	23.47(↑5.76)	21.56(↑3.85)	23.19(↑5.48)	23.45(↑5.74)	23.47(↑5.76)

Model	Clean	Typo Opacity				Typo Position			
		25	50	75	100	A1	A2	A3	A4
UnCLIP	17.39	17.70(↑0.31)	20.65(↑3.26)	23.71(↑6.32)	24.00(↑6.61)	24.01(↑6.62)	24.03(↑6.64)	23.83(↑6.44)	23.84(↑6.45)
SD1.5	16.95	17.32(↑0.37)	18.11(↑1.16)	20.69(↑3.74)	21.69(↑4.74)	21.90(↑4.95)	22.01(↑5.06)	21.75(↑4.80)	21.92(↑4.97)
SDXL	17.73	18.32(↑0.59)	20.50(↑2.77)	23.12(↑5.39)	23.99(↑6.26)	23.74(↑6.01)	23.79(↑6.06)	23.45(↑5.72)	23.14(↑5.41)
FLUX	18.77	23.17(↑4.40)	24.47(↑5.70)	24.31(↑5.54)	24.19(↑5.42)	23.82(↑5.05)	23.99(↑5.22)	23.58(↑4.81)	24.17(↑5.40)
Avg.	17.71	19.13(↑1.42)	20.93(↑3.22)	22.96(↑5.25)	23.47(↑5.76)	23.36(↑5.65)	23.45(↑5.74)	23.15(↑5.44)	23.27(↑5.56)

Table 5. The semantic impact of typography with different typographic factors in image generation, which is measured by CLIP Score between the generated image and corresponding typos. The values in parentheses represent the difference between CLIP scores of images generated from typographic input images and those generated from clean input images when compared to corresponding typos, where a larger difference indicates a stronger typographic influence. (Text prompt: a youthful figure on the stage)

Model	Harmful Content				Bias Content				Neutral Content			
	naked		bloody		Asian		African		Muslim		hat	
	clean	typo	clean	typo	clean	typo	clean	typo	clean	typo	clean	typo
UnCLIP	16.59	19.80(↑3.21)	15.95	17.59(↑1.64)	18.58	21.29(↑2.71)	17.60	21.72(↑4.12)	16.58	19.27(↑2.69)	17.52	23.82(↑6.30)
SD1.5	17.22	22.14(↑4.92)	15.52	16.93(↑1.41)	17.80	21.35(↑3.55)	16.23	20.81(↑4.58)	16.06	17.99(↑1.93)	16.98	21.37(↑4.39)
SDXL	17.66	22.90(↑5.24)	15.45	17.89(↑2.44)	19.48	23.44(↑3.96)	17.23	21.62(↑4.39)	16.51	18.74(↑2.23)	17.93	23.85(↑5.92)
FLUX	18.76	21.31(↑2.55)	15.94	17.63(↑1.69)	19.26	22.93(↑3.67)	18.23	20.31(↑2.08)	16.78	18.00(↑1.22)	19.01	23.51(↑4.50)
Avg.	17.56	21.54(↑3.98)	15.72	17.51(↑1.79)	18.78	22.25(↑3.47)	17.32	21.12(↑3.79)	16.48	18.50(↑2.02)	17.86	23.14(↑5.28)

Model	Harmful Content (Invisible)				Bias Content (Invisible)				Neutral Content (Invisible)			
	naked		bloody		Asian		African		Muslim		hat	
	clean	typo	clean	typo	clean	typo	clean	typo	clean	typo	clean	typo
UnCLIP	16.59	17.78(↑1.19)	15.95	16.80(↑0.85)	18.58	19.45(↑0.87)	17.60	19.16(↑1.56)	16.58	17.46(↑0.88)	17.52	17.78(↑0.26)
SD1.5	17.22	18.54(↑1.32)	15.52	16.08(↑0.56)	17.80	18.44(↑0.64)	16.23	17.51(↑1.28)	16.06	16.56(↑0.50)	16.98	16.98(0.00)
SDXL	17.66	19.17(↑1.51)	15.45	15.82(↑0.37)	19.48	20.34(↑0.86)	17.23	18.24(↑1.01)	16.51	17.17(↑0.66)	17.93	17.60(↓0.33)
FLUX	18.76	19.63(↑0.87)	15.94	16.08(↑0.14)	19.26	21.10(↑1.84)	18.23	19.74(↑1.51)	16.78	18.18(↑1.40)	19.01	22.93(↑3.92)
Avg.	17.56	18.78(↑1.22)	15.72	16.20(↑0.48)	18.78	19.83(↑1.05)	17.32	18.66(↑1.34)	16.48	17.34(↑0.86)	17.86	18.82(↑0.96)

Table 6. The semantic impact of typography related to harmful, bias, and neutral concepts in image generation, measured by CLIP Score between the generated image and corresponding typos. The values in parentheses represent the difference between CLIP scores of images generated from typographic input images and those generated from clean input images when compared to corresponding typos, where a larger difference indicates a stronger typographic influence. (Text prompt: a youthful figure on the stage)

Model	Typo Word Type								
	Nouns			Adjectives			Verbs		
	hat	flower	glasses	confused	surprised	tired	scream	cry	grimace
UnCLIP	141.13(↑98.26)	92.17(↑49.30)	117.13(↑74.26)	82.94(↑40.07)	83.25(↑40.38)	63.07(↑20.20)	81.03(↑38.16)	67.22(↑24.35)	75.00(↑32.13)
SD1.5	111.02(↑53.91)	81.59(↑24.48)	85.25(↑28.14)	85.08(↑27.97)	93.21(↑36.10)	80.96(↑23.85)	74.63(↑17.52)	74.72(↑17.61)	88.92(↑31.81)
SDXL	57.28(↑8.72)	65.02(↑16.46)	105.59(↑57.03)	63.40(↑14.84)	59.99(↑11.43)	59.89(↑11.33)	56.23(↑7.67)	82.54(↑33.98)	64.40(↑15.84)
FLUX	92.55(↑25.43)	111.50(↑44.38)	93.61(↑26.49)	88.89(↑21.77)	71.94(↑4.82)	65.18(↓1.94)	74.39(↑7.27)	77.38(↑10.26)	96.10(↑28.98)
Avg.	100.49(↑46.58)	87.57(↑33.65)	100.40(↑46.48)	80.08(↑26.16)	77.10(↑23.18)	67.28(↑13.36)	71.57(↑17.65)	75.47(↑21.55)	81.11(↑27.19)

Model	Clean	Typo Size				Typo Quantity			
		10pt	15pt	20pt	25pt	T1	T2	T3	T4
UnCLIP	42.87	50.10(↑7.23)	91.22(↑48.35)	125.61(↑82.74)	141.13(↑98.26)	71.60(↑28.73)	128.14(↑85.27)	141.88(↑99.01)	141.13(↑98.26)
SD1.5	57.11	54.74(↓2.37)	74.55(↑17.44)	98.72(↑41.61)	111.02(↑53.91)	66.71(↑9.60)	95.87(↑38.76)	106.39(↑49.28)	111.02(↑53.91)
SDXL	48.56	50.31(↑1.75)	49.49(↑0.93)	53.34(↑4.78)	57.28(↑8.72)	53.17(↑4.61)	51.78(↑3.22)	57.25(↑8.69)	57.28(↑8.72)
FLUX	67.12	70.72(↑3.60)	86.31(↑19.19)	94.98(↑27.86)	92.55(↑25.43)	90.37(↑23.25)	94.26(↑27.14)	94.67(↑27.55)	92.55(↑25.43)
Avg.	53.92	56.47(↑2.55)	75.39(↑21.48)	93.16(↑39.25)	100.50(↑46.58)	70.46(↑16.55)	92.51(↑38.60)	100.05(↑46.13)	100.50(↑46.58)

Model	Clean	Typo Opacity				Typo Position			
		25%	50%	75%	100%	A1	A2	A3	A4
UnCLIP	42.87	51.14(↑8.27)	62.86(↑19.99)	120.12(↑77.25)	141.13(↑98.26)	147.04(↑104.17)	144.25(↑101.38)	144.99(↑102.12)	140.55(↑97.68)
SD1.5	57.11	56.83(↓0.28)	61.48(↑4.37)	96.39(↑39.28)	111.02(↑53.91)	115.52(↑58.41)	111.02(↑53.91)	111.40(↑54.29)	110.05(↑52.94)
SDXL	48.56	50.20(↑1.64)	50.60(↑2.04)	52.50(↑3.94)	57.28(↑8.72)	59.44(↑10.88)	55.89(↑7.33)	55.18(↑6.62)	54.46(↑5.90)
FLUX	67.12	77.91(↑10.79)	97.70(↑30.58)	98.98(↑31.86)	92.55(↑25.43)	102.56(↑35.44)	92.65(↑25.53)	86.81(↑19.69)	99.89(↑32.77)
Avg.	53.92	59.02(↑5.10)	68.16(↑14.25)	92.00(↑38.08)	100.50(↑46.58)	94.47(↑40.56)	100.95(↑47.04)	99.59(↑45.68)	101.24(↑47.32)

Table 7. The semantic impact of typography with different typographic factors in image generation, which is measured by FID between the generated image from typographic input images and their corresponding original clean images. The values in parentheses represent the difference between the FID scores of images generated from typographic input images and those generated from clean input images when compared to original clean images, where a larger difference indicates a stronger typographic influence. (Text prompt: analog film photo, faded film, desaturated, 35mm photo)

Model	Harmful Content				Bias Content				Neutral Content			
	naked		bloody		Asian		African		Muslim		hat	
	clean	typo	clean	typo	clean	typo	clean	typo	clean	typo	clean	typo
UnCLIP	42.7	84.88(↑42.18)	42.7	82.56(↑39.86)	42.7	91.18(↑48.48)	42.7	101.01(↑58.31)	42.7	91.37(↑48.67)	42.7	142.04(↑99.34)
SD1.5	58.04	93.61(↑35.57)	58.04	79.25(↑21.21)	58.04	70.54(↑12.50)	58.04	97.72(↑39.68)	58.04	91.81(↑33.77)	58.04	108.05(↑50.01)
SDXL	50.27	75.57(↑25.30)	50.27	57.09(↑6.82)	50.27	68.95(↑18.68)	50.27	68.52(↑18.25)	50.27	66.60(↑16.33)	50.27	53.93(↑3.66)
FLUX	68.44	68.48(↑0.04)	68.44	78.88(↑10.44)	68.44	72.55(↑4.11)	68.44	73.93(↑5.49)	68.44	74.39(↑5.95)	68.44	92.81(↑24.37)
Avg.	54.86	80.64(↑25.77)	54.86	74.44(↑19.58)	54.86	75.81(↑20.94)	54.86	85.30(↑30.43)	54.86	81.04(↑26.18)	54.86	99.21(↑44.34)

Model	Harmful Content (Invisible)				Bias Content (Invisible)				Neutral Content (Invisible)			
	naked		bloody		Asian		African		Muslim		hat	
	clean	typo	clean	typo	clean	typo	clean	typo	clean	typo	clean	typo
UnCLIP	42.70	48.48(↑5.78)	42.70	53.45(↑10.75)	42.70	52.35(↑9.65)	42.70	50.97(↑8.27)	42.70	56.77(↑14.07)	42.70	55.26(↑12.56)
SD1.5	58.04	65.15(↑7.11)	58.04	59.69(↑1.65)	58.04	61.30(↑3.26)	58.04	64.36(↑6.32)	58.04	62.97(↑4.93)	58.04	60.58(↑2.54)
SDXL	50.27	64.68(↑14.41)	50.27	62.97(↑12.70)	50.27	63.54(↑13.27)	50.27	65.04(↑14.77)	50.27	70.22(↑19.95)	50.27	64.56(↑14.29)
FLUX	68.44	161.69(↑93.25)	68.44	147.02(↑78.58)	68.44	161.05(↑92.61)	68.44	148.37(↑79.93)	68.44	142.66(↑74.22)	68.44	172.52(↑104.08)
Avg.	54.86	85.00(↑30.14)	54.86	80.78(↑25.92)	54.86	84.56(↑29.70)	54.86	82.19(↑27.32)	54.86	83.16(↑28.29)	54.86	88.23(↑33.37)

Table 8. The semantic impact of typography related to harmful, bias, and neutral concepts in image generation, measured by FID between the generated image from typographic input images and their corresponding original clean images. The values in parentheses represent the difference between the FID scores of images generated from typographic input images and those generated from clean input images when compared to original clean images, where a larger difference indicates a stronger typographic influence. (Text prompt: analog film photo, faded film, desaturated, 35mm photo)

Model	Typo Word Type								
	Nouns			Adjectives			Verbs		
	hat	flower	glasses	confused	surprised	tired	scream	cry	grimace
UnCLIP	152.41(↑105.61)	96.27(↑49.47)	109.02(↑62.22)	90.93(↑44.13)	82.46(↑35.66)	68.89(↑22.09)	87.67(↑40.87)	78.83(↑32.03)	90.80(↑44.00)
SD1.5	144.72(↑55.99)	134.51(↑45.78)	143.38(↑54.65)	124.10(↑35.37)	130.13(↑41.40)	140.21(↑51.48)	114.67(↑25.94)	130.73(↑42.00)	130.14(↑41.41)
SDXL	111.89(↓4.36)	99.54(↓16.71)	107.41(↓8.84)	111.41(↓4.84)	114.89(↓1.36)	83.13(↓33.12)	101.65(↓14.60)	81.23(↓35.02)	110.52(↓5.73)
FLUX	116.54(↑21.95)	123.03(↑28.44)	83.99(↓10.60)	128.52(↑33.93)	121.42(↑26.83)	80.57(↓14.02)	111.77(↑17.18)	94.96(↑0.37)	96.12(↑1.53)
Avg.	131.39(↑44.80)	113.34(↑26.75)	110.95(↑24.36)	113.74(↑27.15)	112.22(↑25.63)	93.20(↑6.61)	103.94(↑17.35)	96.44(↑9.84)	106.89(↑20.30)

Model	Clean	Typo Size				Typo Quantity			
		10pt	15pt	20pt	25pt	T1	T2	T3	T4
UnCLIP	46.80	55.47(↑8.67)	99.89(↑53.09)	135.42(↑88.62)	152.41(↑105.61)	76.89(↑30.09)	133.89(↑87.09)	149.58(↑102.78)	152.41(↑105.61)
SD1.5	88.73	85.38(↓3.35)	107.74(↑19.01)	134.81(↑46.08)	144.72(↑55.99)	99.84(↑11.11)	125.05(↑36.32)	139.37(↑50.64)	144.72(↑55.99)
SDXL	116.25	112.46(↓3.79)	97.29(↓18.96)	103.76(↓12.49)	111.89(↓4.36)	113.22(↓3.03)	106.70(↓9.55)	111.45(↓4.80)	111.89(↓4.36)
FLUX	94.59	73.86(↓20.73)	106.77(↑12.18)	117.35(↑22.76)	116.54(↑21.95)	90.56(↓4.03)	106.28(↑11.69)	110.86(↑16.27)	116.54(↑21.95)
Avg.	86.59	81.79(↓4.80)	102.92(↑16.33)	122.84(↑36.24)	131.39(↑44.80)	95.13(↑8.53)	117.98(↑31.39)	127.82(↑41.22)	131.39(↑44.80)

Model	Clean	Typo Opacity				Typo Position			
		25%	50%	75%	100%	A1	A2	A3	A4
UnCLIP	46.80	56.42(↑9.62)	69.16(↑22.36)	125.81(↑79.01)	152.41(↑105.61)	154.32(↑107.52)	154.26(↑107.46)	150.76(↑103.96)	147.24(↑100.44)
SD1.5	88.73	97.36(↑8.63)	111.66(↑22.93)	143.05(↑54.32)	144.72(↑55.99)	148.03(↑59.30)	149.17(↑60.44)	146.03(↑57.30)	143.23(↑54.50)
SDXL	116.25	113.98(↓2.27)	115.67(↓0.58)	109.63(↓6.62)	111.89(↓4.36)	114.00(↓2.25)	114.41(↓1.84)	108.23(↓8.02)	110.38(↓5.87)
FLUX	94.59	89.29(↓5.30)	104.48(↑9.89)	118.43(↑23.84)	116.54(↑21.95)	128.23(↑33.64)	111.62(↑17.03)	104.23(↑9.64)	126.22(↑31.63)
Avg.	86.59	89.26(↑2.67)	100.24(↑13.65)	124.23(↑37.64)	131.39(↑44.80)	132.69(↑46.10)	132.36(↑45.77)	127.31(↑40.72)	131.77(↑45.18)

Table 9. The semantic impact of typography with different typographic factors in image generation, which is measured by FID between the generated image from typographic input images and their corresponding original clean images. The values in parentheses represent the difference between the FID scores of images generated from typographic input images and those generated from clean input images when compared to original clean images, where a larger difference indicates a stronger typographic influence. (Text prompt: a youthful figure on the stage)

Model	Harmful Content				Bias Content				Neutral Content			
	naked		bloody		Asian		African		Muslim		hat	
	clean	typo	clean	typo	clean	typo	clean	typo	clean	typo	clean	typo
UnCLIP	46.99	98.04(↑51.05)	46.99	86.34(↑39.35)	46.99	93.31(↑46.32)	46.99	104.21(↑57.22)	46.99	101.28(↑54.29)	46.99	150.64(↑103.65)
SD1.5	91.00	162.98(↑71.98)	91.00	150.23(↑59.23)	91.00	130.93(↑39.93)	91.00	150.37(↑59.37)	91.00	137.36(↑46.36)	91.00	143.05(↑52.05)
SDXL	113.92	151.94(↑38.02)	113.92	116.15(↑2.23)	113.92	116.88(↑2.96)	113.92	125.63(↑11.71)	113.92	117.55(↑3.63)	113.92	114.35(↑0.43)
FLUX	93.66	99.80(↑6.14)	93.66	105.09(↑11.43)	93.66	105.66(↑12.00)	93.66	102.08(↑8.42)	93.66	94.16(↑0.50)	93.66	111.61(↑17.95)
Avg.	86.39	128.19(↑41.80)	86.39	114.45(↑28.06)	86.39	111.69(↑25.30)	86.39	120.57(↑34.18)	86.39	112.59(↑26.19)	86.39	129.91(↑43.52)

Model	Harmful Content (Invisible)				Bias Content (Invisible)				Neutral Content (Invisible)			
	naked		bloody		Asian		African		Muslim		hat	
	clean	typo	clean	typo	clean	typo	clean	typo	clean	typo	clean	typo
UnCLIP	46.99	53.61(↑6.62)	46.99	57.25(↑10.26)	46.99	56.96(↑9.97)	46.99	54.85(↑7.86)	46.99	59.54(↑12.55)	46.99	58.10(↑11.11)
SD1.5	91.00	102.25(↑11.25)	91.00	99.69(↑8.69)	91.00	96.90(↑5.90)	91.00	94.70(↑3.70)	91.00	99.35(↑8.35)	91.00	92.60(↑1.60)
SDXL	113.92	128.51(↑14.59)	113.92	118.11(↑4.19)	113.92	124.45(↑10.53)	113.92	117.71(↑3.79)	113.92	123.93(↑10.01)	113.92	116.37(↑2.45)
FLUX	93.66	111.25(↑17.59)	93.66	107.26(↑13.60)	93.66	111.05(↑17.39)	93.66	100.88(↑7.22)	93.66	96.60(↑2.94)	93.66	114.93(↑21.27)
Avg.	86.39	98.91(↑12.52)	86.39	95.58(↑9.19)	86.39	97.34(↑10.95)	86.39	92.04(↑5.65)	86.39	94.85(↑8.46)	86.39	95.50(↑9.11)

Table 10. The semantic impact of typography related to harmful, bias, and neutral concepts in image generation, measured by FID between the generated image from typographic input images and their corresponding original clean images. The values in parentheses represent the difference between the FID scores of images generated from typographic input images and those generated from clean input images when compared to original clean images, where a larger difference indicates a stronger typographic influence. (Text prompt: a youthful figure on the stage)