

# A Framework for Evaluating Zero-Shot Image Generation in Concept-based Explainability

## Supplementary Material

### A. Complete List of Concepts

Table 1 contains a detailed overview of the concept-class dataset used in our study. The dataset comprises 41 concepts, each mapped to a unique class label derived from the ImageNet dataset. Real concept images are drawn from four distinct datasets: DTD, ImageNet, FMD, and search engine results.

Type	Dataset	Concept	Relevant Class	Type	Dataset	Concept	Relevant Class
Texture	DTD	Bubbly	Beer glass	Object	ImageNet	Asparagus	Stone wall
Texture	DTD	Chequered	Crossword puzzle	Object	ImageNet	Cloister	Church
Texture	DTD	Cracked	Stone wall	Object	ImageNet	Feline	Leopard
Texture	DTD	Crystalline	Honeycomb	Object	ImageNet	Grandfather	Tench
Texture	DTD	Dotted	Dalmatian	Object	ImageNet	Guitarist	Electric guitar
Texture	DTD	Frilly	Overskirt	Object	ImageNet	Kitchen	Waffle iron
Texture	DTD	Honeycombed	Honeycomb	Object	ImageNet	Lichen	Stone wall
Texture	DTD	Meshed	Honeycomb	Object	ImageNet	Rodent	Leopard
Texture	DTD	Perforated	Honeycomb	Object	ImageNet	Steeple	Church
Texture	DTD	Striped	Zebra	Object	ImageNet	Tattoo	Zebra
Texture	DTD	Waffled	Waffle iron	Object	Search	Cast iron	Waffle iron
Texture	DTD	Woven	Overskirt	Object	Search	Fin	Tench
Texture	DTD	Wrinkled	Overskirt	Object	Search	Grass	Golf ball
Texture	FMD	Fabric	Trench coat	Object	Search	Newspaper	Crossword puzzle
Texture	FMD	Foliage	Spider web	Object	Search	Pen	Fountain pen
Texture	FMD	Glass	Beer glass	Object	Search	Sphere	Golf ball
Texture	FMD	Leather	Trench coat	Texture	Search	Leopard print	Leopard
Texture	FMD	Metal	Waffle iron	Texture	Search	Spotted	Dalmatian
Texture	FMD	Paper	Crossword puzzle				
Texture	FMD	Plastic	Golf ball				
Texture	FMD	Stone	Stone wall				
Texture	FMD	Water	Beer glass				
Texture	FMD	Wood	Wooden spoon				

Table 1. Summary of the concept-class dataset used in our experiments. Each concept is associated with a specific class (e.g., dotted → dalmatian). Concepts are drawn from four distinct datasets: Describable Textures Dataset (DTD), ImageNet, Flickr Material Database (FMD), and search engine results.

## B. Average Cosine Similarity per Concept using Visual-TCAV

Class	beer_glass			church		crossword_puzzle
Concept	bubbly	glass	water	cloister	steeple	chequered
Flux	0.666 ± 0.14	0.453 ± 0.139	0.59 ± 0.069	0.848 ± 0.017	0.726 ± 0.07	0.217 ± 0.141
GPT-Image-1	0.659 ± 0.038	0.389 ± 0.162	0.509 ± 0.14	0.389 ± 0.081	0.842 ± 0.042	0.586 ± 0.096
SD 3.5	0.689 ± 0.1	0.395 ± 0.103	0.523 ± 0.099	0.894 ± 0.026	0.764 ± 0.057	0.664 ± 0.102
Class	crossword_puzzle		dalmatian		electric_guitar	fountain_pen
Concept	newspaper	paper	dotted	spotted	guitarist	pen
Flux	0.589 ± 0.052	0.589 ± 0.067	0.376 ± 0.126	0.375 ± 0.155	0.815 ± 0.04	0.595 ± 0.206
GPT-Image-1	0.702 ± 0.042	0.475 ± 0.061	0.654 ± 0.155	0.554 ± 0.144	0.699 ± 0.065	0.305 ± 0.174
SD 3.5	0.69 ± 0.125	0.533 ± 0.074	0.525 ± 0.098	0.346 ± 0.138	0.818 ± 0.036	0.482 ± 0.226
Class	golf_ball			honeycomb		
Concept	grass	plastic	sphere	crystalline	honeycombed	meshed
Flux	0.411 ± 0.168	0.504 ± 0.092	0.483 ± 0.111	0.486 ± 0.083	0.707 ± 0.188	0.588 ± 0.123
GPT-Image-1	0.328 ± 0.194	0.247 ± 0.085	0.487 ± 0.094	0.482 ± 0.094	0.787 ± 0.07	0.411 ± 0.057
SD 3.5	0.485 ± 0.131	0.435 ± 0.096	0.561 ± 0.103	0.498 ± 0.153	0.866 ± 0.045	0.571 ± 0.145
Class	honeycomb	leopard		overskirt		
Concept	perforated	feline	leopard_print	rodent	frilly	woven
Flux	0.336 ± 0.123	0.537 ± 0.132	0.6 ± 0.127	0.607 ± 0.072	0.473 ± 0.142	0.456 ± 0.182
GPT-Image-1	0.658 ± 0.064	0.732 ± 0.062	0.828 ± 0.13	0.685 ± 0.117	0.512 ± 0.128	0.378 ± 0.082
SD 3.5	0.641 ± 0.045	0.612 ± 0.114	0.682 ± 0.1	0.726 ± 0.095	0.432 ± 0.151	0.691 ± 0.066
Class	overskirt	spider_web	stone_wall			
Concept	wrinkled	foliage	asparagus	cracked	lichen	stone
Flux	0.535 ± 0.133	0.568 ± 0.192	0.622 ± 0.124	0.657 ± 0.041	0.382 ± 0.312	0.542 ± 0.099
GPT-Image-1	0.419 ± 0.106	0.628 ± 0.17	0.279 ± 0.223	0.696 ± 0.057	0.505 ± 0.115	0.559 ± 0.176
SD 3.5	0.559 ± 0.14	0.508 ± 0.217	0.693 ± 0.078	0.678 ± 0.116	0.569 ± 0.027	0.517 ± 0.144
Class	tench		trench_coat		waffle_iron	
Concept	fin	grandfather	fabric	leather	cast_iron	kitchen
Flux	0.151 ± 0.084	0.666 ± 0.085	0.572 ± 0.111	0.602 ± 0.124	0.164 ± 0.114	0.773 ± 0.117
GPT-Image-1	0.282 ± 0.097	0.658 ± 0.155	0.508 ± 0.114	0.572 ± 0.129	0.31 ± 0.104	0.445 ± 0.086
SD 3.5	0.317 ± 0.118	0.681 ± 0.099	0.55 ± 0.102	0.567 ± 0.111	0.282 ± 0.144	0.715 ± 0.069
Class	waffle_iron		wooden_spoon	zebra		
Concept	metal	waffled	wood	striped	tattoo	
Flux	0.458 ± 0.134	0.81 ± 0.047	0.542 ± 0.129	0.583 ± 0.236	0.525 ± 0.058	
GPT-Image-1	0.502 ± 0.108	0.826 ± 0.029	0.656 ± 0.131	0.624 ± 0.149	0.345 ± 0.088	
SD 3.5	0.399 ± 0.206	0.869 ± 0.037	0.572 ± 0.096	0.335 ± 0.151	0.755 ± 0.045	

Table 2. Average cosine similarities between real and generated concepts whose CAVs are computed using Visual-TCAV. Results are aggregated by averaging across four architectures and two layers each (i.e., the outputs of the final two blocks for ResNet-50-V2, Inception-V3, and ConvNeXt, and the outputs of the last two convolutional layers of VGG-16).

### C. Average Cosine Similarity per Concept using TCAV

Class	beer_glass			church		crossword_puzzle
Concept	bubbly	glass	water	cloister	steeple	chequered
Flux	0.416 ± 0.137	0.275 ± 0.06	0.438 ± 0.049	0.689 ± 0.065	0.515 ± 0.108	0.151 ± 0.098
GPT-Image-1	0.563 ± 0.056	0.249 ± 0.046	0.517 ± 0.082	0.11 ± 0.032	0.779 ± 0.047	0.542 ± 0.051
SD 3.5	0.526 ± 0.1	0.258 ± 0.086	0.402 ± 0.056	0.74 ± 0.078	0.742 ± 0.056	0.746 ± 0.025
Class	crossword_puzzle		dalmatian		electric_guitar	fountain_pen
Concept	newspaper	paper	dotted	spotted	guitarist	pen
Flux	0.535 ± 0.055	0.4 ± 0.068	0.448 ± 0.06	0.259 ± 0.119	0.693 ± 0.098	0.655 ± 0.11
GPT-Image-1	0.541 ± 0.04	0.376 ± 0.056	0.597 ± 0.099	0.443 ± 0.12	0.564 ± 0.072	0.237 ± 0.171
SD 3.5	0.759 ± 0.041	0.426 ± 0.037	0.568 ± 0.069	0.395 ± 0.133	0.724 ± 0.094	0.539 ± 0.072
Class	golf_ball			honeycomb		
Concept	grass	plastic	sphere	crystalline	honeycombed	meshed
Flux	0.401 ± 0.159	0.428 ± 0.049	0.282 ± 0.036	0.413 ± 0.052	0.802 ± 0.05	0.584 ± 0.061
GPT-Image-1	0.285 ± 0.181	0.166 ± 0.134	0.263 ± 0.071	0.399 ± 0.05	0.77 ± 0.096	0.271 ± 0.071
SD 3.5	0.444 ± 0.131	0.362 ± 0.051	0.413 ± 0.049	0.376 ± 0.043	0.853 ± 0.037	0.659 ± 0.063
Class	honeycomb	leopard		overskirt		
Concept	perforated	feline	leopard_print	rodent	frilly	woven
Flux	0.372 ± 0.053	0.387 ± 0.104	0.383 ± 0.256	0.296 ± 0.231	0.431 ± 0.09	0.405 ± 0.14
GPT-Image-1	0.61 ± 0.071	0.644 ± 0.053	0.878 ± 0.016	0.477 ± 0.086	0.454 ± 0.106	0.236 ± 0.102
SD 3.5	0.572 ± 0.065	0.485 ± 0.047	0.57 ± 0.202	0.575 ± 0.107	0.413 ± 0.102	0.706 ± 0.038
Class	overskirt	spider_web	stone_wall			
Concept	wrinkled	foliage	asparagus	cracked	lichen	stone
Flux	0.31 ± 0.061	0.366 ± 0.12	0.545 ± 0.071	0.585 ± 0.071	0.255 ± 0.133	0.309 ± 0.057
GPT-Image-1	0.313 ± 0.058	0.676 ± 0.048	0.107 ± 0.105	0.496 ± 0.257	0.523 ± 0.038	0.549 ± 0.045
SD 3.5	0.488 ± 0.06	0.368 ± 0.101	0.631 ± 0.093	0.666 ± 0.073	0.666 ± 0.055	0.318 ± 0.125
Class	tench		trench_coat		waffle_iron	
Concept	fin	grandfather	fabric	leather	cast_iron	kitchen
Flux	0.08 ± 0.046	0.595 ± 0.059	0.565 ± 0.081	0.542 ± 0.11	0.162 ± 0.065	0.797 ± 0.016
GPT-Image-1	0.116 ± 0.071	0.541 ± 0.095	0.523 ± 0.051	0.528 ± 0.041	0.161 ± 0.078	0.25 ± 0.053
SD 3.5	0.099 ± 0.037	0.515 ± 0.062	0.562 ± 0.041	0.537 ± 0.095	0.245 ± 0.166	0.653 ± 0.03
Class	waffle_iron		wooden_spoon	zebra		
Concept	metal	waffled	wood	striped	tattoo	
Flux	0.282 ± 0.022	0.759 ± 0.059	0.333 ± 0.103	0.462 ± 0.092	0.444 ± 0.033	
GPT-Image-1	0.411 ± 0.042	0.818 ± 0.075	0.582 ± 0.053	0.628 ± 0.089	0.193 ± 0.088	
SD 3.5	0.398 ± 0.036	0.835 ± 0.043	0.395 ± 0.075	0.376 ± 0.132	0.662 ± 0.031	

Table 3. Average cosine similarities between real and generated concepts whose CAVs are computed using TCAV. Results are aggregated by averaging across four architectures and two layers each (i.e., the outputs of the final two blocks for ResNet-50-V2, Inception-V3, and ConvNeXt, and the outputs of the last two convolutional layers of VGG-16).

## D. Average Importances per Concept using VisualTCAV

Class	beer_glass			church		crossword_puzzle
Concept	bubbly	glass	water	cloister	steeple	chequered
Real	0.027 ± 0.018	0.104 ± 0.04	0.042 ± 0.039	0.07 ± 0.052	0.111 ± 0.056	0.086 ± 0.077
Flux	0.011 ± 0.011	0.081 ± 0.058	0.03 ± 0.02	0.075 ± 0.055	0.112 ± 0.052	0.02 ± 0.021
GPT-Image-1	0.006 ± 0.007	0.085 ± 0.038	0.001 ± 0.001	0.015 ± 0.012	0.082 ± 0.041	0.022 ± 0.015
SD 3.5	0.012 ± 0.01	0.069 ± 0.058	0.003 ± 0.003	0.054 ± 0.036	0.067 ± 0.034	0.076 ± 0.092
Class	crossword_puzzle		dalmatian		electric_guitar	fountain_pen
Concept	newspaper	paper	dotted	spotted	guitarist	pen
Real	0.123 ± 0.082	0.039 ± 0.032	0.012 ± 0.014	0.171 ± 0.146	0.149 ± 0.113	0.04 ± 0.014
Flux	0.049 ± 0.052	0.058 ± 0.045	0.004 ± 0.009	0.032 ± 0.049	0.134 ± 0.085	0.057 ± 0.027
GPT-Image-1	0.093 ± 0.069	0.013 ± 0.013	0.005 ± 0.006	0.027 ± 0.033	0.089 ± 0.09	0.032 ± 0.011
SD 3.5	0.089 ± 0.089	0.035 ± 0.046	0.0 ± 0.0	0.018 ± 0.028	0.167 ± 0.079	0.055 ± 0.043
Class	golf_ball			honeycomb		
Concept	grass	plastic	sphere	crystalline	honeycombed	meshed
Real	0.008 ± 0.01	0.049 ± 0.028	0.148 ± 0.107	0.03 ± 0.026	0.193 ± 0.136	0.13 ± 0.124
Flux	0.004 ± 0.008	0.014 ± 0.013	0.064 ± 0.038	0.058 ± 0.04	0.156 ± 0.073	0.106 ± 0.117
GPT-Image-1	0.014 ± 0.011	0.008 ± 0.01	0.065 ± 0.025	0.018 ± 0.019	0.11 ± 0.074	0.025 ± 0.035
SD 3.5	0.003 ± 0.002	0.013 ± 0.016	0.051 ± 0.031	0.012 ± 0.006	0.193 ± 0.121	0.028 ± 0.019
Class	honeycomb	leopard		overskirt		
Concept	perforated	feline	leopard_print	rodent	frilly	woven
Real	0.095 ± 0.07	0.125 ± 0.074	0.128 ± 0.063	0.027 ± 0.029	0.025 ± 0.022	0.003 ± 0.004
Flux	0.049 ± 0.034	0.049 ± 0.04	0.063 ± 0.083	0.012 ± 0.011	0.034 ± 0.031	0.008 ± 0.011
GPT-Image-1	0.064 ± 0.05	0.056 ± 0.053	0.067 ± 0.056	0.012 ± 0.016	0.006 ± 0.009	0.003 ± 0.003
SD 3.5	0.04 ± 0.024	0.038 ± 0.032	0.088 ± 0.076	0.012 ± 0.015	0.057 ± 0.035	0.005 ± 0.008
Class	overskirt	spider_web	stone_wall			
Concept	wrinkled	foliage	asparagus	cracked	lichen	stone
Real	0.02 ± 0.024	0.015 ± 0.014	0.02 ± 0.019	0.074 ± 0.061	0.058 ± 0.071	0.1 ± 0.08
Flux	0.03 ± 0.03	0.005 ± 0.006	0.004 ± 0.004	0.049 ± 0.032	0.006 ± 0.005	0.067 ± 0.043
GPT-Image-1	0.003 ± 0.005	0.003 ± 0.006	0.002 ± 0.001	0.034 ± 0.033	0.006 ± 0.004	0.027 ± 0.02
SD 3.5	0.013 ± 0.011	0.008 ± 0.009	0.003 ± 0.003	0.03 ± 0.025	0.013 ± 0.01	0.063 ± 0.019
Class	tench		trench_coat		waffle_iron	
Concept	fin	grandfather	fabric	leather	cast_iron	kitchen
Real	0.055 ± 0.034	0.036 ± 0.036	0.028 ± 0.023	0.015 ± 0.012	0.13 ± 0.081	0.019 ± 0.013
Flux	0.002 ± 0.002	0.013 ± 0.01	0.018 ± 0.016	0.007 ± 0.005	0.041 ± 0.029	0.018 ± 0.011
GPT-Image-1	0.002 ± 0.002	0.023 ± 0.019	0.006 ± 0.007	0.004 ± 0.004	0.028 ± 0.023	0.048 ± 0.041
SD 3.5	0.006 ± 0.011	0.017 ± 0.011	0.005 ± 0.004	0.001 ± 0.0	0.027 ± 0.022	0.024 ± 0.013
Class	waffle_iron		wooden_spoon	zebra		
Concept	metal	waffled	wood	striped	tattoo	
Real	0.052 ± 0.034	0.098 ± 0.059	0.033 ± 0.032	0.195 ± 0.148	0.028 ± 0.026	
Flux	0.027 ± 0.02	0.06 ± 0.044	0.025 ± 0.02	0.111 ± 0.076	0.012 ± 0.009	
GPT-Image-1	0.011 ± 0.011	0.061 ± 0.042	0.025 ± 0.026	0.089 ± 0.08	0.009 ± 0.009	
SD 3.5	0.024 ± 0.022	0.095 ± 0.041	0.014 ± 0.011	0.033 ± 0.035	0.04 ± 0.043	

Table 4. Average importance scores for each concept computed with Visual-TCAV. Results are aggregated by averaging across four architectures and two layers each (i.e., the outputs of the final two blocks for ResNet-50-v2, Inception-V3, and ConvNeXt, and the outputs of the last two convolutional layers of VGG-16).

## E. Average Importances per Concept using TCAV

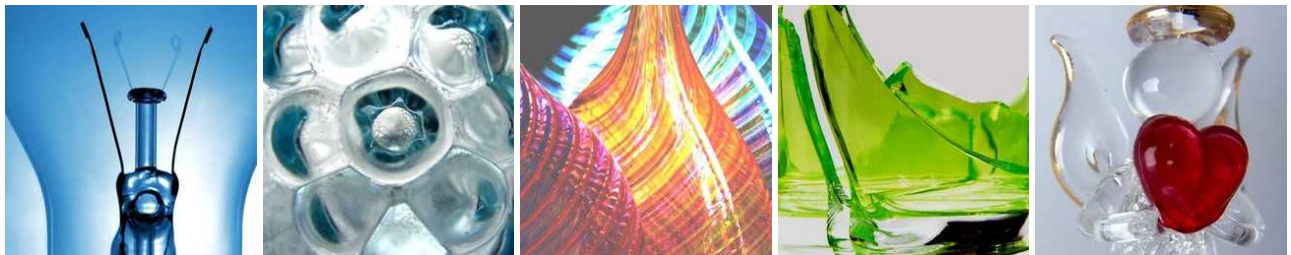
Class	beer_glass			church		crossword_puzzle
Concept	bubbly	glass	water	cloister	steeple	chequered
Real	0.72 ± 0.301	0.872 ± 0.185	0.529 ± 0.452	0.789 ± 0.301	0.995 ± 0.011	0.638 ± 0.401
Flux	0.635 ± 0.486	0.858 ± 0.293	0.676 ± 0.429	0.929 ± 0.137	0.982 ± 0.036	0.81 ± 0.22
GPT-Image-1	0.789 ± 0.323	0.978 ± 0.044	0.122 ± 0.144	0.71 ± 0.347	0.979 ± 0.026	0.65 ± 0.413
SD 3.5	0.721 ± 0.419	0.654 ± 0.407	0.397 ± 0.315	0.84 ± 0.22	0.985 ± 0.039	0.678 ± 0.419
Class	crossword_puzzle		dalmatian		electric_guitar	fountain_pen
Concept	newspaper	paper	dotted	spotted	guitarist	pen
Real	0.863 ± 0.223	0.638 ± 0.416	0.166 ± 0.197	0.939 ± 0.087	0.99 ± 0.015	0.825 ± 0.209
Flux	0.715 ± 0.409	0.678 ± 0.409	0.145 ± 0.164	0.921 ± 0.091	0.983 ± 0.035	0.89 ± 0.148
GPT-Image-1	0.764 ± 0.424	0.702 ± 0.301	0.282 ± 0.363	0.42 ± 0.387	0.948 ± 0.089	0.988 ± 0.028
SD 3.5	0.716 ± 0.433	0.922 ± 0.174	0.007 ± 0.02	0.387 ± 0.401	0.99 ± 0.021	0.828 ± 0.111
Class	golf_ball			honeycomb		
Concept	grass	plastic	sphere	crystalline	honeycombed	meshed
Real	0.398 ± 0.422	0.311 ± 0.36	0.714 ± 0.26	0.61 ± 0.407	0.972 ± 0.037	0.867 ± 0.096
Flux	0.368 ± 0.313	0.45 ± 0.414	0.735 ± 0.454	0.9 ± 0.124	0.968 ± 0.041	0.67 ± 0.396
GPT-Image-1	0.881 ± 0.166	0.508 ± 0.379	0.706 ± 0.442	0.71 ± 0.285	0.972 ± 0.037	0.72 ± 0.387
SD 3.5	0.428 ± 0.328	0.636 ± 0.339	0.833 ± 0.265	0.645 ± 0.45	0.983 ± 0.027	0.802 ± 0.205
Class	honeycomb	leopard		overskirt		
Concept	perforated	feline	leopard_print	rodent	frilly	woven
Real	0.698 ± 0.357	0.968 ± 0.084	0.977 ± 0.023	0.78 ± 0.379	0.474 ± 0.18	0.048 ± 0.061
Flux	0.73 ± 0.44	0.948 ± 0.104	0.5 ± 0.367	0.562 ± 0.476	0.959 ± 0.078	0.628 ± 0.488
GPT-Image-1	0.727 ± 0.337	0.935 ± 0.184	0.97 ± 0.039	0.878 ± 0.293	0.513 ± 0.494	0.61 ± 0.506
SD 3.5	0.695 ± 0.422	0.942 ± 0.147	0.668 ± 0.421	0.893 ± 0.252	0.98 ± 0.05	0.054 ± 0.058
Class	overskirt	spider_web	stone_wall			
Concept	wrinkled	foliage	asparagus	cracked	lichen	stone
Real	0.27 ± 0.255	0.455 ± 0.421	0.468 ± 0.408	0.683 ± 0.429	0.707 ± 0.426	0.675 ± 0.411
Flux	0.645 ± 0.418	0.578 ± 0.397	0.47 ± 0.46	0.68 ± 0.435	0.583 ± 0.453	0.748 ± 0.461
GPT-Image-1	0.117 ± 0.182	0.25 ± 0.337	0.537 ± 0.395	0.852 ± 0.217	0.815 ± 0.299	0.625 ± 0.409
SD 3.5	0.439 ± 0.374	0.46 ± 0.383	0.395 ± 0.402	0.635 ± 0.426	0.708 ± 0.442	0.985 ± 0.035
Class	tench		trench_coat		waffle_iron	
Concept	fin	grandfather	fabric	leather	cast_iron	kitchen
Real	0.572 ± 0.406	0.502 ± 0.382	0.4 ± 0.252	0.375 ± 0.12	0.74 ± 0.231	0.805 ± 0.176
Flux	0.467 ± 0.239	0.535 ± 0.414	0.518 ± 0.405	0.17 ± 0.064	0.938 ± 0.073	0.867 ± 0.133
GPT-Image-1	0.575 ± 0.332	0.565 ± 0.356	0.552 ± 0.344	0.222 ± 0.332	0.84 ± 0.306	0.942 ± 0.074
SD 3.5	0.7 ± 0.331	0.553 ± 0.381	0.32 ± 0.245	0.04 ± 0.037	0.68 ± 0.433	0.93 ± 0.109
Class	waffle_iron		wooden_spoon	zebra		
Concept	metal	waffled	wood	striped	tattoo	
Real	0.69 ± 0.3	0.705 ± 0.269	0.368 ± 0.271	0.925 ± 0.101	0.347 ± 0.374	
Flux	0.685 ± 0.435	0.832 ± 0.224	0.718 ± 0.424	0.993 ± 0.01	0.527 ± 0.4	
GPT-Image-1	0.49 ± 0.392	0.785 ± 0.23	0.5 ± 0.382	0.9 ± 0.115	0.465 ± 0.344	
SD 3.5	0.513 ± 0.356	0.84 ± 0.204	0.55 ± 0.41	0.617 ± 0.396	0.415 ± 0.398	

Table 5. Average importance scores for each concept computed with TCAV. Results are aggregated by averaging across four architectures and two layers each (i.e., the outputs of the final two blocks for ResNet-50-v2, Inception-V3, and ConvNeXt, and the outputs of the last two convolutional layers of VGG-16).

## F. Generated Concepts

Below is a randomly selected subset of the synthetic concepts generated in this study.

### F.1. Glass Concept



(a) Real glass concept images from FMD.



(b) Glass concept images generated with Flux 1.1.



(c) Glass concept images generated with GPT-Image-1.



(d) Glass concept images generated with Stable Diffusion 3.5 Medium.

Figure 10. Comparison of real glass images and generated concept images across various models.

**F.2. Water Concept**



(a) Real water concept images from FMD.



(b) Water concept images generated with Flux 1.1.



(c) Water concept images generated with GPT-Image-1.



(d) Water concept images generated with Stable Diffusion 3.5 Medium.

Figure 11. A comparison of real images (top row) with the corresponding concept images produced by the different generative models: Flux 1.1, GPT-Image-1, Stable Diffusion 3.5 Medium.

**F.3. Bubbly Concept**



(a) Real bubbly concept images from DTD.



(b) Bubbly concept images generated with Flux 1.1.



(c) Bubbly concept images generated with GPT-Image-1.



(d) Bubbly concept images generated with Stable Diffusion 3.5 Medium.

Figure 12. Real bubbly images from the DTD dataset (top row) compared with concept images generated by Flux 1.1, GPT-Image-1, and Stable Diffusion 3.5 Medium (subsequent rows).

#### F.4. Cloister Concept



(a) Real cloister concept images from ImageNet.



(b) Cloister concept images generated with Flux 1.1.



(c) Cloister concept images generated with GPT-Image-1.



(d) Cloister concept images generated with Stable Diffusion 3.5 Medium.

Figure 13. A comparative example for the cloister concept. The first subfigure shows five authentic cloister images from the ImageNet dataset. Subsequent ones display five synthetic renditions of the same concept, generated respectively by Flux 1.1, GPT-Image-1, and Stable Diffusion 3.5 Medium.

## F.5. Steeple Concept



(a) Real steeple concept images from ImageNet.



(b) Steeple concept images generated with Flux 1.1.



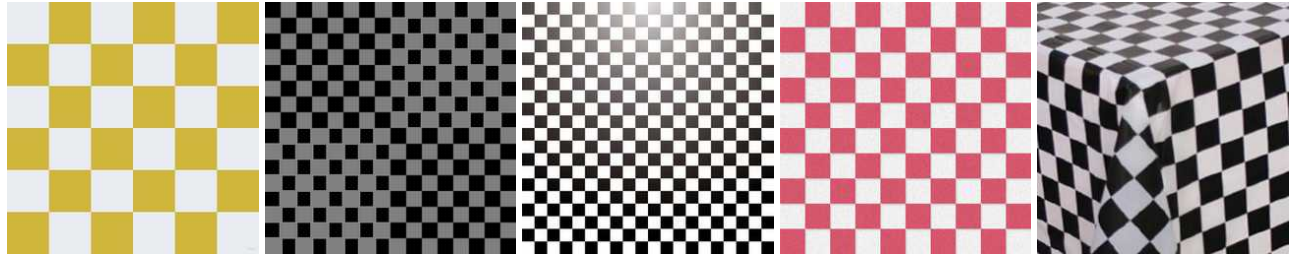
(c) Steeple concept images generated with GPT-Image-1.



(d) Steeple concept images generated with Stable Diffusion 3.5 Medium.

Figure 14. Real steeple images from ImageNet compared with synthetic steeple images generated by Flux 1.1, GPT-Image-1, and Stable Diffusion 3.5 Medium. Each row contains five representative samples of the same concept.

## F.6. Chequered Concept



(a) Real chequered concept images from DTD.



(b) Chequered concept images generated with Flux 1.1.



(c) Chequered concept images generated with GPT-Image-1.



(d) Chequered concept images generated with Stable Diffusion 3.5 Medium.

Figure 15. Chequered concept images. Real DTD examples; images generated by Flux 1.1; images generated by GPT-Image-1; images generated by Stable Diffusion 3.5 Medium.

## F.7. Newspaper Concept



(a) Real newspaper concept images from search engine results.



(b) Newspaper concept images generated with Flux 1.1.



(c) Newspaper concept images generated with GPT-Image-1.



(d) Newspaper concept images generated with Stable Diffusion 3.5 Medium.

Figure 16. Newspaper concept images. Row 1 shows five real samples; rows 2–4 display five synthetic images produced by Flux 1.1, GPT-Image-1, and Stable Diffusion 3.5 Medium, respectively.

**F.8. Paper Concept**



(a) Real paper concept images from FMD.



(b) Paper concept images generated with Flux 1.1.



(c) Paper concept images generated with GPT-Image-1.



(d) Paper concept images generated with Stable Diffusion 3.5 Medium.

Figure 17. Comparison of paper concepts: real FMD examples, Flux 1.1 generated images, GPT-Image-1 generated images, and Stable Diffusion 3.5 Medium generated images.

## F.9. Dotted Concept



(a) Real dotted concept images from DTD.



(b) Dotted concept images generated with Flux 1.1.



(c) Dotted concept images generated with GPT-Image-1.



(d) Dotted concept images generated with Stable Diffusion 3.5 Medium.

Figure 18. Real dotted concept images from the DTD dataset; images produced with Flux 1.1; images produced with GPT-Image-1; images produced with Stable Diffusion 3.5 Medium.

## F.10. Spotted Concept



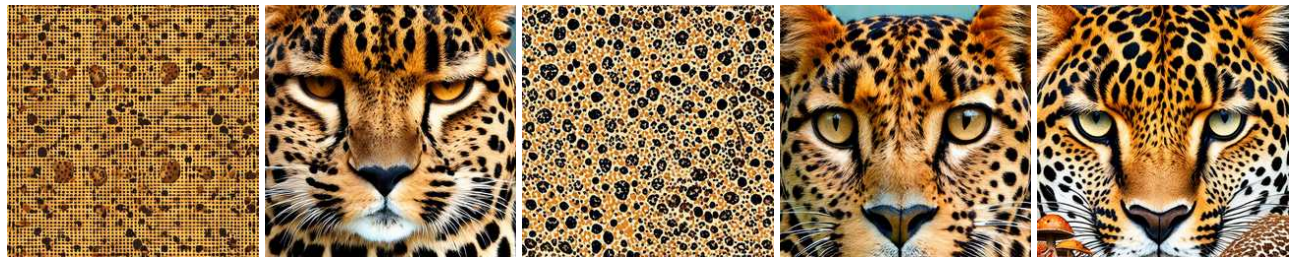
(a) Real spotted concept images from search engine results.



(b) Spotted concept images generated with Flux 1.1.



(c) Spotted concept images generated with GPT-Image-1.



(d) Spotted concept images generated with Stable Diffusion 3.5 Medium.

Figure 19. Real and synthetic spotted-pattern concept images.

## F.11. Grass Concept



(a) Real grass concept images from search engine results.



(b) Grass concept images generated with Flux 1.1.



(c) Grass concept images generated with GPT-Image-1.



(d) Grass concept images generated with Stable Diffusion 3.5 Medium.

Figure 20. Grass concept images: real photographs, Flux 1.1 outputs, GPT-Image-1 outputs, and Stable Diffusion 3.5 Medium outputs.

## F.12. Plastic Concept



(a) Real plastic concept images from FMD.



(b) Plastic concept images generated with Flux 1.1.



(c) Plastic concept images generated with GPT-Image-1.



(d) Plastic concept images generated with Stable Diffusion 3.5 Medium.

Figure 21. Plastic concept images from real photographs, Flux 1.1, GPT-Image-1, and Stable Diffusion 3.5 Medium.

F.13. Sphere Concept



(a) Real sphere concept images from search engine results.



(b) Sphere concept images generated with Flux 1.1.



(c) Sphere concept images generated with GPT-Image-1.



(d) Sphere concept images generated with Stable Diffusion 3.5 Medium.

Figure 22

## G. Ablated Class Images

Below is a randomly selected subset of the ablated class images used in this study.

### G.1. Beer Glass Class



Figure 23. Example of concept removal from a beer glass class image, eliminating bubbly, glass, and water concepts.

### G.2. Church Class



Figure 24. Example of concept removal from a church class image, eliminating the steeple concept and the cloister concept.

### G.3. Crossword Puzzle Class



Figure 25. Example of concept removal from a crossword puzzle class image, eliminating chequered, newspaper, and paper concepts.

### G.4. Dalmatian Class

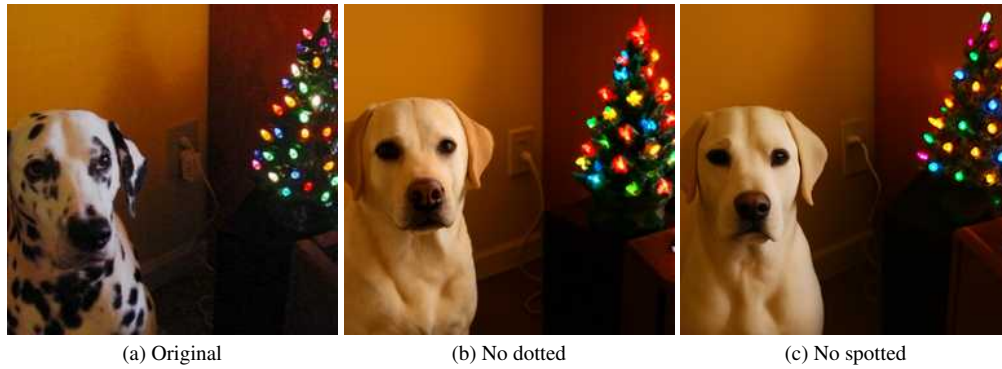


Figure 26. Example of concept removal from a dalmatian class image, eliminating dotted and spotted concepts.