

```

import torch
from diffusers import StableDiffusionPipeline
import os
import re
from PIL import Image
from IPython.display import display

# 1. Initialize Model (Stable Diffusion – Pre-trained Generative Model)
model_id = "runwayml/stable-diffusion-v1-5"

pipe = StableDiffusionPipeline.from_pretrained(
    model_id,
    torch_dtype=torch.float16
).to("cuda")

# 2. Define Input Prompts / Domains
labels = [
    "Normal anatomy – healthy lungs, age & gender variations",
    "Infectious patterns – bacterial/viral pneumonia, COVID-like opacities",
    "Lung opacities – focal, diffuse, ground-glass, consolidations",
    "Pleural conditions – pleural effusion, pneumothorax",
    "Structural lesions – nodules, masses, fibrosis",
    "Cardiac findings – cardiomegaly, vascular congestion",
    "Medical devices – tubes, catheters, pacemakers",
    "Imaging artifacts – noise, motion blur, exposure issues",
    "View & positioning – PA/AP views, rotation, supine/erect",
    "Domain shift – scanner, hospital, and resolution variations"
]

variation_descriptions = [
    "very mild and early-stage findings",
    "mild but clearly visible clinical findings",
    "moderate severity with noticeable abnormalities",
    "severe and advanced pathological findings",
    "critical condition with extensive abnormalities"
]

# 3. Setup Dataset Storage
base_dir = "/content/xray_dataset" # Change path if not using Colab
os.makedirs(base_dir, exist_ok=True)

def clean_label(label):
    return re.sub(r'^[a-zA-Z0-9_]+', '_', label.split("-")[0].strip())

def build_prompt(label, i):
    return f"""
        A high-resolution clinical chest X-ray image showing {label}.
        The case represents {variation_descriptions[i-1]}.
    """

```

```

The image should be medically accurate, realistic, and suitable
for diagnostic use.
Hospital imaging style, sharp anatomical details, proper contrast.
No artistic effects, no cartoons, professional medical scan.
"""

# 4. Generate Synthetic Data
print("Starting Synthetic Data Generation...")

for label in labels:
    folder = clean_label(label)
    folder_path = os.path.join(base_dir, folder)
    os.makedirs(folder_path, exist_ok=True)

    # Generate 5 variations per category
    for i in range(1, 6):
        prompt = build_prompt(label, i)

        image = pipe(
            prompt,
            num_inference_steps=50,
            guidance_scale=8.5
        ).images[0]

        # 5. Save Output
        filename = f"{folder}_{i}.png"
        path = os.path.join(folder_path, filename)
        image.save(path)

        print(f"Saved: {path}")

    # Display first image of each category
    if i == 1:
        print(f"Sample Output for {label}:")
        display(image)

print("\nLab Experiment Complete: Synthetic dataset generated and
stored.")

```

Flax classes are deprecated and will be removed in Diffusers v1.0.0. We recommend migrating to PyTorch classes or pinning your version of Diffusers.

Flax classes are deprecated and will be removed in Diffusers v1.0.0. We recommend migrating to PyTorch classes or pinning your version of Diffusers.

/usr/local/lib/python3.12/dist-packages/huggingface_hub/utils/_auth.py :94: UserWarning:

The secret `HF_TOKEN` does not exist in your Colab secrets.

To authenticate with the Hugging Face Hub, create a token in your settings tab (<https://huggingface.co/settings/tokens>), set it as

```
secret in your Google Colab and restart your session.  
You will be able to reuse this secret in all of your notebooks.  
Please note that authentication is recommended but still optional to  
access public models or datasets.  
warnings.warn(  
  
{ "model_id": "322f49d26c4343288553948ccd7c1fef", "version_major": 2, "ver-  
sion_minor": 0}  
  
{ "model_id": "310152d742da4600a43d5ceae6dc5d84", "version_major": 2, "ver-  
sion_minor": 0}  
  
{ "model_id": "d0d1fc8da8884a8e8b4b2edd12c9ac95", "version_major": 2, "ver-  
sion_minor": 0}  
  
{ "model_id": "05844efdf0144c8abd5dcaf85791cfed", "version_major": 2, "ver-  
sion_minor": 0}  
  
{ "model_id": "f2e83a0c281e4ffd8226d004dc72fd29", "version_major": 2, "ver-  
sion_minor": 0}  
  
{ "model_id": "4e42d144a6c34elabe23636a8fa6ef7d", "version_major": 2, "ver-  
sion_minor": 0}  
  
{ "model_id": "875dc1d9b7f240a6bdcb08e8d24cf299", "version_major": 2, "ver-  
sion_minor": 0}  
  
{ "model_id": "11409742e4944d98b9474becbde2e26d", "version_major": 2, "ver-  
sion_minor": 0}  
  
{ "model_id": "c2b64f5881904182932b6206647ae94d", "version_major": 2, "ver-  
sion_minor": 0}  
  
{ "model_id": "6eb0f9031fd8469aa6e075383ced3fa8", "version_major": 2, "ver-  
sion_minor": 0}  
  
{ "model_id": "803522699f814b598b5b76abd0e956bc", "version_major": 2, "ver-  
sion_minor": 0}  
  
{ "model_id": "ccb7d46f263c47ab8dfd156b19230623", "version_major": 2, "ver-  
sion_minor": 0}  
  
{ "model_id": "3c14187b435e430da84226236a6cb663", "version_major": 2, "ver-  
sion_minor": 0}  
  
{ "model_id": "1ef1d15ab9cf40979f161b8350cfe332", "version_major": 2, "ver-  
sion_minor": 0}  
  
{ "model_id": "c05fda597c944edf80f2b40c4abe19da", "version_major": 2, "ver-  
sion_minor": 0}  
  
{ "model_id": "3f3107758c164deca4b9919d6ad7b68c", "version_major": 2, "ver-  
sion_minor": 0}
```

```
{"model_id": "00b7f785967346b0b8255a94ba0f4bb4", "version_major": 2, "version_minor": 0}
```

```
`torch_dtype` is deprecated! Use `dtype` instead!
```

```
Starting Synthetic Data Generation...
```

```
{"model_id": "67ec093b38ef440c8de58d267153e273", "version_major": 2, "version_minor": 0}
```

```
Saved: /content/xray_dataset/Normal_anatomy/Normal_anatomy_1.png  
Sample Output for Normal anatomy – healthy lungs, age & gender variations:
```



```
{"model_id": "492409b18deb4e1a8da46b6318acc49a", "version_major": 2, "version_minor": 0}
```

Potential NSFW content was detected in one or more images. A black image will be returned instead. Try again with a different prompt and/or seed.

```
Saved: /content/xray_dataset/Normal_anatomy/Normal_anatomy_2.png
```

```
{"model_id": "4139b52a9a9445448aca499aa7e2cff9", "version_major": 2, "version_minor": 0}
```

```
Saved: /content/xray_dataset/Normal_anatomy/Normal_anatomy_3.png
```

```
{"model_id": "a59bbada448b4eecb479cb8031005179", "version_major": 2, "version_minor": 0}
```

```
Saved: /content/xray_dataset/Normal_anatomy/Normal_anatomy_4.png
```

```
{"model_id": "02e0f6eeb6e844bb8731c2679e0b646d", "version_major": 2, "version_minor": 0}
```

Token indices sequence length is longer than the specified maximum sequence length for this model (79 > 77). Running this sequence through the model will result in indexing errors

The following part of your input was truncated because CLIP can only handle sequences up to 77 tokens: ['scan .']

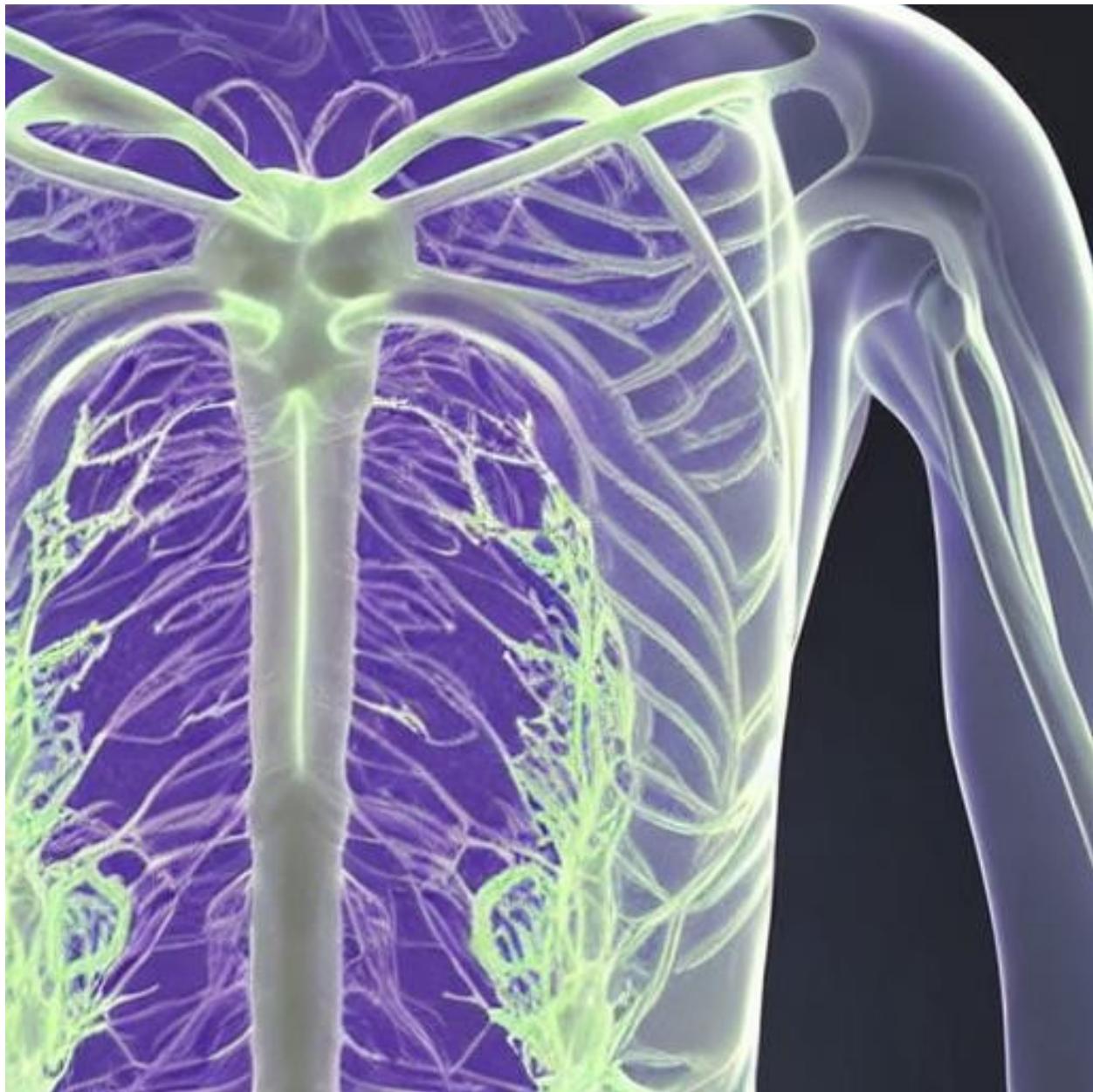
```
Saved: /content/xray_dataset/Normal_anatomy/Normal_anatomy_5.png
```

```
{"model_id": "ff69f493c5a14403931cdf6b14f667bd", "version_major": 2, "version_minor": 0}
```

```
Saved:
```

```
/content/xray_dataset/Infectious_patterns/Infectious_patterns_1.png
```

Sample Output for Infectious patterns – bacterial/viral pneumonia, COVID-like opacities:



The following part of your input was truncated because CLIP can only handle sequences up to 77 tokens: [']

```
{"model_id": "5a014f342cd6497b81d27c40949880d7", "version_major": 2, "version_minor": 0}
```

The following part of your input was truncated because CLIP can only handle sequences up to 77 tokens: ['medical scan .']

Saved:

/content/xray_dataset/Infectious_patterns/Infectious_patterns_2.png

```
{"model_id": "4651d485dad048a28882f33c50826ede", "version_major": 2, "version_minor": 0}
```

The following part of your input was truncated because CLIP can only handle sequences up to 77 tokens: ['.']

Saved:

```
/content/xray_dataset/Infectious_patterns/Infectious_patterns_3.png
```

```
{"model_id": "9d5ab881a7614c64alec94374e0eb0ff", "version_major": 2, "version_minor": 0}
```

The following part of your input was truncated because CLIP can only handle sequences up to 77 tokens: ['scan .']

Saved:

```
/content/xray_dataset/Infectious_patterns/Infectious_patterns_4.png
```

```
{"model_id": "552c7584ef0845d0b705a1116a6b3bdd", "version_major": 2, "version_minor": 0}
```

The following part of your input was truncated because CLIP can only handle sequences up to 77 tokens: ['professional medical scan .']

Saved:

```
/content/xray_dataset/Infectious_patterns/Infectious_patterns_5.png
```

```
{"model_id": "2b3335678bb7417a96d645601d9d06c5", "version_major": 2, "version_minor": 0}
```

Saved: /content/xray_dataset/Lung_opacities/Lung_opacities_1.png
Sample Output for Lung opacities – focal, diffuse, ground-glass, consolidations:



The following part of your input was truncated because CLIP can only handle sequences up to 77 tokens: ['medical scan .']

```
{"model_id": "af0e0c0df19243b8886214cde4cab7a2", "version_major": 2, "version_minor": 0}
```

The following part of your input was truncated because CLIP can only handle sequences up to 77 tokens: [, professional medical scan .']

Saved: /content/xray_dataset/Lung_opacities/Lung_opacities_2.png

```
{"model_id": "e579392650ed487ea3803e0d474e0505", "version_major": 2, "version_minor": 0}
```

```
The following part of your input was truncated because CLIP can only handle sequences up to 77 tokens: ['medical scan .']
```

```
Saved: /content/xray_dataset/Lung_opacities/Lung_opacities_3.png
```

```
{"model_id": "dc7734a8b39f451c8098e5ce07e0e232", "version_major": 2, "version_minor": 0}
```

```
The following part of your input was truncated because CLIP can only handle sequences up to 77 tokens: ['professional medical scan .']
```

```
Saved: /content/xray_dataset/Lung_opacities/Lung_opacities_4.png
```

```
{"model_id": "476a19255c5d4ec9a9ad5c148cd401ca", "version_major": 2, "version_minor": 0}
```

```
Saved: /content/xray_dataset/Lung_opacities/Lung_opacities_5.png
```

```
{"model_id": "c9a942a1942145eba6c9993586972e65", "version_major": 2, "version_minor": 0}
```

```
Saved:
```

```
/content/xray_dataset/Pleural_conditions/Pleural_conditions_1.png
```

```
Sample Output for Pleural conditions – pleural effusion, pneumothorax:
```



```
{"model_id": "0ca28e119fdd45cab79620dbcb70affd", "version_major": 2, "version_minor": 0}
```

The following part of your input was truncated because CLIP can only handle sequences up to 77 tokens: [']

Saved:

/content/xray_dataset/Pleural_conditions/Pleural_conditions_2.png

```
{"model_id": "c9146a8823db437c8f1466d94c7f1c02", "version_major": 2, "version_minor": 0}
```

```
Saved:  
/content/xray_dataset/Pleural_conditions/Pleural_conditions_3.png  
{"model_id": "0c6d6a7f4faa473f92696c5590024d1e", "version_major": 2, "version_minor": 0}  
  
Saved:  
/content/xray_dataset/Pleural_conditions/Pleural_conditions_4.png  
{"model_id": "65424f18de6d47998ebf068e6af30412", "version_major": 2, "version_minor": 0}  
  
Saved:  
/content/xray_dataset/Pleural_conditions/Pleural_conditions_5.png  
{"model_id": "1730f9eb0795464b8f90a5c534014e36", "version_major": 2, "version_minor": 0}  
  
Saved:  
/content/xray_dataset/Structural_lesions/Structural_lesions_1.png  
Sample Output for Structural lesions – nodules, masses, fibrosis:
```



```
{"model_id": "c6113ccf439b4349af5b8299d33671e7", "version_major": 2, "version_minor": 0}
```

Saved:

```
/content/xray_dataset/Structural_lesions/Structural_lesions_2.png
```

```
{"model_id": "9bf9ea035a0a40628f2aeeac882a9146", "version_major": 2, "version_minor": 0}
```

Saved:

```
/content/xray_dataset/Structural_lesions/Structural_lesions_3.png
```

```
{"model_id": "a92674dfc9c944fda0ee2664d2af5c7a", "version_major": 2, "version_minor": 0}
```

Saved:

```
/content/xray_dataset/Structural_lesions/Structural_lesions_4.png
```

```
{"model_id": "44e1d9b3b692436d8cfbba9705d26ecb", "version_major": 2, "version_minor": 0}
```

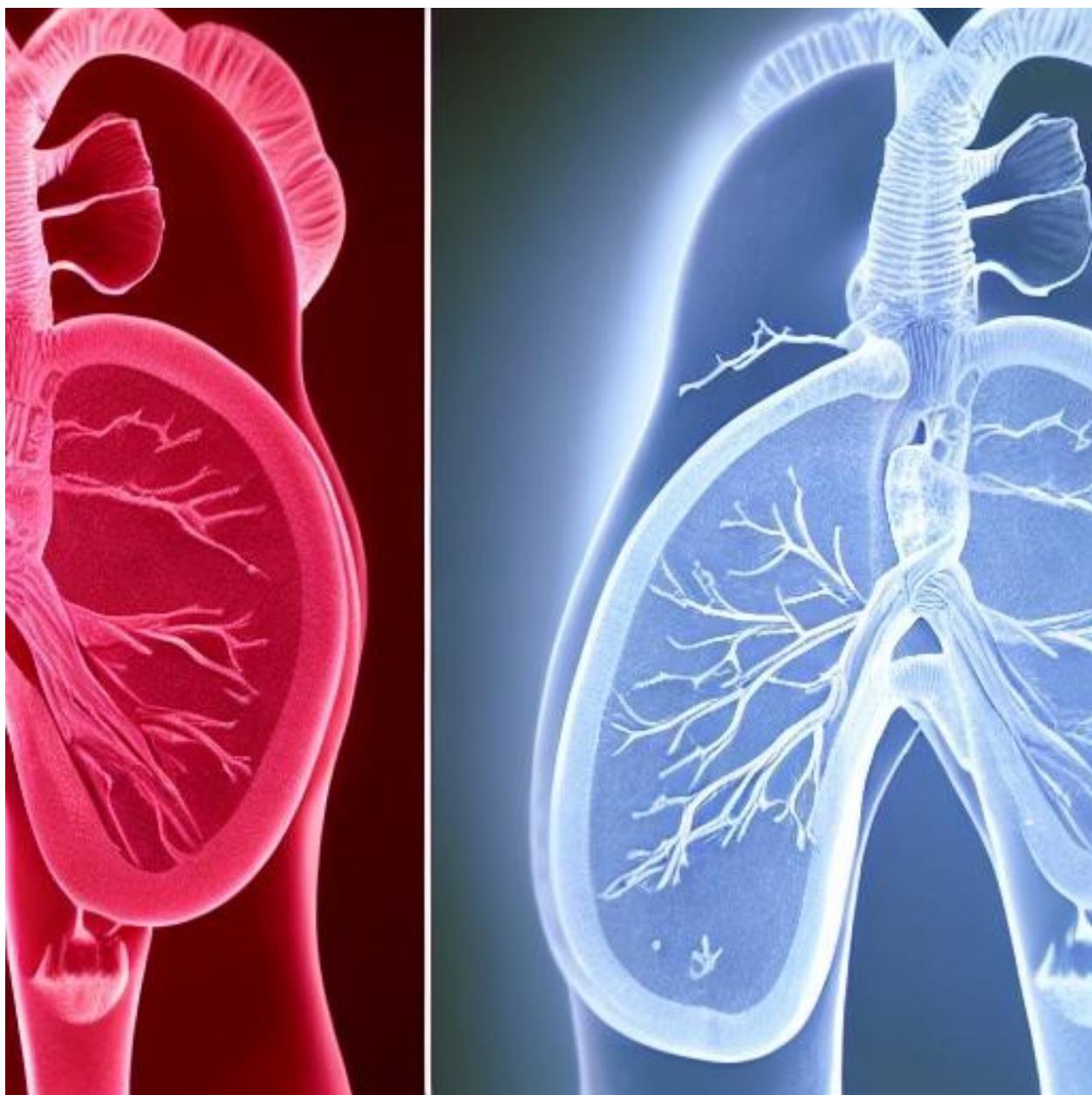
Saved:

```
/content/xray_dataset/Structural_lesions/Structural_lesions_5.png
```

```
{"model_id": "3a29f6f84c8d4823a2b9f334e3f254ef", "version_major": 2, "version_minor": 0}
```

Saved: /content/xray_dataset/Cardiac_findings/Cardiac_findings_1.png

Sample Output for Cardiac findings – cardiomegaly, vascular congestion:



```
{"model_id": "e199f5a14855406bb87d79bcfae11a72", "version_major": 2, "version_minor": 0}
```

Saved: /content/xray_dataset/Cardiac_findings/Cardiac_findings_2.png

```
{"model_id": "2dea4af375f84aeba7eb78b7947dff8c", "version_major": 2, "version_minor": 0}
```

Saved: /content/xray_dataset/Cardiac_findings/Cardiac_findings_3.png

```
{"model_id": "2c904e30c46f441a8adde3e859b620cf", "version_major": 2, "version_minor": 0}
```

```
Saved: /content/xray_dataset/Cardiac_findings/Cardiac_findings_4.png
{"model_id": "b1b46ba622a646bb8c352f3c25584c99", "version_major": 2, "version_minor": 0}

Saved: /content/xray_dataset/Cardiac_findings/Cardiac_findings_5.png
{"model_id": "0d178e98e006463db8f45abfef0cbe61", "version_major": 2, "version_minor": 0}

Saved: /content/xray_dataset/Medical_devices/Medical_devices_1.png
Sample Output for Medical devices – tubes, catheters, pacemakers:
```



```
{"model_id": "938d3f025f9446ad9d2e73124299d752", "version_major": 2, "version_minor": 0}

Saved: /content/xray_dataset/Medical_devices/Medical_devices_2.png

{"model_id": "77b82acf2b6c4264840e6342eb3a33a9", "version_major": 2, "version_minor": 0}

Saved: /content/xray_dataset/Medical_devices/Medical_devices_3.png

{"model_id": "8af43c23a02b47a3948bf8ddd7da102c", "version_major": 2, "version_minor": 0}

Saved: /content/xray_dataset/Medical_devices/Medical_devices_4.png

{"model_id": "e71e82e9ca244128ab8b96c9b3d91510", "version_major": 2, "version_minor": 0}

Saved: /content/xray_dataset/Medical_devices/Medical_devices_5.png

{"model_id": "1e6059fef69549b6ab6f8c29cdalced9", "version_major": 2, "version_minor": 0}

Saved: /content/xray_dataset/Imaging_artifacts/Imaging_artifacts_1.png
Sample Output for Imaging artifacts – noise, motion blur, exposure issues:
```



```
{"model_id": "5a422a923a2846e28d4713d495db4924", "version_major": 2, "version_minor": 0}
```

```
Saved: /content/xray_dataset/Imaging_artifacts/Imaging_artifacts_2.png
```

```
{"model_id": "00aec8bab868406fa3ad3e56dd4c2d0f", "version_major": 2, "version_minor": 0}
```

```
Saved: /content/xray_dataset/Imaging_artifacts/Imaging_artifacts_3.png
```

```
{"model_id": "4662da8eba23457894c258e0ef0167f0", "version_major": 2, "version_minor": 0}
```

```
Saved: /content/xray_dataset/Imaging_artifacts/Imaging_artifacts_4.png
```

```
{"model_id":"f1a211bb1a9e4961821925080052c457","version_major":2,"version_minor":0}
```

```
The following part of your input was truncated because CLIP can only handle sequences up to 77 tokens: ['medical scan .']
```

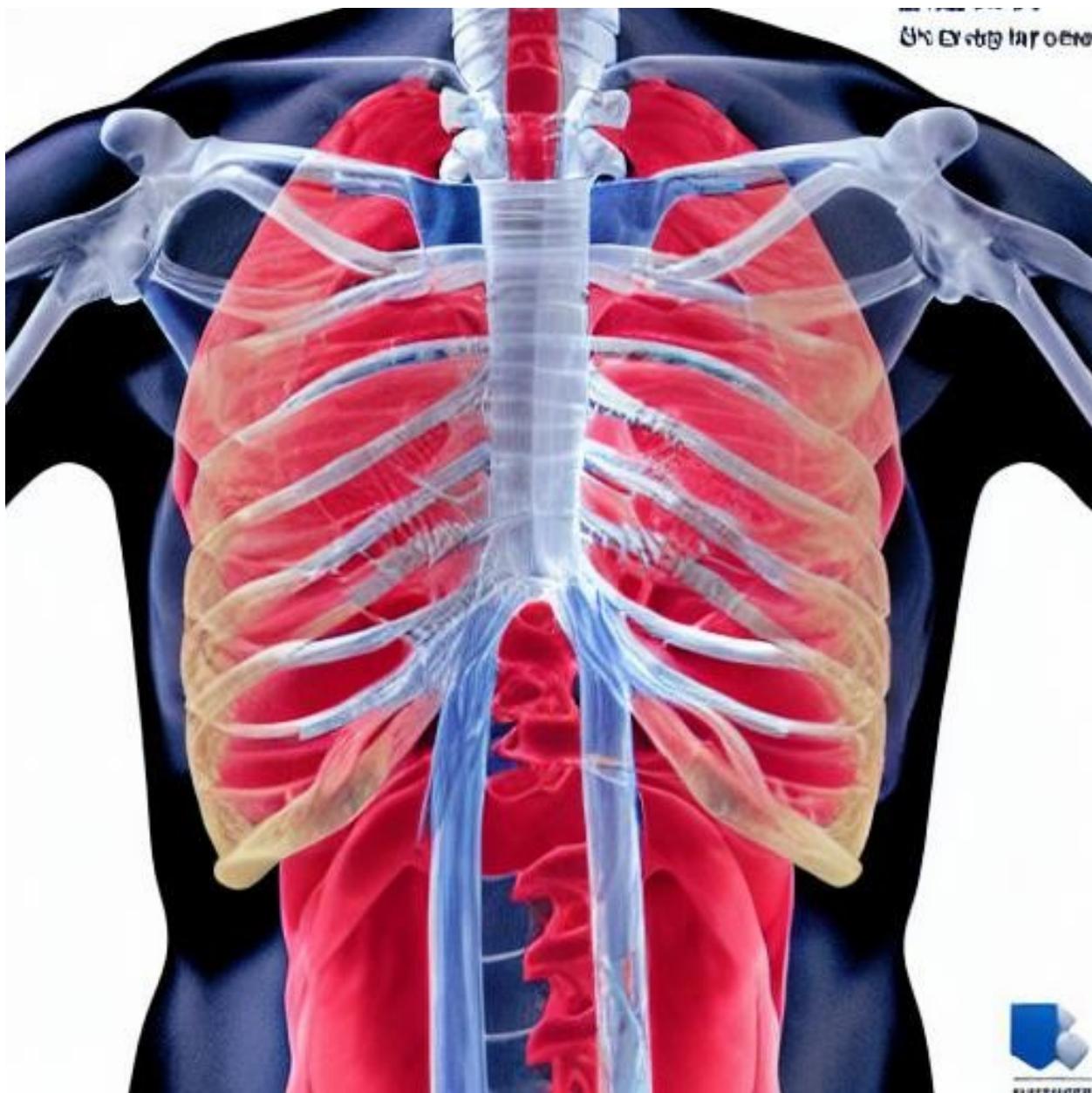
```
Saved: /content/xray_dataset/Imaging_artifacts/Imaging_artifacts_5.png
```

```
{"model_id":"c1a0d4de58c743258db549289eba9ddd","version_major":2,"version_minor":0}
```

```
Saved:
```

```
/content/xray_dataset/View__positioning/View__positioning_1.png
```

```
Sample Output for View & positioning – PA/AP views, rotation, supine/erect:
```



The following part of your input was truncated because CLIP can only handle sequences up to 77 tokens: ['scan .']

```
{"model_id": "444ab30810734871bab1266b83ef5491", "version_major": 2, "version_minor": 0}
```

The following part of your input was truncated because CLIP can only handle sequences up to 77 tokens: ['professional medical scan .']

Saved:

/content/xray_dataset/View__positioning/View__positioning_2.png

```
{"model_id": "285e0a3e0b434182b25bb8911b4609f3", "version_major": 2, "version_minor": 0}
```

The following part of your input was truncated because CLIP can only handle sequences up to 77 tokens: ['scan .']

Saved:

```
/content/xray_dataset/View__positioning/View__positioning_3.png
```

```
{"model_id": "bbc286ba348444a38420f139635d12cc", "version_major": 2, "version_minor": 0}
```

The following part of your input was truncated because CLIP can only handle sequences up to 77 tokens: ['medical scan .']

Saved:

```
/content/xray_dataset/View__positioning/View__positioning_4.png
```

```
{"model_id": "18ba78b583034f23887df88af8fea0a6", "version_major": 2, "version_minor": 0}
```

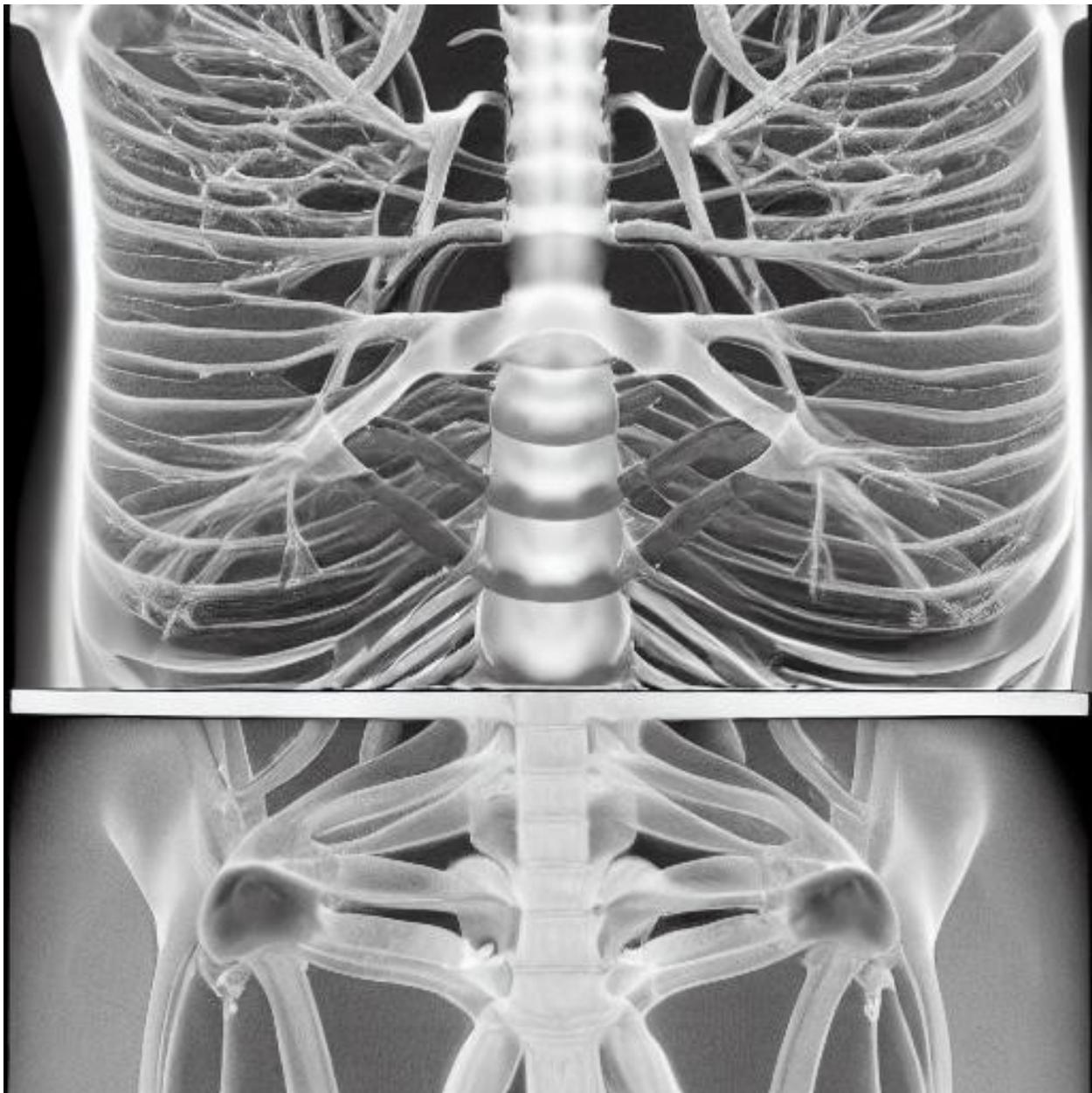
Saved:

```
/content/xray_dataset/View__positioning/View__positioning_5.png
```

```
{"model_id": "515ea99169a24e989ffcca02d3e3880d", "version_major": 2, "version_minor": 0}
```

Saved: /content/xray_dataset/Domain_shift/Domain_shift_1.png

Sample Output for Domain shift – scanner, hospital, and resolution variations:



```
{"model_id": "4d2d8e8ca9e04070a8dcba75781bada", "version_major": 2, "version_minor": 0}
```

Saved: /content/xray_dataset/Domain_shift/Domain_shift_2.png

```
{"model_id": "7f35143c14b848a0a814e0d79340e32e", "version_major": 2, "version_minor": 0}
```

Saved: /content/xray_dataset/Domain_shift/Domain_shift_3.png

```
{"model_id": "c40aeelcd5b14103ad436156607b2664", "version_major": 2, "version_minor": 0}
```

```
Saved: /content/xray_dataset/Domain_shift/Domain_shift_4.png
{"model_id": "22367f29cc144277a3430093c2c10a4b", "version_major": 2, "version_minor": 0}
Saved: /content/xray_dataset/Domain_shift/Domain_shift_5.png
Lab Experiment Complete: Synthetic dataset generated and stored.
```