Technical Blueprint — StackGAN-HIV (From Pitch to Prototype)

Date: Oct 18, 2025

A) Problem Context & Project Summary

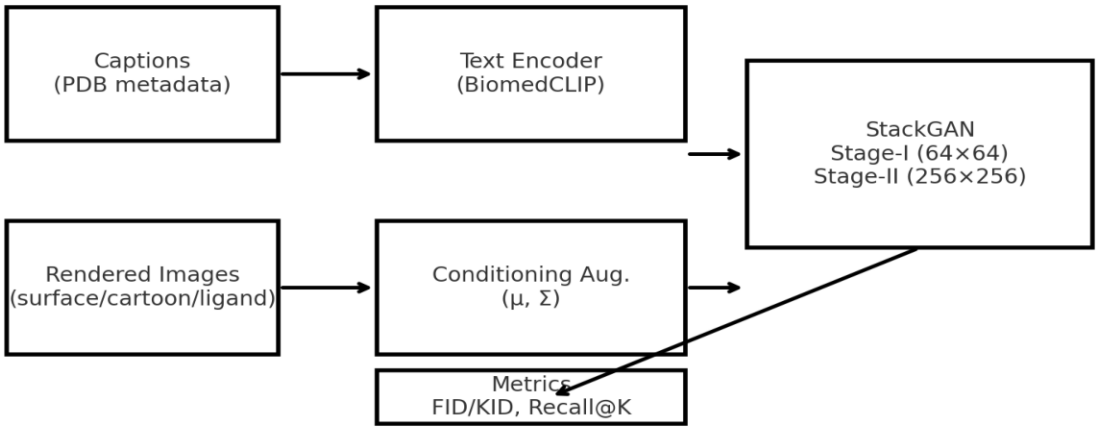
We will adapt the original StackGAN (two-stage GAN with conditioning augmentation) to synthesize publication-style molecular renderings from short HIV-related captions (CCR5/CXCR4, Env–CD4, RT/PR/IN complexes). Training pairs are created by auto-rendering CC0 Protein Data Bank (PDB) structures and converting PDB metadata into concise captions. The goal is an interactive tool where users type a caption and receive 256×256 figure panels that visually match HIV entry/drug-target descriptions, with quantitative evaluation via FID/KID and text–image Recall@K.

B) Dataset

• Core dataset: RCSB Protein Data Bank (PDB). Type: images (rendered PNGs) + text (auto captions). Size (first pass): ~300–600 images from 30–60 PDB entries × 6–10 views each. Access: public domain (CC0). Format: PNG images and a CSV pairs file: id, pdb_id, caption, image_path, view. • Optional: Human Protein Atlas IF/IHC images for CCR5/CXCR4 (CC BY-SA 4.0) for a small microscopy pack. Preprocessing: auto-render surfaces/cartoon/ligand sticks at 64×64 and 256×256; normalize filenames; generate captions via templates; split into train/val. Ethical/privacy: no human PHI; structural data is public-domain. For HPA images, include attribution and respect CC BY-SA.

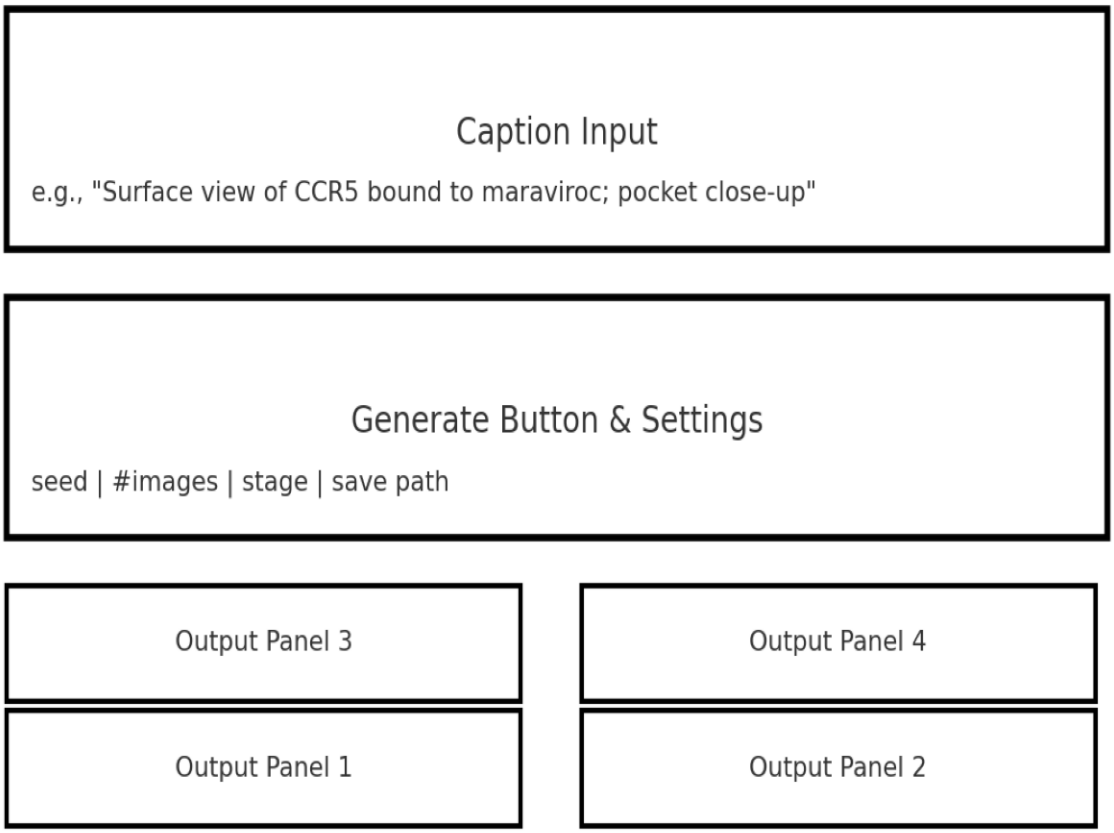
C) Planned Architecture

Data flow: Captions → Text Encoder (BiomedCLIP) → Conditioning Augmentation (μ, Σ) + noise → StackGAN Stage-I (64×64) → Stage-II (256×256) → UI. We will retain the original StackGAN losses and two-stage training. The interface (Streamlit/Gradio) accepts a caption and seed, then displays 2×2 synthesized panels with options to save.



D) User Interface Plan

Inputs: short caption, seed, number of images. Outputs: 2×2 grid of 256×256 images, plus nearest real reference renders. Usability: fast iteration on captions; compare against real renders to assess alignment; simple controls for reproducibility.



E) Innovation and Anticipated Challenges

Innovation: applying classic StackGAN to biomedical figure synthesis with a domain text tower (BiomedCLIP) and clean, auto-captioned data. Challenges: (1) Mode collapse on small datasets; mitigations: DiffAugment/ADA, TTUR, spectral norm, careful LR; (2) Caption drift; mitigations: templated captions + held-out validation; (3) GAN vs diffusion

F) Implementation Timeline

Week	Focus	Expected Outcome
Oct 20–26	Data scripts & pairs; environment setup	See README; Stage-I configs ready
Oct 27–Nov 2	Train Stage-I (64×64); baseline FID	See README for commands; checkpoints
Nov 3–9	Train Stage-II (256×256); add metrics	See README for evaluation; (Recall@K)
Nov 10–16	Tuning & stability; ablations	See README; frozen validation set
Nov 17–30	UI polish; nearest-real viewer; exports	See README for app flows
Dec 1–11	Final demo, ablations, write-up & presentation	See README & presentation

G) Responsible AI Reflection

Fairness & transparency: Images are synthetic and should be watermarked as such; we will display captions and random seeds for reproducibility and provide nearest real references. Licensing: PDB is CC0; any HPA images require CC BY-SA attribution. Environmental impact: Limit epochs/batch sizes; checkpoint and early-stop; prefer mixed precision where applicable.