1. What generative model(s) will your project focus on?

**Variational Autoencoders (VAEs)**

1. Enter 1-2 paragraphs describing which variants or aspects of these generative model(s) your project will focus on, and which learning algorithm(s) you plan to use or compare.

Our project focuses on **Variational Autoencoders (VAEs)**, beginning with the original formulation in Kingma & Welling (2014). We will start by implementing this standard VAE on the dSprites dataset, using a convolutional encoder/decoder.

We then extend the objective to the **β-VAE** framework (Higgins et al., 2017), weighting the KL divergence term by a factor β to encourage disentangled latent factors.All models will be trained with stochastic gradient descent (Adam optimizer), and we will compare constant-β vs. capacity-scheduling schemes.

1. Enter 1-2 paragraphs describing how the generative model(s) your project will focus on, and the learning algorithm(s) you plan to use, connect to the CS275P course material.  *Strong projects should have clear links to material from the CS275P lectures and readings.*

Both our baseline VAE and β-VAE extensions build directly on the **Variational Inference** and **Amortized Inference**material from the CS275P lecture on deep generative models.  In particular:

* **Variational Autoencoders** introduced the ELBO derivation and the “reparameterization trick” that we implement in our encoder network.
* **β-VAEs and Disentanglement** covers how scaling the KL term trades off reconstruction fidelity vs. latent factor independence, which we reproduce and extend.

1. What is the primary reference that your project will build on?  For *generative model application* projects, this could be a section in one of the course textbooks.  For *paper validation/extension* projects, this should be a citation to a published paper (with URL).

**Primary reference (paper extension):**Kingma, D. P. & Welling, M. (2014). *Auto-Encoding Variational Bayes*. arXiv:1312.6114.  
<https://arxiv.org/abs/1312.6114>

1. Provide at least one additional reference (with URL) related to the generative models and learning algorithms your project will use.  These may be published papers or textbook sections, but *not* blog posts.

**Additional reference (generative models & learning algorithms):**Higgins, I. et al. (2017). *beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework*.  
https://openreview.net/forum?id=Sy2fzU9gl

1. Provide at least one reference (with URL) related to the data that your project will use.  These may be published papers or technical reports, or a website where the data is discussed.

**Data reference:**Matthey, L. et al. (2017). *dSprites: Disentanglement testing Sprites dataset*.  
<https://github.com/deepmind/dsprites-dataset>

1. Enter 1-2 paragraphs reviewing prior work related to your project.  This could be either work that applies the same generative models to similar data, or work that applies different machine learning methods to the same data.

Early work on VAEs (Kingma & Welling, 2014; Rezende et al., 2014) showed that deep amortized inference can scale probabilistic models to high-dimensional data. But the standard VAE often learns entangled latents, making it hard to understand and control. Higgins et al. (2017) introduced the β-VAE, which shows that up-weighting the KL term encourages latent factor independence on simple image datasets (e.g. dSprites). Burgess et al. (2018) built on this by scheduling a target KL capacity, which improves the trade-off between disentanglement and reconstruction. Recent metrics like the Mutual Information Gap (Chen et al., 2018) have given us quantitative ways to evaluate disentanglement, which we’ll use in our experiments.

1. Enter 1-2 paragraphs providing more details about the dataset(s) you plan to use, and why they are appropriate for the proposed generative models.  Also provide a high-level description of how you will evaluate your results.  *Both quantitative (computing a performance metric on test data) and qualitative (visualizing the learned generative model structure) evaluations are useful.*

We will use the **dSprites** dataset, which consists of 64×64 binary images of 2D shapes (square, ellipse, heart) parameterized by five ground-truth factors (shape, scale, orientation, x-pos, y-pos).  This synthetic dataset is ideal for disentanglement experiments because the true generative factors are known and easily visualized.

**Evaluation plan:**

* **Quantitative**: Compute disentanglement metrics (β-VAE metric, MIG, FactorVAE score) on a held-out subset.  Track reconstruction error (binary cross-entropy) and KL divergence separately.
* **Qualitative**: Perform latent-space traversals by varying one latent dimension at a time and observing the decoded image grid.  We will visualize changes in each generative factor to confirm disentanglement.

1. Upload a figure in PDF format illustrating a preliminary experiment with some data related to your project. This could be some sort of visualization of the raw data, or the results of running a baseline machine learning method. *The preliminary experiment may use the generative model that will be the main focus of your project, but it does not have to.*

Enter 1-2 paragraphs describing your preliminary experiment, and your interpretation of the results in the figure.

[pending: Conclusion needs to be drawn from plot]

* + How fast the reconstruction loss decreases
  + The growth pattern of the KL term
  + What the combined ELBO curve tells you about convergence
  + Any early signs of latent‐factor control in your traversal visualizations

**1. Train Loss per Epoch**

* **What it shows:**
* **Curve behavior:**
* **Interpretation:**

**2. Test Average ELBO per Epoch**

* **What it shows:** the estimated ELBO (i.e. reconstruction – KL, acting as a lower bound on log-likelihood) on the *held-out* test set after each epoch.
* **Curve behavior:**
* **Interpretation:**

(To be referenced from) an increasing ELBO means the model’s bound on the data log-likelihood is improving— the VAE is not just memorizing the training data but also generalizing better to unseen examples.  Like the training loss, the biggest improvement is early on, with smaller gains by the higher epoch.