

Appendix A

Glossary of Terms

AI	Artificial Intelligence.
ADD	Adversarial Diffusion Distillation. A combination of aspects of Generative Adversarial Networks (GANs) and Diffusion Models.
API	Application Programming Interface. A defined protocol that allows two computer programs to communicate and exchange data.
Autoencoder	An autoencoder is a system that can encode an image from a “pixel space” representation of an image into a representation in “latent space” (the encoder role) or from latent space back into pixel space (the decoder role).
AWS Cluster	Amazon Web Services Inc cloud hosting and processing resources.
Batch/Batch size	In the context of training Stable Diffusion, the number of training images processed by the model during a single training step or iteration.
Bias	See also “weights”. A constant added to the sum of the weighted inputs of a neuron (or node) in a neural network which affects the output. The weights and biases are sometimes just called the “weights” of a model.
Checkpoint	A file recording the current state of training of a machine-learning model. Further training can continue from that checkpoint. Versions of Stable Diffusion may be released at a particular “checkpoint” and then training continued for later versions. Versions v1.1, v1.2, v1.3 and v1.4 of Stable Diffusion are both checkpoints and versions.
CLIP/CLIP text encoder	Contrastive Language-Image Pre-Training. It is a deep-learning model that has learned the relationship between images and text. It has an image encoder and a text encoder.
Diffusion Model	An AI system that generates an output, such as an image, by a process of starting with an image consisting of noise, or an image to which noise has been added, and progressively removing the noise to generate an output image.
Deep Learning	A procedure to optimise both the decision function and the underlying input representation based on data.
Denoising	The process of removing noise from an image.
Epoch	A complete pass of all the data in a training dataset during a training process.

GANs	Generative Adversarial Networks. Unlike Diffusion Models, GANs can produce images in a single step but include drawbacks such as lower sample quality.
Generalisation	A network such as Stable Diffusion must be able to perform well on new data. This means it must generalise – i.e. recognise and understand the general patterns and rules in the training data and be able to apply them in a different context.
Generative AI	Machine Learning models that generate content such as images or text as their output.
GPU	Graphics Processing Unit.
Gradient Descent	A method of training in which a model plots a route from a given starting point (i.e., the model's unoptimized starting loss function) to the lowest point (i.e. a minimum of the cost function) by taking the steepest descending local gradient.
Inference	The process of running a machine learning model to produce outputs (images in the case of Stable Diffusion).
Iteration	The process by which the model being trained generates an output from a training data sample (or batch of samples), calculates the loss function, calculates the gradient of the loss function on a layer-by-layer basis, and updates the weights to achieve the gradient descent step.
Latent space	An image in “latent space” is an image that has been converted from an image in pixel space using an autoencoder. Latent space is a compressed, representative form of the pixel space image that is more memory- and computationally-efficient.
Layers	A collection of neurons or nodes. In a neural network, information flows from layer to layer. Each layer is composed of a collection of neurons or nodes. There is an “input layer” and an “output layer”, and there may be intermediate layers, known as “hidden layers”.
Loss Function	A quantification of the difference between the target or expected output of a model and the actual output. Loss function is used in a model's training process in determining adjustments to its weights that should result in a reduced loss function.
Machine Learning	Given a task to solve, a machine-learning system is designed to infer a solution from data without being instructed explicitly on how to solve the problem. The system is trained by providing it with data from which it can “learn” to carry out tasks not explicitly provided for by the software engineer's design. Machine learning is part of the wider field of AI.
Memorisation	This process occurs in image models such as Stable Diffusion when the model is over-exposed to one specific training image linked to a particular concept. The process of memorisation undermines the model's ability

	to generalize and generate new content when prompted with a text prompt for that concept.
Model Card	A short, standardised document that provides detailed information about the Model.
Neural Network	A machine learning model consisting of a network of neurons (or nodes).
Noise	In the context of an image, noise can take the form of random pixel values. An existing image can be partially corrupted with noise, or an image can consist entirely of noise.
Open source	Software whose source code is made freely available for anyone to use.
Overfitting	A situation where the model produces outputs that fit too closely to the training samples rather than generalising from them.
Pixel/Pixel space	A digital image suitable for human viewing may be represented by a two-dimensional array of “pixels”. An image in “pixel space” is an image in this form.
Prompt	A descriptive text caption or an image (or both) input by a user to condition the output of Stable Diffusion.
Sample/Sampling	The value drawn from the probability distribution. Drawing a value from a specified probability distribution.
Stochastic system	A system in which the same input may not produce the same output because of some randomness or probabilistic behaviour of the system.
Seed/Seed value	In Stable Diffusion a “seed” is a latent image used as the starting point for the inference process. The seed may be pure random noise or a user image to which noise has been added. In the case of text-to-image inference, the user can specify the “seed value” (a numeric value) in which case the same seed image will be used each time, and the output will be identical or practically so.
Step	In the context of training a model like Stable Diffusion a “step” is the process of adjusting the model weights at the end of each iteration. It can also be used synonymously with “iteration”. See also “Iteration”.
Text Encoder	The text encoder takes a sequence of tokens (numbers representing words or parts of words) and converts them using a deep learning model to form a “text-embedding” or an “embedding” which is designed to represent the semantic meaning of the words.
Training	The process by which a machine-learning model “learns” its model weights from user-provided data.
Tokens/Tokeniser	A tokeniser converts text into tokens, which are numbers representing words or parts of words.
U-Net	A neural network that, in the case of Stable Diffusion, is responsible for “denoising” the image in latent space. The denoising process may be conditioned by

	user prompts to influence the resulting image.
URL	Uniform Resource Locator – a hyperlink (via a web address) to the original internet location at which the referenced image is made available to the public by the website operator.
Vector File	A resolution independent depiction of a visual image that can be scaled up or down and is often used for illustrations.
VRAM	Video Random Access Memory.
Weights	In a neural network a neuron (or node) has inputs (which may come from outside the network or from a previous layer of the network) and an output. A “weight” determines the strength of the effect a particular input has on the output. See also “bias”. The weights and biases are sometimes just called the “weights” of a model. The aim of training is to adapt the weights and biases such that the model produces desired outputs.