| --- | --- | --- | --- | --- | --- | --- | --- | --- |
1	N/A	cell-6-output2.png	hogaldm_01in01.png	As-is. We may want to size it down to fit on the prior page				
2	An example of character-level tokenization. no draft, see Col F for description	hogaldm_02in01.png	To be drawn	An image with a sentence "A llama learns running a llama farm", with each letter having a number assigned. The same letter always has the same number. For this case, use Character "A", ID: 1 / Character " ", ID: 0 / Character "l", ID: 12 / Character "l", ID: 12 / Character "a", ID: 13 / Character "m", ID: 1 / Character "a", ID: 13 / Character "w", ID: 5 / Character "a", ID: 1 / Character "l", ID: 18 / Character "r", ID: 14 / Character "n", ID: 19 / Character "s", ID: 0 / Character "r", ID: 18 / Character "u", ID: 21 / Character "n", ID: 7 / Character "n", ID: 19 / Character "i", ID: 0 / Character "n", ID: 12 / Character "g", ID: 12 / Character "a", ID: 1 / Character "l", ID: 13 / Character "l", ID: 1 / Character "a", ID: 6 / Character "m", ID: 18 / Character "a", ID: 13 / Character "f", ID: 0			
2	An example of word-level tokenization. The no draft, see Col F for description	hogaldm_02in02.png	Redraw	An image with a sentence "A llama learns running a llama farm", with each word having a number assigned. The same word always has the same number. For this case, use A - 8428 / llama - 30913 / learns - 12019 / running - 7432 / a - 49 / llama - 30913 / farm - 21835			
2	An example of sub-word tokenization. Llama no draft, see Col F for description	hogaldm_02in03.png	Redraw	An image with a sentence "A llama learns running a llama farm", with the sentence split in "a", "lla", "ma", "learn, "s", "running", "a", "llama", "farm". A - 49 / lla - 303 / ma - 4530 / learn - 16312 / s - 35 / running - 4108 / a - 81 / lla - 304 / ma - 4530 / farm - 2769			
2	Another greedy decoding example. The do no draft, see Col F for description	hogaldm_02in04.png	Redraw	An image showing a tree of potential generations with probabilities assigned to each sequence. The tree root is "the". The three words are "dog" (0.5), "horse" (0.4), and "llama" (0.1). For "dog", next words are "barks" (0.4), "runs" (0.3), and "bites" (0.3). For horse, next words are "runs" (0.9), "head" (0.05) and has (0.05). For llama, next words are "spits" (0.6), "runs" (0.3) and "eats" (0.1). The sequence "The dog barks" is highlighted with a total probability of 0.2 (0.5*0.4). Inspired/similar to [link]	Please see PDF		
2	A beam search example which finds a mo no draft, see Col F for description	hogaldm_02in05.png	Redraw	Same tree as before, but highlighting "The horse runs" Inspired/similar to [link]	Please see PDF		
2	Greedy generation will always pick the mo no draft, see Col F for description	hogaldm_02in06.png	Redraw	An image of a probability distribution chart with bars representing the probabilities of potential next tokens: 60% for "night," 35% for "day," and 5% for "apple." Below the chart, 1. there is a label that reads "Greedy," indicating that the greedy decoding method always selects "night" (the highest probability). 2. below it, there's another label that reads "Sampling", it has three sentences starting with "the night" and one sentence starting with "the day" Not the same, but this img can help [link]	Please see PDF		
2	Effect of temperature on the token probabil no draft, see Col F for description	hogaldm_02in07.png	Redraw	An image with a distribution similar to the one at the left of [link] (but with no x-axis). There are three arrows leading to new distributions. The top one, labeled with t=0, has all probability in the leftmost token. The one at the middle, with t=0.4, has more probability in the left tokens (makes distribution sharper). The bottom one, with t=3, has a more uniform distribution.) To make an idea, please see our generated images temperature_zero, temperature_0.4, and temperature_3 at [link] Similar to above, but now just two arrows. The top one, labeled with top_k=5, has a box at the left for the most likely 5 tokens. The bottom one, labeled with top_p=0.94, has a box at the left for the tokens that sum up to 0.94 probability. To make an idea, use our generated images top_k_sampling and top_p_sampling in the directory [link]		Instructions pending (requested from authors)	
2	Effect of top_k and top_p on the token probno draft, see Col F for description	hogaldm_02in08.png	Redraw		Please see PDF		
2	Architecture of a transformer-based langua transformers/transformer_arch.png	hogaldm_02in09.png	Redraw		Please see PDF		
2	A diagram of the encoder-decoder transfor no draft, see Col F for description	hogaldm_02in10.png	Redraw	The idea is to do something along the lines of Figure 1 from [link] Here is an example [link] A sequence (with one token in the middle being MASK) goes through a transformer encoder, which outputs an embedding for each token. The embedding corresponding to MASK goes through an additional layer and finally a softmax, from which it outputs the word.	Please see PDF		
2	Encoder models output semantic embeddin no draft, see Col F for description	hogaldm_02in11.png	Redraw	For example, input at the bottom: "A [MASK] was barking a lot" From this text, a line to a (tall) block that says "transformer" From the block, the output is a small block that says "embedding". The embedding is really a list of numbers, e.g. 0.173, 0.013, ... From the embedding, there's a row to a single row block. And from the block there's a text that says softmax Finally, from the softmax there's an arrow and the word "dog"		Please see PDF.	
2	Popular model releases no draft, see Col F for description	hogaldm_02in12.png	Redraw	We want to discuss a bit more here on what could be done. TL;DR we have a timeline of model releases similar to [link] but we want to see what can be done here A redrawn version of [link]		Please see PDF.	
2	While pre-training a base model can require no draft, see Col F for description	hogaldm_02in13.png	Redraw	We just want to make sure it doesn't take too much space on the page	Please see PDF		
2	N/A	image-classification-input.jpeg	hogaldm_02in01.png	Redraw	A gallery of 3 images. In the first one, we showcase image classification. There's an image with a cat and a label "cat". In the second image, there are multiple objects, and there's a segmentation on top of it. The third image shows captioning. For inspiration Image classification: see image at top of [link]. The input is an image (e.g. a cat), which goes through a model, and the model outputs a label (e.g. "cat" for simplicity). I would suggest to just have the input image (a cat) and a label "label: cat" Image segmentation: This can be the image output at the right of [link] (https://huggingface.co/tasks/image-segmentation) (https://huggingface.co/datasets/huggingface/documentation-images/resolve/main/image-segmentation/image-segmentation-output.png) Image captioning: this could be an image (e.g. a picture such as the one in [link]) with a text below with a description. Unlike classification, the label should actually be a "description": a herd of giraffes and zebras grazing in a field			
2	Transformers models can be used for differ need file for this	hogaldm_02in14.png	Redraw	Three images in one. At the left, we have an image at the left, an Autoencoder at the middle, and reconstructed image at the right. In the middle, we have an image, which goes to an encoder, which goes to a single layer network which outputs a class, e.g. "dog". Finally, the third contains just a decoder; its input is pure noise and the output is a dog. Draft sketch in first slide of [link]	Please see PDF		
1	Efficient data representation methods (left) need image for this, it sounds like there are image components so this isn't actually a 100% redraw	hogaldm_03in01.png	Redraw		Please see PDF		
1	Autoencoder: conceptual architecture diagr Autoencoder.png	hogaldm_03in02.png	Redraw				
3	N/A	cell-6-output-1	hogaldm_03in01.png		Per authors, this can be sized down a bit, so it doesn't take up so much space on the page.			
3	N/A	cell-7-output-1.png	hogaldm_03in02.png	As-is				
3	N/A	cell-9-output-1.png	hogaldm_03in03.png	As-is				
3	N/A	cell-14-output-1	hogaldm_03in04.png	As-is (pdf)				
3	N/A	2-7.pdf	hogaldm_03in05.png	As-is (pdf)				
3	N/A	3-8.pdf	hogaldm_03in06.png	As-is (pdf)				
3	Activation functions: ReLU vs Sigmoid	3-9.pdf	hogaldm_03in03.png	As-is (doesn't have to be as big)	Please see PDF			
3	N/A	3-10.pdf	hogaldm_03in07.png	As-is (pdf)				
3	N/A	3-11.pdf	hogaldm_03in08.png	As-is (pdf)				
3	N/A	3-12.pdf	hogaldm_03in09.png	As-is (pdf)			Replacement provided by ALbi (3-12.pdf); see also PDF	
3	N/A	3-13.pdf	hogaldm_03in10.png	As-is (pdf)			Replacement provided by ALbi (3-13.pdf); see also PDF	
3	N/A	3-14.pdf	hogaldm_03in11.png	As-is (pdf)			Replacement provided by ALbi (3-14.pdf); see also PDF	
3	Conceptual diagram of a VAE. VAEs learn compressing_info/vae-diagram.png	hogaldm_03in04.png	Redraw		Please see PDF		
3	N/A	3-16.pdf	hogaldm_03in12.png	As-is				
3	N/A	3-18.pdf and 3-19.pdf	hogaldm_03in05.png	As-is	This is 3-18.pdf merged with 3-19.pdf; 3-17.pdf was cut for space	Please see PDF		
3	N/A	3-20.pdf	hogaldm_03in13.png	As-is				
3	N/A	3-21.pdf	hogaldm_03in14.png	As-is			Replacement provided by ALbi (3-21.pdf); see also PDF	
3	N/A	3-22.pdf	hogaldm_03in15.png	Redraw, stack images together			Replacement provided by ALbi (3-23.pdf); see also PDF	
3	N/A	3-23.pdf	hogaldm_03in16.png	As-is				
3	N/A	3-24.pdf	hogaldm_03in17.png	As-is				
3	N/A	3-25.pdf	hogaldm_03in18.png	As-is				
3	CLIP: Contrastive Pre-Training Process. In need image file...what post?	hogaldm_03in06.png	Redraw / permission to use	Refers to the "contrastive pre-training" figure from the post (https://openai.com/index/clip/) The image is a [link]			
3	Photo of a cute lion cub behind a branch	compressing_info/lion.jpg	hogaldm_03in07.png	As-is				
3	N/A	cell-74-output-1	hogaldm_03in19.png	As-is				
4	Progressive denoising process.	iterative_denoising	hogaldm_04in01.png	As-is	An image of iterative refinement. There's a noisy input image on the left, and progressively there is less noise until there's a nice denoised image			
4	N/A	4-3-1.pdf / 4-3-2.pdf / 4-3-3.pdf / 4-3-4.pdf	hogaldm_04in02.png	Redraw, stack 4 images together		Please see PDF		
4	Augmentation creates more data from the need image file	hogaldm_04in02.png	Redraw	An image undergoing 3 different transformations: in one we horizontally flip the image, in the second we rotate it 90 degrees, and in the third we apply a translation operation. The source image is a butterfly, with a different arrow to each of the transformed versions. We suggest to take a single butterfly (e.g. from figure 4-5 or one such as [link]) That's the source image. From it, there should be 3 arrows: 1. Horizontal flip: this is the same image but flipped 2. Rotation: same original image but rotated 90 degrees 3. Translation: same image but translated This is similar to the next image, although ours should have just the 3 above [link]	Please see PDF		
4	N/A	4-5.pdf	hogaldm_04in03.png	As-is				
4	N/A	4-6.pdf	hogaldm_04in04.png	As-is	Links for more UNet diagrams: https://www.assemblyai.com/blog/minimagen-build-your-own-imagen-text-to-image-model https://www.assemblyai.com/blog/content/2022/07/arch-4.png https://media.geeksforgeeks.org/wp-content/uploads/20200619121251/Group14.jpg https://miro.medium.com/v2/resize/fit:1400/0*NjmV21H42iWsx-4.jpg Same template as 4-21, perhaps this one should be a simplified version. AU: How specifically do you want it simplified? And 4-21 will be the full version? Note from Omar: I would do same as 4-21			
4	Architecture of a simplified UNet.	diffusion_models/unet.png	hogaldm_04in03.png	Redraw		Please see PDF		
4	N/A	4-8.pdf	hogaldm_04in05.png	As-is		Please see PDF		
4	N/A	cell-16-output-2	hogaldm_04in06.png	As-is				
4	N/A	cell-17-output-2	hogaldm_04in07.png	As-is				
4	CNN network used to extract feature maps diffusion_models/film.png	hogaldm_04in04.png	Redraw				
4	The general principles of diffusion work fro no draft, see Col F for description	hogaldm_04in05.png	Redraw / permission to use	An image from figure 1 of the Cold Diffusion paper (https://arxiv.org/pdf/2208.09392). We need to ask authors for permission	Please see PDF		
4	N/A	4-20-output-1	hogaldm_04in08.png	As-is				
4	N/A	4-21-output-1	hogaldm_04in09.png	As-is				
4	N/A	cell-22-output-1	hogaldm_04in10.png	As-is				
4	N/A	4-16.pdf	hogaldm_04in11.png	As-is				
4	N/A	4-17.pdf	hogaldm_04in12.png	As-is				
4	N/A	4-18.pdf	hogaldm_04in13.png	As-is				
4	N/A	4-19.pdf	hogaldm_04in14.png	As-is		Please see PDF		
4	N/A	4-20.pdf	hogaldm_04in15.png	As-is		Please see PDF		
4	Architecture of a basic UNET.	diffusion_models/unet.png	hogaldm_04in07.png	Redraw	duplicate or modification of 4-7, details TBD note from omar: same!	Please see PDF		
4	Loss and generations of a basic UNet.	4-22.pdf	hogaldm_04in08.png	As-is		Please see PDF		
4	Loss and generations from the _diffusers_ l 4-23.pdf	hogaldm_04in09.png	As-is				
4	Comparing UNet with UViit and RIN. need image file	hogaldm_04in10.png	Redraw	Template / guidelines to be provided. Initial ideas in slide 2 of [link]	Please see PDF		
4	Comparing +ape+ vs +all+ vs +u+ objectiv cell-29.pdf	hogaldm_04in11.png	As-is		Please see PDF		
5	N/A	5-1.pdf	hogaldm_05in01.png	As-is				
5	N/A	cell-7-output-1	hogaldm_05in02.png	As-is				
5	N/A	5-3.pdf	hogaldm_05in03.png	As-is				
5	N/A	cell-12-output-2	hogaldm_05in04.png	As-is				
5	N/A	cell-13-output-2	hogaldm_05in05.png	As-is				
5	N/A	cell-14-output-2	hogaldm_05in06.png	As-is				

Chapter #	Caption	Original Filename	New Filename	Redraw, or use As-Is?	Redraw / Label Notes (if applicable)	Initial PE Notes	Pre-QC1 edits 10/23	QC1 edits 11/6
5	The Latent Diffusion process. Note the VAE	no draft, see Col F for description	hogaldm_0501.png	Redraw	Placeholder: Image from Latent Diffusion Models - https://github... Note the VAE encoder and decoder on the left for translating between pixel space and latent space			
5	N/A	cell-16-output-2	hogaldm_0507.png	As-is			Please see PDF	
5	The UNet can be conditioned on multiple	ch_4_stable_diffusion/simplified_unet.png	hogaldm_0502.png	Redraw	Redraw for consistency with the rest of the book		Please see PDF	
5	The text encoding process transforms the	ch_4_stable_diffusion/text_encoder.png	hogaldm_0503.png	Redraw	Positional Embeddings should be renamed to Positional Encodings		Please see PDF	
5	The VAE architecture	ch_4_stable_diffusion/vae.png	hogaldm_0504.png	Redraw	Redraw for consistency with the rest of the book		Please see PDF	
5	N/A	cell-19-output-1	hogaldm_0508.png	As-is				
5	N/A	5-13.pdf	hogaldm_0509.png	As-is				
5	N/A	cell-22-output-1	hogaldm_0510.png	As-is				
5	Conditioned UNet architecture. At the left	ch_4_stable_diffusion/unet.png	hogaldm_0505.png	Redraw	To be simplified and consistent		Please see PDF	
5	Images generated with SDXL	cat_moon.jpeg	hogaldm_0506.png	As-is	Caption: "SDXL generated image with the prompt 'a cat playing with a miniature moon toy'"			
5	N/A	5-17.pdf	hogaldm_0511.png	As-is				
5	N/A	5-18.pdf	hogaldm_0512.png	As-is				
5	An explosion of creativity in the text-to-image	cell-30-output-1	hogaldm_0507.png	As-is				
5	N/A	ch_4_stable_diffusion/gradio.png	hogaldm_05or13.png	As-is				
6	The usual steps of a fine-tuning workflow.	no draft, see Col F for description	hogaldm_0601.png	Redraw	A flow chart with the following 6 steps 1. Identify dataset 2. Define model type 3. Select base model 4. Pre-process dataset 5. Define evaluation metrics 6. Train 7. Share		Please see PDF	
6	BERT with a classification head. In practice	no draft, see Col F for description	hogaldm_0602.png	Redraw	A BERT model with a layer on top of it. The input is the text, then there is the transformer block, which has embeddings at the end, and finally a layer			
6	N/A	6-3.pdf	hogaldm_06or01.png	As-is				
6	With instruction-tuning, we can format many	no draft, see Col F for description	hogaldm_0603.png	Redraw	Image similar to figure 1 from https://arxiv.org/pdf/2210.11416		Please see PDF	
6	LoRA reduces the number of trainable weights	n-diagram.png	hogaldm_0604.png	Redraw	To be simplified. This is same as 7-6! Note from Omar: They should be the same		Please see PDF	
6	Different precisions.	no draft, see Col F for description	hogaldm_0605.png	Redraw	Image of different precisions (float 32, float 16, bfloat16, int8, int4). Have FP32, FP16, and bfloat16 based on...		Please see PDF	
6	Model offloading.	no draft, see Col F for description	hogaldm_0606.png	Redraw	For int8 and int4, just have green 4 or 8 bits + 1 for sign. Image showing a model that requires 50GB...		Please see PDF	
6	A screenshot of the LMSYS Leaderboard thing	lmsys.png	hogaldm_0607.png	As-is				
6	A RAG pipeline.	no draft, see Col F for description	hogaldm_0608.png	Redraw	At the left, a user inputs a question about a movie. The query goes to a retrieval box connected to a database box. The retrieval box is connected to an LLM box, which outputs an answer.		Please see PDF	
7	Stable Diffusion Fine-Tuning architecture diag	dsd_fine_tuning.png	hogaldm_0701.png	As-is / Redraw	Can be redrawn for harmonization with the rest of the book, but also can be kept as is (original)		Please see PDF	
7	N/A	cell-6-output-2	hogaldm_07or01.png	As-is				
7	Technical diagram for Dreambooth architecture	dreambooth_flow.png	hogaldm_0702.png	As-is / Redraw	Can be redrawn for harmonization with the rest of the book, but also can be kept as is (we have authorization from the authors)		Please see PDF	
7	Training set for faces of "Apollonio Passerotti"	grid.png	hogaldm_0703.png	As-is				
7	Output images generated from the model "fpol	dreambooth.png	hogaldm_0704.png	As-is	Can be redrawn for harmonization with the rest of the book, but also can be kept as is (we have authorization from the authors). This is same as 6-7! Details TBD whether this should be a more complex version of 6-7.			
7	Technical diagram for LoRA architecture	lora-diagram.png	hogaldm_0705.png	As-is	Note from Omar: this should be exactly the same as 6-7		Please see PDF	
7	N/A	cell-10-output-2	hogaldm_07or02.png	As-is				
7	Example inference of Stable Diffusion 2 fine	depth_model_example.png	hogaldm_0706.png	As-is / Redraw	Can be redrawn for harmonization with the rest of the book, but also can be kept as is (original)		Please see PDF	
8	Images showcasing image-to-image with denoi	strengths.png	hogaldm_0801.png	As-is			Please see PDF	
8	N/A	cell-6-output-1	hogaldm_08or01.png	As-is				
8	N/A	cell-9-output-4	hogaldm_08or02.png	As-is				
8	N/A	cell-11-output-2	hogaldm_08or03.png	As-is				
8	N/A	cell-12-output-2	hogaldm_08or04.png	As-is				
8	N/A	cell-13-output-2	hogaldm_08or05.png	As-is				
8	N/A	cell-14-output-3	hogaldm_08or06.png	As-is				
8	Examples of editing with the DDIM Inversion	inversion_examples.png	hogaldm_0802.png	As-is / Redraw	Can be redrawn for harmonization with the rest of the book, but also can be kept as is (we have authorization from the authors)		Please see PDF	
8	N/A	textix_edited_glasses.png	hogaldm_0807.png	As-is	Note from Omar: we needed to export this			
8	Examples of edits with the InstructPix2Pix	p2p1_pixel_instruct.png	hogaldm_0803.png	As-is / Redraw	Can be redrawn for harmonization with the rest of the book, but also can be kept as is (original) - redrawing could be preferred to make it clearer		Please see PDF	
8	N/A	coral_edited_mountain.png	hogaldm_0808.png	As-is	Note from Omar: we needed to export this			
8	ControlNet examples with image input, canny	controlnet_examples.png	hogaldm_0804.png	As-is / Redraw	Can be redrawn for harmonization with the rest of the book, but also can be kept as is (original) - redrawing could be preferred to make it clearer		Please see PDF	Please see PDF
8	Duck ControlNet Canny example	controlnet_canny.png	hogaldm_0805.png	As-is				
8	N/A	controlnet_masked.png	hogaldm_0809.png	As-is				
8	A diagram of the IP-Adapter architecture.	ip_adapter_features.png	hogaldm_0806.png	As-is / Redraw	Can be redrawn for harmonization with the rest of the book, but also can be kept as is (as the license of the image is Apache 2.0)			
8	N/A	camera_flower_jp_variation.png	hogaldm_0810.png	As-is				
8	N/A	cell-21-output-2	hogaldm_0811.png	As-is				
8	N/A	cell-22-output-3	hogaldm_08or12.png	As-is				
9	Some examples of audio tasks	Description / guidance to be provided	hogaldm_0901.png	Redraw	Similar idea to 2-15. I would suggest three three 1. Automatic Speech Recognition: a waveform at the left, an arrow, text 2. Text to Speech: same as above, but the other way around 3. Voice enhancement: waveform in both sides		Please see PDF	
9	A waveform sampled using a sampling rate	no draft, see Col F for description	hogaldm_0902.png	Redraw	An image of a waveform. The x axis is time and goes from 0 to 1 second. The sampling rate is 6, so there is a snapshot (a red mark) at 0, 0.2, 0.4, 0.6, 0.8 and 1			
9	The amplitude of some sounds in the dB scale	no draft, see Col F for description	hogaldm_0903.png	Redraw	A plot bar. At the left, 0dB. At 10, a person breathing. At 120, an intense concert. At 210, a volcano eruption.		Please see PDF	
9	N/A	9-4.pdf	hogaldm_09or01.png	As-is				
9	Spectrogram (left) and mel spectrogram (right)	cell-13-output-1	hogaldm_0904.png	As-is		higher-res.pdf to be provided	Please see PDF	
9	The same note played by several instruments	no draft, see Col F for description	hogaldm_0905.png	Redraw	An image that explains amplitude and period. An idea here is to show an image like the one from this post: http://newtdesign.blogspot.			
9	Some sinusoidal functions and their frequency	9-7.pdf	hogaldm_09or02.png	As-is	Note about caption: We want pitch to be italized			
9	A complex sound wave can be decomposed	9-8.pdf	hogaldm_09or05.png	As-is/redraw	Replace/remove text below image			
9	N/A	9-10.pdf	hogaldm_09or07.png	As-is				
9	N/A	9-11.pdf	hogaldm_09or03.png	As-is				
9	N/A	9-12.pdf	hogaldm_09or04.png	As-is				
9	Whisper training is modeled in a sequence	no draft, see Col F for description	hogaldm_0906.png	Redraw	Image from the Whisper paper about the sequence-to-sequence audio task. We are thinking of a simplified version of figure 1 in https://arxiv.org/pdf/2212. 04356		Please see PDF	Please see PDF
9	Architecture of the SpeechT5 model. Image	no draft, see Col F for description	hogaldm_0909.png	Redraw	An image of the full SpeechT5 model, inspired on https://huggingface.co/blog/speecht5		Please see PDF	
9	N/A	9-15.pdf	hogaldm_09or06.png	As-is				
9	The Bark pipeline uses three components	Template / guidance to be provided	hogaldm_0910.png	Redraw	Text at the left. Things then go through the following components 1. The text goes into "Semantic Model", which outputs semantic text tokens 2. The output of 1 goes into "Coarse Model" which outputs two blocks (called codebooks) 3. A "Fine Acoustic Model" uses 2 and itself (arrow to itself) t predict the next 6 codebooks 4. All codebooks from 2 and 3 go into "Encoder Model" which outputs the audio		Please see PDF	
9	The AudioLM model pipeline. AudioLM con	no draft, see Col F for description	hogaldm_0911.png	Redraw	Based on https://arxiv.org/pdf/2209.03143. At the bottom, we have a waveform. There is SoundStream -> Acoustic Tokens, and w2v-BERT -> Semantic Tokens coming out of it.		Please see PDF	Please see PDF
9	Music M incorporates text conditioning to the	no draft, see Col F for description	hogaldm_0912.png	Redraw	Image based on Figure 2 from https://arxiv.org/pdf/2301.11325			
9	N/A	9-19.pdf	hogaldm_09or07.png	As-is				
9	N/A	cell-58-output-1	hogaldm_09or06.png	As-is				
10	RLHF three core components: pre-training	no draft, see Col F for description	hogaldm_1001.png	Redraw	A diagram with 3 components: human feedback, pre-training, and fine-tuning. A simplified version of https://huggingface.		Please see PDF	
10	RLHF applied to diffusion models.	no draft, see Col F for description	hogaldm_1002.png	Redraw	The before and after results from https://huggingface.co/blog/dreambooth-lora		Please see PDF	
10	Needle In A Haystack is a method to evaluate	no draft, see Col F for description	hogaldm_1003.png	Redraw	Image of Needle in A Haystack eval, similar to https://github.		Please see PDF	
10	A simplified traditional transformer block	final_chapter/blok.png	hogaldm_1004.png	Redraw	To make nicer			
10	Example of Object Detection	more detail needed here	hogaldm_1005.png	Redraw	To re-use parts of 2-15			See Jira ticket or PDF for AU instruc for redraw
10	Example of Depth Estimation	more detail needed here	hogaldm_1006.png	Redraw	To re-use parts of 2-15			See Jira ticket or PDF for AU instruc for redraw
10	A mesh can be generated even from a sing	Template / guidance to be provided	hogaldm_1007.png	Redraw	We can use an example of image to 3D. See https://huggingface.			See Jira ticket or PDF for AU instruc for redraw
10	Example of video generation	Template / guidance to be provided	hogaldm_1008.png	Redraw	From Omar: any ideas here are appreciated. One option could be to have an input some text and then have multiple images as frames of a video, but I'm not sure			See Jira ticket or PDF for AU instruc for redraw
10	Example of BLIP captioning	Template / guidance to be provided	hogaldm_1009.png	Redraw	The following image: https://help-blol.nl... With a text below that says "caption: a pizza with pepperoni and green leaves on top"		Please see PDF	
10	VLM interface.	To add a screenshot of some nice space	hogaldm_1010.png	As-is	A screenshot from https://huggingface.co/spaces/llava/llava-vl/llava or https://nvlm.store.co (happy to provide more options)		Please see PDF	
App	A RAG pipeline	no draft, see Col F for description	hogaldm_1301.png	Redraw	A setup in which a query goes to a retriever, which then gets info from a memory, and it's taken to the model. Similar/inspired on https://huggingface.co			

Original names	New names
cell-6-output-2.png	hogaitdm_01in01.png
placeholder.png	hogaitdm_0201.png
placeholder.png	hogaitdm_0202.png
placeholder.png	hogaitdm_0203.png
placeholder.png	hogaitdm_0204.png
placeholder.png	hogaitdm_0205.png
placeholder.png	hogaitdm_0206.png
placeholder.png	hogaitdm_0207.png
placeholder.png	hogaitdm_0208.png
transformer_arch.png	hogaitdm_0209.png
placeholder.png	hogaitdm_0210.png
placeholder.png	hogaitdm_0211.png
placeholder.png	hogaitdm_0212.png
placeholder.png	hogaitdm_0213.png
image-classification-input.jpeg	hogaitdm_02in01.png
placeholder.png	hogaitdm_0214.png
placeholder.png	hogaitdm_0301.png
Autoencoder.png	hogaitdm_0302.png
cell-6-output-1.png	hogaitdm_03in01.png
cell-7-output-1.png	hogaitdm_03in02.png
cell-9-output-1.png	hogaitdm_03in03.png
cell-14-output-1.png	hogaitdm_03in04.png
cell-29-output-1.png	hogaitdm_03in05.png
cell-32-output-1.png	hogaitdm_03in06.png
cell-33-output-1.png	hogaitdm_0303.png
cell-40-output-1.png	hogaitdm_03in07.png
cell-43-output-1.png	hogaitdm_03in08.png
cell-46-output-1.png	hogaitdm_03in09.png
cell-48-output-1.png	hogaitdm_03in10.png
cell-49-output-1.png	hogaitdm_03in11.png
vae-diagram.png	hogaitdm_0304.png
cell-56-output-1.png	hogaitdm_03in12.png
cell-57-output-1.png	hogaitdm_03in13.png
cell-58-output-1.png	hogaitdm_03in14.png
cell-59-output-1.png	hogaitdm_03in15.png
cell-62-output-1.png	hogaitdm_03in16.png
cell-64-output-1.png	hogaitdm_03in17.png
cell-66-output-1.png	hogaitdm_03in18.png
cell-66-output-2.png	hogaitdm_03in19.png
cell-67-output-1.png	hogaitdm_03in20.png
cell-68-output-1.png	hogaitdm_03in21.png
cell-70-output-1.png	hogaitdm_03in22.png

placeholder.png	hogaitdm_0304.png
lion.jpg	hogaitdm_0305.png
cell-74-output-1.png	hogaitdm_03in24.png
iterative_denoising.png	hogaitdm_0401.png
cell-3-output-3.png	hogaitdm_04in01.png
cell-4-output-1.png	hogaitdm_04in02.png
cell-4-output-2.png	hogaitdm_04in03.png
cell-4-output-3.png	hogaitdm_04in04.png
cell-4-output-4.png	hogaitdm_04in05.png
placeholder.png	hogaitdm_0402.png
cell-9-output-1.png	hogaitdm_04in06.png
cell-10-output-1.png	hogaitdm_04in07.png
unet.png	hogaitdm_0403.png
cell-14-output-1.png	hogaitdm_04in08.png
cell-16-output-2.png	hogaitdm_04in09.png
cell-17-output-1.png	hogaitdm_04in10.png
fim.png	hogaitdm_0404.png
placeholder.png	hogaitdm_0405.png
cell-20-output-1.png	hogaitdm_04in11.png
cell-21-output-1.png	hogaitdm_04in12.png
cell-22-output-1.png	hogaitdm_04in13.png
cell-23-output-1.png	hogaitdm_04in14.png
cell-24-output-1.png	hogaitdm_04in15.png
cell-25-output-1.png	hogaitdm_04in16.png
cell-26-output-1.png	hogaitdm_04in17.png
cell-27-output-1.png	hogaitdm_04in18.png
unet.png	hogaitdm_0406.png
basic_unet_generations.png	hogaitdm_0407.png
advanced_unet_detail.png	hogaitdm_0408.png
placeholder.png	hogaitdm_0409.png
cell-29-output-1.png	hogaitdm_04in19.png
cell-3-output-1.png	hogaitdm_05in01.png
cell-7-output-1.png	hogaitdm_05in02.png
cell-9-output-1.png	hogaitdm_05in03.png
cell-12-output-2.png	hogaitdm_05in04.png
cell-13-output-2.png	hogaitdm_05in05.png
cell-14-output-2.png	hogaitdm_05in06.png
placeholder.png	hogaitdm_0501.png
cell-16-output-2.png	hogaitdm_05in07.png
simplified_unet.png	hogaitdm_0502.png
text_encoder.png	hogaitdm_0503.png
vae.png	hogaitdm_0504.png
cell-19-output-1.png	hogaitdm_05in08.png

cell-21-output-1.png	hogaitdm_05in09.png
cell-22-output-1.png	hogaitdm_05in10.png
sd_unet.png	hogaitdm_0505.png
placeholder.png	hogaitdm_0506.png
cell-24-output-1.png	hogaitdm_05in11.png
cell-29-output-1.png	hogaitdm_05in12.png
cell-30-output-1.png	hogaitdm_0507.png
gradio.png	hogaitdm_05in13.png
placeholder.png	hogaitdm_0601.png
placeholder.png	hogaitdm_0602.png
cell-23-output-1.png	hogaitdm_06in01.png
placeholder.png	hogaitdm_0603.png
placeholder.png	hogaitdm_0604.png
lora-diagram.png	hogaitdm_0605.png
placeholder.png	hogaitdm_0606.png
placeholder.png	hogaitdm_0607.png
lmsys.png	hogaitdm_0608.png
placeholder.png	hogaitdm_0609.png
sd_fine_tuning.png	hogaitdm_0701.png
cell-6-output-2.png	hogaitdm_07in01.png
dreambooth_flow.png	hogaitdm_0702.png
poli_grid.png	hogaitdm_0703.png
poli_dreambooth.png	hogaitdm_0704.png
lora-diagram.png	hogaitdm_0705.png
cell-10-output-2.png	hogaitdm_07in02.png
depth_model_example.png	hogaitdm_0706.png
multiple_strenghts.png	hogaitdm_0801.png
cell-6-output-1.png	hogaitdm_08in01.png
cell-9-output-4.png	hogaitdm_08in02.png
cell-11-output-2.png	hogaitdm_08in03.png
cell-12-output-2.png	hogaitdm_08in04.png
cell-13-output-2.png	hogaitdm_08in05.png
cell-14-output-3.png	hogaitdm_08in06.png
ddim_inversion_examples.png	hogaitdm_0802.png
ledits_edited_glasses.png	hogaitdm_08in07.png
girl_pearl_instruct.png	hogaitdm_0803.png
cosxl_edited_mountain.png	hogaitdm_08in08.png
controlnet_examples.png	hogaitdm_0804.png
controlnet_canny.png	hogaitdm_0805.png
controlnet_masked.png	hogaitdm_08in09.png
ip_adapter_features.png	hogaitdm_0806.png
camera_flower_ip_variation.png	hogaitdm_08in10.png
cell-21-output-3.png	hogaitdm_08in11.png

cell-22-output-3.png	hogaitdm_08in12.png
placeholder.png	hogaitdm_0901.png
placeholder.png	hogaitdm_0902.png
placeholder.png	hogaitdm_0903.png
cell-12-output-1.png	hogaitdm_09in01.png
cell-13-output-1.png	hogaitdm_0904.png
placeholder.png	hogaitdm_0905.png
cell-14-output-1.png	hogaitdm_09in02.png
cell-15-output-1.png	hogaitdm_0906.png
cell-16-output-1.png	hogaitdm_0907.png
cell-17-output-1.png	hogaitdm_09in03.png
cell-18-output-1.png	hogaitdm_09in04.png
cell-19-output-1.png	hogaitdm_09in05.png
placeholder.png	hogaitdm_0908.png
placeholder.png	hogaitdm_0909.png
cell-39-output-1.png	hogaitdm_09in06.png
placeholder.png	hogaitdm_0910.png
placeholder.png	hogaitdm_0911.png
placeholder.png	hogaitdm_0912.png
cell-55-output-1.png	hogaitdm_09in07.png
cell-58-output-1.png	hogaitdm_09in08.png
hugging_face.png	hugging_face.png
hugging_face.png	hugging_face.png
placeholder.png	hogaitdm_1001.png
placeholder.png	hogaitdm_1002.png
placeholder.png	hogaitdm_1003.png
MoE.png	hogaitdm_1004.png
placeholder.png	hogaitdm_1005.png
placeholder.png	hogaitdm_1006.png
placeholder.png	hogaitdm_1007.png
placeholder.png	hogaitdm_1008.png
placeholder.png	hogaitdm_1009.png
placeholder.png	hogaitdm_1010.png
placeholder.png	hogaitdm_1301.png

Chapter	Page	Issue	Intended					
Global		Quote marks seem backwards most of the time						
Global		There are some log parts we'll want to remove (e.g. see page 17)						
Global		Q: should we add break pages between the 3 big sections of the book?						
Global		Some images are automatically getting "Figure ..." automatically by exporting in Atlas, but the images generated by code don't get the same treatment. Some help here to harmonize would be appreciated, we're not sure how to best tackle this						
2	36	Multiplication sign is removed from caption						
3	99	~~The code anchors are not rendering properly~~	Fixed in the latest PDF					
3	101	Should equations be centered?						
3	115	~~LaTeX in the callout is not rendered properly (it worked in the latest early p~~	Fixed in the latest PDF					
4		A text before a footnote couldn't be italized	"Cold Diffusion" should be italized (fixed in latest pdf)					
4		A text before a footnote couldn't be italized	"Common Diffusion Noise Schedules and Sample Steps are Flawed" should be italized					
5		Multiplications in parenthesis (786,432) should use proper multiplication sign						

5	160	Duplicate line `with torch.inference_mode()` in last code block	code should read:
			
			`x = torch.randn((1, 1, 32, 32))`
			`with torch.inference_mode():`
			`    out = model(x, timestep=7, class_labels=torch.tensor([2])).sample`
			`out.shape`					

5	174	Missing comma in first code block	code should read:
			
			`from genaibook.core import load_image, show_image, SampleURL`
			
			`im = load_image(`
			`    SampleURL.LlamaExample,`
			`    size=(512, 512),`
			`)`
			`show_image(im);`					

5	181	Typo in first code block	line should read `image_grid(images, 1, 4)` instead of `image_grid(images, 1, 5)`					
6		We'll want some help to make table 6-2 nicer. It looks ok in jupyter but not in the export						
6		We'll want some help to make 6-3 look nice						
6	204	Typo in code block (missing .core)	Third line should read `from genaibook.core import get_device`, not `from genaibook import get_device`					
8		A triple "+" sign is being deleted	"The is equ" should be "The +++ is equivalent". Same in "Besides using the +++"					
8		Paper could not be italized in footnote	"An Image is Worth One Word..." should be italized					
8		A text before a footnote couldn't have a ++ sign before	"LEDITS" should be "LEDITS++"					
13		Some weird formatting "cell..."						