Preface

> What I cannot create, I do not understand.
>
> —Richard Feynman

Generative AI is one of the most revolutionary technologies of our time, transforming the way we interact with machines. Its potential to revolutionize the way we live, work, and play has been the subject of countless conversations, debates, and predictions. But what if there was an even greater potential to this powerful technology? What if the possibilities of generative AI extend beyond our current imagination? The future of generative AI may be more exciting than we ever thought possible…

Since our earliest days, we have sought opportunities to generate original and beautiful creations. For early humans, this took the form of cave paintings depicting wild animals and abstract patterns, created with pigments placed carefully and methodically onto rock. The Romantic Era gave us the mastery of Tchaikovsky symphonies, with their ability to inspire feelings of triumph and tragedy through sound waves, woven together to form beautiful melodies and harmonies. And in recent times, we have found ourselves rushing to bookshops at midnight to buy stories about a fictional wizard, because the combination of letters creates a narrative that wills us to turn the page and find out what happens to our hero.

It is therefore not surprising that humanity has started to ask the ultimate question of creativity: can we create something that is in itself creative?

This is the question that generative AI aims to answer. With recent advances in methodology and technology, we are now able to build machines that can paint original artwork in a given style, write coherent blocks of text with long-term structure, compose music that is pleasant to listen to, and develop winning strategies for complex games by generating imaginary future scenarios. This is just the start of a generative revolution that will leave us with no choice but to find answers to some of the biggest questions about the mechanics of creativity, and ultimately, what it means to be human.

In short, there has never been a better time to learn about generative AI—so let's get started!

Objective and Approach

This book assumes no prior knowledge of generative AI. We will build up all of the key concepts from scratch in a way that is intuitive and easy to follow, so don't worry if you have no experience with generative AI. You have come to the right place!

Rather than only covering the techniques that are currently in vogue, this book serves as a complete guide to generative modeling that covers a broad range of model families. There is no one technique that is objectively *better* or *worse* than any other—in fact, many state-of-the-art models now mix together ideas from across the broad spectrum of approaches to generative modeling. For this reason, it is important to keep abreast of developments across all areas of generative AI, rather than focusing on one particular kind of technique. One thing is certain: the field of generative AI is moving fast, and you never know where the next groundbreaking idea will come from!

With this in mind, the approach I will take is to show you how to train your own generative models on your own data, rather than relying on pre-trained off-the-shelf models. While there are now many impressive open source generative models that can be downloaded and run in a few lines of code, the aim of this book is to dig deeper into their architecture and design from first principles, so that you gain a complete understanding of how they work and can code up examples of each technique from scratch using Python and Keras.

In summary, this book can be thought of as a map of the current generative AI landscape that covers both theory and practical applications, including full working examples of key models from the literature. We will walk through the code for each step by step, with clear signposts that show how the code implements the theory underpinning each technique. This book can be read cover to cover or used as a reference book that you can dip into. Above all, I hope you find it a useful and enjoyable read!

Throughout the book, you will find short, allegorical stories that help explain the mechanics of some of the models we will be building. I believe that one of the best ways to teach a new abstract theory is to first convert it into something that isn't quite so abstract, such as a story, before diving into the technical explanation. The story and the model explanation are just the same mechanics explained in two different domains—you might therefore find it useful to refer back to the relevant story while learning about the technical details of each model!

Prerequisites

This book assumes that you have experience coding in Python. If you are not familiar with Python, the best place to start is through LearnPython.org. There are many free resources online that will allow you to develop enough Python knowledge to work with the examples in this book.

Also, since some of the models are described using mathematical notation, it will be useful to have a solid understanding of linear algebra (for example, matrix multiplication) and general probability theory. A useful resource is Deisenroth et al.'s book *Mathematics for Machine Learning* (Cambridge University Press), which is freely available.

The book assumes no prior knowledge of generative modeling (we will examine the key concepts in Chapter 1) or TensorFlow and Keras (these libraries will be introduced in Chapter 2).

Roadmap

This book is divided into three parts.

Part I is a general introduction to generative modeling and deep learning, where we explore the core concepts that underpin all of the techniques in later parts of the book:

- In Chapter 1, "Generative Modeling", we define generative modeling and consider a toy example that we can use to understand some of the key concepts that are important to all generative models. We also lay out the taxonomy of generative model families that we will explore in Part II of this book.
- In Chapter 2, "Deep Learning", we begin our exploration of deep learning and neural networks by building our first example of a multilayer perceptron (MLP) using Keras. We then adapt this to include convolutional layers and other improvements, to observe the difference in performance.

Part II walks through the six key techniques that we will be using to build generative models, with practical examples for each:

- In Chapter 3, "Variational Autoencoders", we consider the variational autoencoder (VAE) and see how it can be used to generate images of faces and morph between faces in the model's latent space.
- In Chapter 4, "Generative Adversarial Networks", we explore generative adversarial networks (GANs) for image generation, including deep convolutional GANs, conditional GANs, and improvements such as the Wasserstein GAN that make the training process more stable.

- In Chapter 5, "Autoregressive Models", we turn our attention to autoregressive models, starting with an introduction to recurrent neural networks such as long short-term memory networks (LSTMs) for text generation and PixelCNN for image generation.
- In Chapter 6, "Normalizing Flow Models", we focus on normalizing flows, including an intuitive theoretical exploration of the technique and a practical example of how to build a RealNVP model to generate images.
- In Chapter 7, "Energy-Based Models", we cover energy-based models, including important methods such as how to train using contrastive divergence and sample using Langevin dynamics.
- In Chapter 8, "Diffusion Models", we dive into a practical guide to building diffusion models, which drive many state-of-the-art image generation models such as DALL.E 2 and Stable Diffusion.

Finally, in Part III we build on these foundations to explore the inner workings of state-of-the-art models for image generation, writing, composing music, and model-based reinforcement learning:

- In Chapter 9, "Transformers", we explore the lineage and technical details of the StyleGAN models, as well as other state-of-the-art GANs for image generation such as VQ-GAN.
- In Chapter 10, "Advanced GANs", we consider the Transformer architecture, including a practical walkthrough for building your own version of GPT for text generation.
- In Chapter 11, "Music Generation", we turn our attention to music generation, including a guide to working with music data and application of techniques such as Transformers and MuseGAN.
- In Chapter 12, "World Models", we see how generative models can be used in the context of reinforcement learning, with the application of world models and Transformer-based methods.
- In Chapter 13, "Multimodal Models", we explain the inner workings of four state-of-the-art multimodal models that incorporate more than one type of data, including DALL.E 2, Imagen, and Stable Diffusion for text-to-image generation and Flamingo, a visual language model.
- In Chapter 14, "Conclusion", we recap the key milestones of generative AI to date and discuss the ways in which generative AI will revolutionize our daily lives in years to come.

Changes in the Second Edition

Thank you to everyone who read the first edition of this book—I am really pleased that so many of you have found it a useful resource and provided feedback on things that you would like to see in the second edition. The field of generative deep learning has progressed significantly since the first edition was published in 2019, so as well as refreshing the existing content I have added several new chapters to bring the material in line with the current state of the art.

The following is a summary of the main updates, in terms of the individual chapters and general book improvements:

- Chapter 1 now includes a section on the different families of generative models and a taxonomy of how they are related.
- Chapter 2 contains improved diagrams and more detailed explanations of key concepts.
- Chapter 3 is refreshed with a new worked example and accompanying explanations.
- Chapter 4 now includes an explanation of conditional GAN architectures.
- Chapter 5 now includes a section on autoregressive models for images (e.g., PixelCNN).
- Chapter 6 is an entirely new chapter, describing the RealNVP model.
- Chapter 7 is also a new chapter, focusing on techniques such as Langevin dynamics and contrastive divergence.
- Chapter 8 is a newly written chapter on denoising the diffusion models that power many of today's state-of-the-art applications.
- Chapter 9 is an expansion of the material provided in the conclusion of the first edition, with deeper focus on architectures of the various StyleGAN models and new material on VQ-GAN.
- Chapter 10 is a new chapter that explores the Transformer architecture in detail.
- Chapter 11 includes modern Transformer architectures, replacing the LSTM models from the first edition.
- Chapter 12 includes updated diagrams and descriptions, with a section on how this approach is informing state-of-the-art reinforcement learning today.
- Chapter 13 is a new chapter that explains in detail how impressive models like DALL.E 2, Imagen, Stable Diffusion, and Flamingo work.
- Chapter 14 is updated to reflect the outstanding progress in the field since the first edition and give a more complete and detailed view of where generative AI is heading in the future.

- All comments given as feedback to the first edition and typos identified have been addressed (to the best of my knowledge!).
- Chapter goals have been added at the start of each chapter, so that you can see the key topics covered in the chapter before you start reading.
- Some of the allegorical stories have been rewritten to be more concise and clear—I am pleased that so many readers have said that the stories have helped them to better understand the key concepts!
- The headings and subheadings of each chapter have been aligned so that is it clear which parts of the chapter are focused on explanation and which are focused on building your own models.

Other Resources

I highly recommend the following books as general introductions to machine learning and deep learning:

- *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems* by Aurélien Géron (O'Reilly)
- *Deep Learning with Python* by Francois Chollet (Manning)

Most of the papers in this book are sourced through arXiv, a free repository of scientific research papers. It is now common for authors to post papers to arXiv before they are fully peer-reviewed. Reviewing the recent submissions is a great way to keep on top of the most cutting-edge developments in the field.

I also highly recommend the website Papers with Code, where you can find the latest state-of-the-art results in a variety of machine learning tasks, alongside links to the papers and official GitHub repositories. It is an excellent resource for anyone wanting to quickly understand which techniques are currently achieving the highest scores in a range of tasks and has certainly helped me to decide which techniques to include in this book.

Conventions Used in This Book

The following typographical conventions are used in this book:

Italic
: Indicates new terms, URLs, email addresses, filenames, and file extensions.

`Constant width`
: Used for commands and program listings, as well as within paragraphs to refer to program elements such as variable or function names.

`Constant width italic`
: Shows text that should be replaced with user-supplied values or by values determined by context.

This element signifies a tip or suggestion.

This element signifies a general note.

This element signifies a warning or caution.

Codebase

The code examples in this book can be found in a GitHub repository. I have deliberately ensured that none of the models require prohibitively large amounts of computational resources to train, so that you can start training your own models without having to spend lots of time or money on expensive hardware. There is a comprehensive guide in the repository on how to get started with Docker and set up cloud resources with GPUs on Google Cloud if required.

The following changes have been made to the codebase since the first edition:

- All examples are now runnable from within a single notebook, instead of some code being imported from modules across the codebase. This is so that you can run each example cell by cell and delve into exactly how each model is built, piece by piece.
- The sections of each notebook are now broadly aligned between examples.
- Many of the examples in this book now utilize code snippets from the amazing open source Keras repository—this is to avoid creating a completely detached open source repository of Keras generative AI examples, when there already exist excellent implementations available through the Keras website. I have added references and links to the original authors of code that I have utilized from the Keras website throughout this book and in the repository.

- I have added new data sources and improved the data collection process from the first edition—now, there is a script that can be easily run to collect data from the required sources in order to train the examples in the book, using tools such as the Kaggle API.

Using Code Examples

Supplemental material (code examples, exercises, etc.) is available for download at *https://github.com/davidADSP/Generative_Deep_Learning_2nd_Edition*.

If you have a technical question or a problem using the code examples, please send email to *bookquestions@oreilly.com*.

This book is here to help you get your job done. In general, if example code is offered with this book, you may use it in your programs and documentation. You do not need to contact us for permission unless you're reproducing a significant portion of the code. For example, writing a program that uses several chunks of code from this book does not require permission. Selling or distributing examples from O'Reilly books does require permission. Answering a question by citing this book and quoting example code does not require permission. Incorporating a significant amount of example code from this book into your product's documentation does require permission.

We appreciate, but do not require, attribution. An attribution usually includes the title, author, publisher, and ISBN. For example: "*Generative Deep Learning*, 2nd edition, by David Foster (O'Reilly). Copyright 2023 Applied Data Science Partners Ltd., 978-1-098-13418-1."

If you feel your use of code examples falls outside fair use or the permission given above, feel free to contact us at *permissions@oreilly.com*.

O'Reilly Online Learning

For more than 40 years, *O'Reilly Media* has provided technology and business training, knowledge, and insight to help companies succeed.

Our unique network of experts and innovators share their knowledge and expertise through books, articles, and our online learning platform. O'Reilly's online learning platform gives you on-demand access to live training courses, in-depth learning paths, interactive coding environments, and a vast collection of text and video from O'Reilly and 200+ other publishers. For more information, visit *https://oreilly.com*.

How to Contact Us

Please address comments and questions concerning this book to the publisher:

O'Reilly Media, Inc.
1005 Gravenstein Highway North
Sebastopol, CA 95472
800-998-9938 (in the United States or Canada)
707-829-0515 (international or local)
707-829-0104 (fax)

We have a web page for this book, where we list errata, examples, and any additional information. You can access this page at *https://oreil.ly/generative-dl*.

Email *bookquestions@oreilly.com* to comment or ask technical questions about this book.

For news and information about our books and courses, visit *https://oreilly.com*.

Find us on LinkedIn: *https://linkedin.com/company/oreilly-media*

Follow us on Twitter: *https://twitter.com/oreillymedia*

Watch us on YouTube: *https://youtube.com/oreillymedia*

Acknowledgments

There are so many people I would like to thank for helping me write this book.

First, I would like to thank everyone who has taken time to technically review the book—in particular Vishwesh Ravi Shrimali, Lipi Deepaakshi Patnaik, Luba Elliot, and Lorna Barclay. Thanks also to Samir Bico for helping to review and test the codebase that accompanies this book. Your input has been invaluable.

Also, a huge thanks to my colleagues at Applied Data Science Partners, Ross Witeszczak, Amy Bull, Ali Parandeh, Zine Eddine, Joe Rowe, Gerta Salillari, Aleshia Parkes, Evelina Kireilyte, Riccardo Tolli, Mai Do, Khaleel Syed, and Will Holmes. Your patience with me while I have taken time to finish the book is hugely appreciated, and I am greatly looking forward to all the machine learning projects we will complete together in the future! Particular thanks to Ross—had we not decided to start a business together, this book might never have taken shape, so thank you for believing in me as your business partner!

I also want to thank anyone who has ever taught me anything mathematical—I was extremely fortunate to have fantastic math teachers at school, who developed my interest in the subject and encouraged me to pursue it further at university. I would

like to thank you for your commitment and for going out of your way to share your knowledge of the subject with me.

A huge thank you goes to the staff at O'Reilly for guiding me through the process of writing this book. A special thanks goes to Michele Cronin, who has been there at each step, providing useful feedback and sending me friendly reminders to keep completing chapters! Also to Nicole Butterfield, Christopher Faucher, Charles Roumeliotis, and Suzanne Huston for getting the book into production, and Mike Loukides for first reaching out to ask if I'd be interested in writing a book. You have all been so supportive of this project from the start, and I want to thank you for providing me with a platform on which to write about something that I love.

Throughout the writing process, my family has been a constant source of encouragement and support. A huge thank you goes to my mum, Gillian Foster, for checking every single line of text for typos and for teaching me how to add up in the first place! Your attention to detail has been extremely helpful while proofreading this book, and I'm really grateful for all the opportunities that both you and dad have given me. My dad, Clive Foster, originally taught me how to program a computer—this book is full of practical examples, and that's thanks to his early patience while I fumbled around in BASIC trying to make football games as a teenager. My brother, Rob Foster, is the most modest genius you will ever find, particularly within linguistics—chatting with him about AI and the future of text-based machine learning has been amazingly helpful. Last, I would like to thank my Nana, who was always a constant source of inspiration and fun for all of us. Her love of literature was one of the reasons I first decided that writing a book would be an exciting thing to do.

I would also like to thank my wife, Lorna Barclay. As well as providing me with endless support and cups of tea throughout the writing process, you have rigorously checked every word of this book in meticulous detail. I couldn't have done it without you. Thank you for always being there for me, and for making this journey so much more enjoyable. I promise I won't talk about generative AI at the dinner table for at least a few days after the book is published.

Lastly, I would like to thank our beautiful baby daughter Alina for providing endless entertainment during the long nights of book-writing. Your adorable giggles have been the perfect background music to my typing. Thanks for being my inspiration and for always keeping me on my toes. You're the real brains behind this operation.