

Introduction to Transformers for NLP

**With the Hugging Face Library
and Models to Solve Problems**

Shashank Mohan Jain

Apress®

Introduction to Transformers for NLP: With the Hugging Face Library and Models to Solve Problems

Shashank Mohan Jain
Bangalore, India

ISBN-13 (pbk): 978-1-4842-8843-6
<https://doi.org/10.1007/978-1-4842-8844-3>

ISBN-13 (electronic): 978-1-4842-8844-3

Copyright © 2022 by Shashank Mohan Jain

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

Trademarked names, logos, and images may appear in this book. Rather than use a trademark symbol with every occurrence of a trademarked name, logo, or image we use the names, logos, and images only in an editorial fashion and to the benefit of the trademark owner, with no intention of infringement of the trademark.

The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Managing Director, Apress Media LLC: Welmoed Spahr
Acquisitions Editor: Celestin Suresh John
Development Editor: James Markham
Coordinating Editor: Shrikant Vishwakarma

Cover designed by eStudioCalamar

Cover image by and machines on Unsplash (www.unsplash.com)

Distributed to the book trade worldwide by Apress Media, LLC, 1 New York Plaza, New York, NY 10004, U.S.A. Phone 1-800-SPRINGER, fax (201) 348-4505, e-mail orders-ny@springer-sbm.com, or visit www.springeronline.com. Apress Media, LLC is a California LLC and the sole member (owner) is Springer Science + Business Media Finance Inc (SSBM Finance Inc). SSBM Finance Inc is a **Delaware** corporation.

For information on translations, please e-mail booktranslations@springernature.com; for reprint, paperback, or audio rights, please e-mail bookpermissions@springernature.com.

Apress titles may be purchased in bulk for academic, corporate, or promotional use. eBook versions and licenses are also available for most titles. For more information, reference our Print and eBook Bulk Sales web page at <http://www.apress.com/bulk-sales>.

Any source code or other supplementary material referenced by the author in this book is available to readers on GitHub (<https://github.com/Apress>). For more detailed information, please visit <http://www.apress.com/source-code>.

Printed on acid-free paper

Table of Contents

About the Authorvii

About the Technical Reviewerix

Introductionxi

Chapter 1: Introduction to Language Models1

 History of NLP2

 Bag of Words4

 n-grams6

 Recurrent Neural Networks8

 Language Models11

 Summary.....16

Chapter 2: Introduction to Transformers19

 What Is a Seq2Seq Neural Network?20

 The Transformer21

 Transformers22

 Summary.....36

Chapter 3: BERT37

 Workings of BERT38

 Masked LM (MLM)38

 Next Sentence Prediction41

 Inference in NSP43

TABLE OF CONTENTS

BERT Pretrained Models44

BERT Input Representations.....45

Use Cases for BERT46

 Sentiment Analysis on Tweets47

Performance of BERT on a Variety of Common Language Tasks48

Summary.....49

Chapter 4: Hugging Face51

 Features of the Hugging Face Platform53

 Components of Hugging Face54

 Summary.....67

Chapter 5: Tasks Using the Hugging Face Library69

 Gradio: An Introduction.....69

 Creating a Space on Hugging Face70

 Hugging Face Tasks72

 Question and Answering.....72

 Translation78

 Summary84

 Zero-Shot Learning.....90

 Text Generation Task/Models.....95

 Text-to-Text Generation106

 Chatbot/Dialog Bot.....123

 Code and Code Comment Generation.....126

 Code Comment Generator131

 Summary.....136

Chapter 6: Fine-Tuning Pretrained Models	137
Datasets	139
Fine-Tuning a Pretrained Model	142
Training for Fine-Tuning	142
Inference	150
Summary	151
Appendix A: Vision Transformers	153
Self-Attention and Vision Transformers	153
Summary	157
Index	159

About the Author



Shashank Mohan Jain has been working in the IT industry for around 22 years mainly in the areas of cloud computing, machine learning, and distributed systems. He has keen interests in virtualization techniques, security, and complex systems. Shashank has many software patents to his name in the area of cloud computing, IoT, and machine learning. He is a speaker at multiple reputed cloud conferences. Shashank holds Sun, Microsoft, and Linux kernel certifications.

About the Technical Reviewer



Akshay Kulkarni is a renowned AI and machine learning evangelist and thought leader. He has consulted several Fortune 500 and global enterprises on driving AI and data science-led strategic transformation. Akshay has rich experience in building and scaling AI and machine learning businesses and creating significant impact. He is currently a data science and AI manager at Publicis Sapient, where he is a part of strategy and transformation interventions through AI. He manages high-priority growth initiatives around data science and works on various artificial intelligence engagements by applying state-of-the-art techniques to this space. Akshay is also a Google Developers Expert in machine learning, a published author of books on NLP and deep learning, and a regular speaker at major AI and data science conferences. In 2019, Akshay was named one of the top “40 under 40 data scientists” in India. In his spare time, he enjoys reading, writing, coding, and mentoring aspiring data scientists. He lives in Bangalore, India, with his family.

Introduction

This book takes the user through the journey of natural language processing starting from n-gram models to neural network architectures like RNN before it moves to the state-of-the-art technology today, which is known as the transformers. The book details out the transformer architecture and mainly explains the self-attention mechanism, which is the foundation of the transformer concept.

The book deals with the topic of transformers in depth with examples from different NLP areas like text generation, sentiment analysis, zero-shot learning, text summarization, etc. The book takes a deep dive into huggingface APIs and their usage to create simple Gradio-based applications. We will delve into details of not only using pretrained models but also how to fine-tune the existing models with our own datasets.

We cover models like BERT, GPT2, T5, etc., and showcase how these models can be used directly to create a different range of applications in the area of natural language processing and understanding.

The book doesn't just limit the knowledge and exploration of transformers to NLP but also covers at a high level how transformers are being used in areas like vision.

Source Code

All source code used in this book can be found at github.com/apress/intro-transformers-nlp.