

CONTENTS

INTRODUCTION.....	6
CHAPTER 1. Principles of large language models and pre-training engineering	7
1.1. Overview of large model technology	7
1.2. Data preprocessing.....	24
1.3. LLM pre-training and transfer learning technologies overview.....	32
1.4. Training parallelism on GPUs for LLM.....	57
1.5. Mixed-precision training technology	70
1.6. Other training techniques, such as data loading and batching, gradient checkpoints, gradient accumulation, etc.	73
Questions	83
Practical Tasks	84
CHAPTER 2. Mainstream large language models: trends and technologies.....	86
2.1. A detailed explanation of the core principles of GPT series models: The core principles of GPT series models (GPT1/GPT2/GPT3/GPT4/ InstructGPT).....	86
2.2. Current situation and trend of mainstream LLMs.....	92
2.3. Overview of pre-training technology in the LLaMA 2.....	100
2.4. LLaMA 2 data source: data source, data type, data quality	104
2.5. LLaMA 2 distributed training project.....	112
2.6. LLaMA 2 fundamental model core principles explained.....	113
2.7. LLaMA 2 model source code detailed explanation	115
2.8. API test and trial evaluation of LLaMa 2 large model.....	161
Questions	164
Practical Tasks	165
CHAPTER 3. Basic LLM'S industry large model training practice	167
3.1. Practice Part 1	167
3.2. Practice Part 2 The large model training process in practice	170
3.2.1. Model training environment construction.....	170
3.2.2. Model CPU/GPU training code development.....	172
3.2.3. Model Multi-GPU parallel training code development	174
3.2.4. Model effect evaluation: Modeling quantitative indicators, more commonly [PPL], [BPC], etc.,.....	175
Questions	177
Practical Tasks	178
CHAPTER 4. Large Language Model Fine-Tuning Technologies: Theory and Practice	180
4.1. Overview and core elements of large model fine-tuning techniques	180
4.2. A detailed explanation of large model fine-tuning technology.....	181
4.3. A detailed explanation of large model PEFT technology	182
4.4. Large model RLHF technology in detail	184
4.5. Fine-Tuning Strategies for Specific NLP Tasks	187
4.6. A detailed explanation of large model fine-tuning hyperparameters	188
4.7. Fine-Tuning Configurations: Learning Rate Scheduling, Regularization Techniques, Optimizer Selection, Handling Domain Shifts	193
4.8. Large model fine-tuning code in detail (Practice).....	196
Questions	197
Practical Tasks	198
VOCABLUARY	199