

Contenido

Prefacio	xi
1. Los cinco principios del <i>prompting</i>.....	1
Visión general de los cinco principios del <i>prompting</i>	4
1. Proporcionar orientación	8
2. Especificar el formato.....	13
3. Proporcionar ejemplos.....	16
4. Evaluar la calidad.....	19
5. Dividir la tarea.....	32
Resumen.....	38
2. Introducción a los modelos de lenguaje de gran tamaño para la generación de texto	41
¿Qué son los modelos de generación de texto?.....	41
Representaciones vectoriales: la esencia numérica del lenguaje	42
Arquitectura <i>transformer</i> : orquestación de relaciones contextuales	43
Generación probabilística de texto: el mecanismo de decisión.....	45
Bases históricas: el auge de las arquitecturas de transformadores	46
Transformadores preentrenados generativos de OpenAI.....	48
GPT-3.5-turbo y ChatGPT	48
GPT-4.....	51
Gemini de Google.....	51
Llama de Meta y el código abierto.....	52
Aprovechamiento de la cuantización y LoRA	53
Mistral	53
Anthropic: Claude	54
GPT-4V(ision)	54
Comparación de modelos.....	55
Resumen.....	56
3. Prácticas estándar para la generación de texto con ChatGPT	57
Generación de listas.....	57
Generación de listas jerárquicas	59
Cuándo evitar el uso de expresiones regulares	64
Generación de JSON	64

YAML.....	66
Filtrado de <i>payloads</i> de YAML	68
Tratamiento de <i>payloads</i> no válidos en YAML	69
Generación de distintos formatos con ChatGPT	72
Datos CSV simulados.....	73
Explíquemelo como si tuviera cinco años.....	74
Traducción universal mediante los LLM.....	74
Preguntar por el contexto	76
Desagregación de estilos del texto	80
Identificación de las características deseadas del texto.....	80
Generación de contenido nuevo utilizando las características extraídas.....	82
Extracción de características del texto específicas mediante los LLM.....	82
Elaboración de resúmenes.....	83
Realización de resumen teniendo en cuenta las limitaciones de la ventana contextual.....	84
División del texto en fragmentos.....	85
Ventajas de la división de texto en fragmentos	86
Escenarios para la división del texto en fragmentos	86
Mal ejemplo de la división del texto en fragmentos.....	87
Estrategias de la división de texto en fragmentos.....	88
Detección de oraciones con SpaCy	90
Creación de un algoritmo sencillo de división de texto en fragmentos en Python	91
Fragmentado por ventana deslizante.....	92
Paquetes de división de texto en fragmentos.....	95
División de texto en fragmentos con tiktoken.....	95
Codificaciones.....	95
Comprensión de la tokenización de cadenas.....	96
Estimación del uso de <i>tokens</i> para llamadas a la API de chat	97
Análisis de sentimiento.....	99
Técnicas para mejorar el análisis de sentimiento.....	101
Limitaciones y retos del análisis de sentimiento	101
De menos a más	102
Planificación de la arquitectura	102
Codificación de funciones individuales.....	102
Adición de pruebas.....	103
Ventajas de la técnica de menos a más	104
Retos de la técnica de menos a más.....	104
<i>Role prompting</i>	105
Ventajas del <i>role prompting</i>	106
Retos del <i>role prompting</i>	106
Cuándo utilizar el <i>role prompting</i>	107

Tácticas del <i>prompting</i> de GPT	107
Evitación de alucinaciones con referencias	108
Concesión a los GPT de "tiempo para pensar"	110
Táctica del monólogo interior.....	111
Autoevaluación de las respuestas del LLM.....	112
Clasificación con los LLM	114
Creación de un modelo de clasificación	115
Voto mayoritario para la clasificación	116
Evaluación de criterios.....	117
<i>Meta prompting</i>	120
Resumen.....	124
4. Técnicas avanzadas para la generación de texto con LangChain.....	125
Introducción a LangChain	125
Configuración del entorno	127
Modelos de chat	128
Modelos de chat en <i>streaming</i>	130
Creación de múltiples generaciones de LLM.....	130
Plantillas de <i>prompt</i> en LangChain	131
Lenguaje de expresión LangChain (LCEL)	132
Uso de PromptTemplate con modelos de chat.....	134
Analizadores de salida.....	134
Evaluaciones de LangChain	138
Llamadas a funciones de OpenAI.....	145
Llamadas a funciones en paralelo	149
Llamadas a funciones en LangChain.....	150
Extracción de datos con LangChain.....	152
Planificación de consultas.....	152
Creación de plantillas de <i>prompt few-shot</i>	154
Creación con un número fijo de ejemplos	155
Formateo de los ejemplos	155
Selección de ejemplos <i>few-shot</i> por longitud	156
Limitaciones de los ejemplos <i>few-shot</i>	158
Guardar y cargar <i>prompts</i> para los LLM.....	159
Conexión de datos	160
Cargadores de documentos	162
Divisores de texto	164
División de texto por longitud y tamaño del <i>tokens</i>	165
División de texto con divisiones recursivas de caracteres.....	167
Descomposición de tareas	168
Encadenamiento de <i>prompts</i>	171

Cadena secuencial.....	171
itemgetter y extracción de claves de diccionario.....	173
Estructuración de cadenas LCEL.....	178
Cadenas de documentos.....	179
Stuff.....	181
Refinamiento.....	181
Map reduce.....	182
Map re-rank.....	182
Resumen.....	183
5. Bases de datos vectoriales con FAISS y Pinecone.....	185
Generación mejorada por recuperación (RAG).....	188
Introducción a los embeddings.....	189
Carga de documentos.....	199
Recuperación de memoria con FAISS.....	201
RAG con LangChain.....	206
Bases de datos vectoriales alojadas con Pinecone.....	208
Consulta automática.....	216
Mecanismos alternativos de recuperación.....	221
Resumen.....	222
6. Agentes autónomos con memoria y herramientas.....	223
Cadena de razonamiento.....	223
Agentes.....	225
Reason and Act (ReAct).....	227
Implementación de Reason and Act.....	230
Utilización de las herramientas.....	236
Utilización de los LLM como una API.....	238
Comparación entre las funciones de OpenAI y ReAct.....	242
Casos de uso de las funciones de OpenAI.....	243
ReAct.....	243
Casos de uso para ReAct.....	243
Paquetes de herramientas para agentes.....	244
Personalización de agentes estándar.....	246
Agentes personalizados en LCEL.....	247
Comprensión y uso del contenido de la memoria.....	250
Memoria a largo plazo.....	250
Memoria a corto plazo.....	250
Memoria a corto plazo en agentes de conversación QA.....	251
Memoria en LangChain.....	251
Preservación del estado.....	252

Consulta del estado	253
ConversationBufferMemory	253
Otros tipos de memoria populares en LangChain	256
ConversationBufferWindowMemory	256
ConversationSummaryMemory	256
ConversationSummaryBufferMemory	257
ConversationTokenBufferMemory	257
Agentes de funciones de OpenAI con memoria.....	258
<i>Frameworks</i> avanzados para agentes.....	260
Agentes de planificación y ejecución	260
Árbol de razonamiento	261
<i>Callbacks</i> (funciones de retorno).....	263
<i>Callbacks</i> globales (constructor)	264
<i>Callbacks</i> específicos de solicitud	265
El argumento verbose.....	265
¿Cuándo utilizar Which?	265
Recuento de <i>tokens</i> con LangChain	266
Resumen.....	267

7. Introducción a los modelos de difusión para la generación de imágenes.....269

DALL-E de OpenAI.....	272
Midjourney	274
Stable Diffusion.....	277
Gemini de Google.....	280
Texto a vídeo	280
Comparación de modelos.....	281
Resumen.....	281

8. Prácticas estándar para la generación de imágenes con Midjourney283

Modificadores de formato	283
Modificadores de estilos artísticos.....	287
Ingeniería inversa de <i>prompts</i>	289
Potenciadores de la calidad	290
<i>Prompts</i> negativos.....	292
Términos ponderados	295
<i>Prompts</i> con una imagen	297
<i>Inpainting</i> (retoque de imágenes).....	300
<i>Outpainting</i> (ampliación de imágenes)	302

Personajes consistentes	304
Reescritura de <i>prompts</i>	306
Desagregación de memes.....	309
Mapeo de memes	314
Análisis de <i>prompts</i>	316
Resumen.....	317
9. Técnicas avanzadas de generación de imágenes con Stable Diffusion	319
Ejecución de Stable Diffusion	319
Interfaz web de usuario AUTOMATIC1111	326
Img2Img.....	333
Escalado de imágenes.....	336
Interrogate CLIP	339
SD inpainting y outpainting.....	339
ControlNet.....	343
Modelo de segmentación de todo (SAM).....	352
Ajuste fino de DreamBooth.....	355
Stable Diffusion XL Refiner.....	362
Resumen.....	365
10. Creación de aplicaciones impulsadas por IA.....	367
Redacción de blogs con IA	367
Investigación del tema	369
Entrevista con un experto.....	371
Generación del esquema.....	373
Generación de texto.....	375
Estilo de escritura.....	377
Optimización del título	381
Imágenes de blogs con IA.....	382
Interfaz de usuario.....	387
Resumen.....	390
Acerca de los autores.....	393
Colofón	393