

Curso Completo de Engenharia de Prompt

Da Fundamentação à Engenharia de Agentes

VISÃO GERAL DO CURSO

Objetivo Principal

Capacitar profissionais e iniciantes a dominar a arte e ciência da engenharia de prompt, evoluindo naturalmente desde conceitos fundamentais até técnicas avançadas de engenharia de contexto e agentes de IA, preparando-os para o futuro da interação com sistemas de inteligência artificial.

Público-Alvo

- Iniciantes em IA e LLMs sem conhecimento prévio
- Profissionais que desejam otimizar uso de ferramentas de IA
- Desenvolvedores interessados em automação e agentes
- Gestores e tomadores de decisão em transformação digital

Metodologia Pedagógica

Progressão Natural e Prática:

- Aprender fazendo** - Cada conceito acompanhado de exemplos práticos
- Evolução gradual** - Do simples ao complexo, sem saltos bruscos
- Aplicação imediata** - Exercícios que simulam casos reais
- Iteração e refinamento** - Cultura de melhoria contínua

ARQUITETURA DO CURSO

MÓDULO 1: FUNDAMENTOS DE ENGENHARIA DE PROMPT

Duração estimada: 4-6 horas | Nível: Iniciante

1.1 Introdução ao Mundo dos LLMs

- O que são Large Language Models (LLMs)

- Como funcionam: tokens, contexto e inferência
- Principais modelos em 2025: GPT-4o, Claude 4, Gemini 1.5 Pro
- Diferenças entre modelos e quando usar cada um
- Limitações e capacidades dos LLMs

1.2 O que é Engenharia de Prompt?

- Definição e importância
- Diferença entre prompt casual e prompt engenheirado
- Por que prompt engineering importa para negócios
- Casos de uso reais: legal tech, customer support, healthcare
- Prompt engineering vs fine-tuning vs RAG

1.3 Anatomia de um Prompt

- Componentes básicos: instrução, contexto, input, output
- Estrutura clara e direta
- Importância da especificidade
- Ambiguidade: o maior inimigo dos prompts
- Exemplos práticos de prompts bem e mal estruturados

1.4 Primeiros Passos Práticos

- Escrevendo seu primeiro prompt efetivo
- Técnica: seja claro e direto
- Técnica: use delimitadores (""", ###, XML tags)
- Técnica: especifique formato de saída
- Exercícios práticos com feedback

MÓDULO 2: TÉCNICAS FUNDAMENTAIS

Duração estimada: 6-8 horas | **Nível:** Iniciante-Intermediário

2.1 Zero-Shot Prompting

- Conceito e funcionamento
- Quando usar zero-shot

- Exemplos práticos de classificação
- Exemplos práticos de geração de texto
- Limitações do zero-shot
- Exercícios progressivos

2.2 Few-Shot Prompting (Aprendizado por Exemplos)

- Conceito de in-context learning
- Estrutura de few-shot prompts
- Quantos exemplos usar? (1-shot, 3-shot, 5-shot)
- Qualidade vs quantidade de exemplos
- Método Tell and Show
- Método Talent Show (mostrar o bom e o ruim)
- Exercícios práticos com diferentes domínios

2.3 Chain of Thought (CoT) - Cadeia de Pensamento

- O que é raciocínio passo a passo
- Por que CoT melhora resultados complexos
- CoT explícito vs implícito
- Prompt: "Vamos pensar passo a passo"
- Aplicações em matemática, lógica e análise
- Exercícios de problemas complexos

2.4 Uso de Papéis e Personas (Role Prompting)

- Atribuindo papéis ao modelo
- System prompts vs user prompts
- Criando personas efetivas
- Tom, estilo e expertise
- Exemplos: especialista, professor, analista
- Exercícios de criação de personas

2.5 Formatação e Estruturação

- Uso de XML tags para organização

- Uso de Markdown headers
 - Separação clara de seções
 - Templates reutilizáveis
 - Método de Empilhagem (Stacking Method)
 - Exercícios de estruturação
-

MÓDULO 3: TÉCNICAS INTERMEDIÁRIAS

Duração estimada: 8-10 horas | **Nível:** Intermediário

3.1 Instruções Negativas (Anti-Keyword Staining)

- O que evitar vs o que fazer
- Direcionando comportamento por negação
- Casos de uso: evitar hashtags, emojis, jargões
- Balanceamento entre positivo e negativo
- Exercícios práticos

3.2 Prefilling e Controle de Saída

- Técnica de prefill da resposta
- Forçando formatos específicos
- Controle de tom desde o início
- Aplicações em geração estruturada
- Exercícios de prefilling

3.3 Decomposição de Tarefas Complexas

- Método da Decomposição (Deconstruction Method)
- Quebrando problemas grandes em pequenos
- Sequenciamento lógico de subtarefas
- Quando decompor vs quando manter junto
- Exercícios de decomposição

3.4 Encadeamento de Prompts (Chaining)

- Método de Encadeamento (Chaining Method)

- Conectando múltiplos prompts
- Passagem de contexto entre etapas
- Pipelines de processamento
- Exercícios de criação de chains

3.5 Método de Importação (Import Method)

- Reutilizando conhecimento externo
- Incorporando dados e referências
- Contexto dinâmico
- Exercícios de importação

3.6 Ajuste de Parâmetros

- Temperatura: criatividade vs determinismo
 - Top_p (nucleus sampling)
 - Max tokens e stop sequences
 - Quando ajustar cada parâmetro
 - Exercícios de otimização
-

MÓDULO 4: TÉCNICAS AVANÇADAS

Duração estimada: 10-12 horas | **Nível:** Avançado

4.1 Prompting para Diferentes Modelos

- Diferenças entre GPT-4o, Claude 4, Gemini
- Otimização específica por modelo
- Não existe "melhor prática universal"
- Testes comparativos
- Exercícios multi-modelo

4.2 Long Context Tips

- Trabalhando com contextos extensos
- Context rot e degradação de atenção
- Estratégias para contextos longos

- Posicionamento de informação crítica
- Exercícios com documentos longos

4.3 Structured Outputs

- Forçando JSON, XML, YAML
- Schemas e validação
- Extração de dados estruturados
- LlamaExtract e ferramentas similares
- Exercícios de extração estruturada

4.4 Multimodal Prompting

- Trabalhando com imagens
- Trabalhando com áudio/voz
- Combinando modalidades
- Melhores práticas para cada tipo
- Exercícios multimodais

4.5 Prompt Optimization e Testing

- Criando avaliações empíricas
- Métricas de sucesso
- A/B testing de prompts
- Iteração baseada em dados
- Documentação de versões
- Exercícios de otimização

MÓDULO 5: ENGENHARIA DE CONTEXTO

Duração estimada: 8-10 horas | **Nível:** Avançado

5.1 Da Engenharia de Prompt à Engenharia de Contexto

- Evolução do campo
- Contexto como recurso finito
- "Orçamento de atenção" dos LLMs

- Pensar em contexto (thinking in context)
- Curadoria iterativa vs prompt estático

5.2 Anatomia de Contexto Efetivo

- Princípio: menor conjunto de tokens de alto sinal
- Componentes do contexto: system, tools, MCP, data, history
- Altitude certa: específico vs flexível
- Organização em seções
- Exercícios de otimização de contexto

5.3 Context Rot e Limitações Arquiteturais

- O que é context rot
- Arquitetura transformer e atenção n^2
- Position encoding interpolation
- Degradação gradual vs cliff
- Estratégias de mitigação

5.4 Técnicas de Context Engineering

- **Ordenação de contexto:** ranking por relevância
- **Compressão de contexto:** summarization
- **Memória de longo prazo:** VectorMemoryBlock, FactExtraction
- **Informação estruturada:** schemas e condensação
- **Workflow engineering:** sequências otimizadas
- Exercícios práticos de cada técnica

5.5 RAG (Retrieval Augmented Generation)

- Conceito e arquitetura
 - Quando usar RAG vs prompt engineering
 - Vector stores e embeddings
 - Retrieval strategies
 - Integração com context engineering
 - Exercícios de implementação RAG
-

MÓDULO 6: AGENTES DE IA - FUNDAMENTOS

Duração estimada: 10-12 horas | **Nível:** Avançado

6.1 Introdução aos Agentes de IA

- O que são agentes de IA
- Diferença entre LLM e agente
- Componentes de um agente: percepção, ação, memória
- Agentic workflows
- Casos de uso de agentes

6.2 Ferramentas (Tools) para Agentes

- Definição e design de tools
- Contratos entre agentes e ambiente
- Eficiência de tokens em tools
- Descrições claras de ferramentas
- Function calling
- Exercícios de criação de tools

6.3 Model Context Protocol (MCP)

- O que é MCP
- Arquitetura e componentes
- Integração com agentes
- Casos de uso práticos
- Exercícios com MCP

6.4 Memória e Estado em Agentes

- Memória de curto prazo (working memory)
- Memória de longo prazo (persistent memory)
- Gerenciamento de estado
- Context window como memória de trabalho
- Exercícios de implementação

6.5 Loops Agênticos

- Percepção → Raciocínio → Ação → Observação
 - Controle de loops
 - Condições de parada
 - Error handling em loops
 - Exercícios de loops
-

MÓDULO 7: AGENTES AVANÇADOS E SKILLS

Duração estimada: 10-12 horas | **Nível:** Avançado-Expert

7.1 Claude Skills (Agent Skills)

- O que são Skills da Anthropic
- Arquitetura de Skills: pastas, instruções, scripts, recursos
- Composabilidade de Skills
- Portabilidade entre produtos Claude
- Criação de Skills customizadas
- Skill-creator skill
- Exercícios de criação de Skills

7.2 Extended Thinking

- O que é extended thinking mode
- Claude 3.7 Sonnet e pensamento visível
- Quando usar extended thinking
- Técnicas de prompting para extended thinking
- Think tool
- Exercícios práticos

7.3 Agentes de Automação

- Workflows automatizados
- Integração com APIs e sistemas externos
- Orquestração de tarefas
- Monitoramento e logging

- Casos de uso: RPA, processamento de documentos
- Exercícios de automação

7.4 Agentes de Voz (Voice AI Agents)

- Arquitetura de agentes de voz
- Latência e tempo real (sub-100ms)
- Conversational AI
- Plataformas: ElevenLabs, Voiceflow, Google Dialogflow
- Casos de uso: atendimento, qualificação de leads
- Exercícios de voice agents

7.5 Multi-Agent Systems

- Coordenação entre múltiplos agentes
 - Especialização de agentes
 - Comunicação inter-agentes
 - Casos de uso avançados
 - Exercícios de sistemas multi-agente
-

MÓDULO 8: MASTERCLASSES

Duração estimada: 12-15 horas | **Nível:** Expert

Masterclass 1: Engenharia de Contexto Avançada

- Deep dive em context optimization
- Casos complexos de curadoria de contexto
- Debugging de problemas de contexto
- Arquiteturas escaláveis
- Projeto prático completo

Masterclass 2: Prompt Engineering para Produção

- Da prototipação à produção
- Versionamento e CI/CD de prompts
- Monitoramento e observabilidade

- Segurança: prompt injection, jailbreaking
- Mitigação de riscos
- Projeto prático completo

Masterclass 3: Construindo Agentes Robustos

- Arquitetura de agentes para produção
- Reliability e error handling
- Testing de agentes
- Deployment e escalabilidade
- Casos de estudo reais
- Projeto prático completo

Masterclass 4: O Futuro da Engenharia de Agentes

- Tendências emergentes
 - Agentic AI e o futuro do trabalho
 - Considerações éticas
 - Preparação para próximas gerações de modelos
 - Discussão e networking
-

RECURSOS COMPLEMENTARES

Materiais de Apoio

- Biblioteca de templates de prompts
- Repositório de exemplos práticos
- Checklists de boas práticas
- Glossário completo de termos
- Referências e leituras adicionais

Ferramentas Recomendadas

- Claude Console (Anthropic)
- OpenAI Playground
- LangChain / LlamaIndex

- Prompt engineering frameworks
- Testing e evaluation tools

Comunidade e Suporte

- Fórum de discussão
 - Sessões de Q&A ao vivo
 - Feedback em projetos
 - Networking com outros alunos
-

AVALIAÇÃO E CERTIFICAÇÃO

Critérios de Avaliação

- Exercícios práticos em cada módulo (40%)
- Projetos de aplicação real (30%)
- Masterclass projects (20%)
- Participação e engajamento (10%)

Certificação

- Certificado de conclusão para quem completar 80% do curso
 - Certificado de excelência para quem atingir 90%+ nas avaliações
 - Badge digital para compartilhar em redes profissionais
-

CARGA HORÁRIA TOTAL

Estimativa: 68-85 horas de conteúdo + exercícios práticos

Sugestão de ritmo:

- **Intensivo:** 4-6 semanas (15-20h/semana)
 - **Regular:** 8-12 semanas (7-10h/semana)
 - **Flexível:** Ao seu próprio ritmo
-

DIFERENCIAIS DO CURSO

1. **Progressão Natural:** Evolução gradual e lógica do básico ao avançado
2. **Prática Intensiva:** Cada conceito acompanhado de exercícios reais
3. **Conteúdo Atualizado:** Baseado em 2025, incluindo últimas técnicas
4. **Foco em Aplicação:** Não apenas teoria, mas aplicação prática imediata
5. **Preparação Completa:** Do prompt básico à engenharia de agentes
6. **Visão Holística:** Integração entre prompt, contexto e agentes
7. **Materiais Ricos:** Templates, exemplos, checklists e recursos
8. **Design Elegante:** Materiais visuais profissionais e atraentes