Glossário de Inteligência Artificial

# Glossário de Inteligência Artificial

## Conceitos Fundamentais de IA

1. **Inteligência Artificial (IA)**: Campo da ciência da computação que busca criar sistemas capazes de realizar tarefas que normalmente exigem inteligência humana, como reconhecer imagens, compreender linguagem ou tomar decisões.
2. **Aprendizado de Máquina (Machine Learning)**: Subárea da IA em que os computadores “aprendem” a partir de dados, sem serem explicitamente programados para cada tarefa.
3. **Redes Neurais Artificiais (Artificial Neural Networks)**: Modelos inspirados no cérebro humano, compostos por “neurônios artificiais” conectados, capazes de reconhecer padrões complexos em dados.
4. **Dados**: A matéria-prima da IA. Quanto mais numerosos, variados e de qualidade forem os dados, melhor tende a ser o desempenho do modelo.
5. **Algoritmo**: Conjunto de instruções ou regras que o computador segue para resolver um problema ou executar uma tarefa.
6. **Treinamento**: Processo de ajustar os parâmetros de um modelo de IA até que ele aprenda a realizar bem determinada tarefa.
7. **Inferência**: Etapa em que o modelo já treinado aplica o que aprendeu para analisar novos dados e dar respostas ou previsões.
8. **Generalização**: Capacidade do modelo de aplicar o que aprendeu no treinamento para situações novas, diferentes dos exemplos originais.
9. **Sobreajuste (Overfitting)**: Quando o modelo “decorou” demais os dados de treino e não consegue lidar bem com exemplos novos.
10. **Subajuste (Underfitting)**: Quando o modelo é simples ou pouco treinado e não consegue nem aprender direito os dados de treino.
11. **Modelo**: Uma Inteligência Artificial específica, implementada / configurada para uso.
12. **Aprendizado Supervisionado**: Usa dados rotulados (já identificados previamente) para ensinar o modelo a realizar uma tarefa.
13. **Aprendizado Não Supervisionado**: Busca padrões, estruturas ou agrupamentos em dados sem rótulos ou exemplos específicos. Exemplos: agrupar clientes similares, detectar anomalias, reduzir dimensionalidade de dados.
14. **Aprendizado por Reforço**: Método onde um agente aprende através de tentativa e erro, recebendo recompensas ou penalidades baseadas em suas ações. O agente interage com um ambiente e aprende a maximizar recompensas ao longo do tempo. Exemplos: jogos (AlphaGo, Dota 2), robótica, sistemas de recomendação adaptativos.
15. **Classificação**: Tarefa em que o modelo aprende a colocar dados em categorias ou classes específicas. Exemplos: identificar se um email é spam ou não spam, classificar imagens como gato ou cachorro, categorizar sentimentos como positivo, negativo ou neutro.
16. **Regressão**: Tarefa em que o modelo aprende a prever valores numéricos contínuos. Exemplos: prever o preço de uma casa baseado em suas características, estimar a temperatura de amanhã, calcular o tempo de entrega de um produto.
17. **Viés e Variância**: Duas fontes de erro em modelos: viés é simplificação exagerada, variância é excesso de sensibilidade aos dados. O ideal é equilibrar ambos.
18. **Parâmetro**: Variável ajustável em um modelo de IA que é otimizada durante o treinamento para melhorar o desempenho e a precisão das previsões.
19. **Aprendizado Profundo (Deep Learning)**: Subárea do machine learning que utiliza redes neurais artificiais com múltiplas camadas para aprender representações hierárquicas e complexas de dados, permitindo tarefas como reconhecimento de imagens e processamento de linguagem natural.
20. **Processamento de Linguagem Natural (Natural Language Processing, NLP)**: Campo da IA que desenvolve sistemas capazes de entender, interpretar e gerar linguagem humana. Inclui tarefas como tradução automática, análise de sentimentos, chatbots, sumarização de textos e compreensão de perguntas.
21. **Visão Computacional (Computer Vision)**: Campo da IA que permite que computadores “vejam” e interpretem imagens e vídeos, realizando tarefas como reconhecimento de objetos, detecção de faces, segmentação de imagens e análise de movimento. Fundamental para aplicações como carros autônomos, sistemas de segurança e diagnóstico médico.
22. **Pesos e Vieses (Weights and Biases, W&B)**: Parâmetros ajustáveis em modelos de IA: os **pesos (weights)** multiplicam entradas/ativação nas camadas, enquanto os **vieses (biases)** deslocam a função de ativação, permitindo ajustes finos. Juntos, determinam como o modelo transforma entradas em saídas durante treinamento e inferência.

## Conceitos Relacionados à IA Generativa

1. **IA Generativa**: Área da IA voltada para a criação de novos conteúdos — textos, imagens, sons, vídeos — a partir de padrões aprendidos em grandes volumes de dados.
2. **Modelos de Linguagem (Language Models, LMs)**: Modelos capazes de compreender e gerar texto em linguagem natural.
3. **Modelos de Linguagem de Grande Porte (Large Language Models, LLMs)**: Modelos de linguagem com bilhões de parâmetros, capazes de compreender e gerar texto em linguagem natural com alta qualidade. Exemplos: GPT-4, Claude, LLaMA. Requerem infraestrutura computacional significativa para treinamento e inferência.
4. **Modelos de Linguagem de Pequeno Porte (Small Language Models, SLMs)**: Modelos de linguagem com menos parâmetros (milhões a poucos bilhões), otimizados para eficiência computacional e execução em dispositivos com recursos limitados. Equilibram qualidade e velocidade para aplicações específicas.
5. **Modelos Multimodais (Multimodal Models)**: Sistemas de IA capazes de processar e gerar múltiplos tipos de dados simultaneamente (texto, imagem, áudio, vídeo). Permitem tarefas como descrição de imagens, geração de conteúdo baseado em diferentes mídias e compreensão contextual rica.
6. **Modelos de Raciocínio (Reasoning Models)**: Modelos de linguagem que podem realizar tarefas de raciocínio complexo por meio de técnicas como *cadeia‑de‑pensamento* e *auto-reflexão* que produzem uma série de tokens internos, que são então usados para gerar a resposta final. O modelo utiliza mais tempo de computação para gerar respostas mais precisas para tarefas complexas.
7. **Mistura de Especialistas (Mixture of Experts, MoE)**: Arquitetura em que vários “especialistas” (sub-redes) coexistem e um roteador escolhe apenas alguns para cada entrada/token. Isso permite um modelo grande (com muitos parâmetros) rodar com um custo de computação menor.
8. **Prompt**: A entrada (pergunta, instrução ou comando) que o usuário fornece para que o modelo generativo produza uma resposta.
9. **Token**: Unidade mínima em que o texto é dividido para processamento. Pode ser uma palavra, parte dela ou até um caractere.
10. **Vetorização Semântica (Embeddings)**: Representação matemática de palavras, frases ou documentos em vetores, permitindo que modelos entendam relações semânticas entre eles.
11. **Transformer**: Arquitetura de rede neural que revolucionou a IA generativa, permitindo o processamento eficiente de sequências longas de dados.
12. **Treinamento Prévio (Pretraining)**: Fase inicial em que o modelo aprende padrões gerais da linguagem ou de imagens a partir de enormes conjuntos de dados.
13. **Ajuste Fino (Fine-Tuning)**: Etapa em que o modelo é refinado com dados específicos, adaptando-o a tarefas ou contextos particulares.
14. **Ajuste por Instrução (Instruction Tuning)**: Técnica para treinar modelos a seguirem melhor instruções dadas em linguagem natural.
15. **Alucinação (Hallucination)**: Quando um modelo generativo cria informações falsas ou inventadas que parecem plausíveis.
16. **Geração Aumentada por Recuperação (Retrieval-Augmented Generation, RAG)**: Técnica que combina geração de texto com recuperação de informações em bases externas (ex.: documentos, web, bancos de dados), aumentando a precisão das respostas.
17. **Modelos de Difusão (Diffusion Models)**: Modelos de IA usados na geração de imagens, que criam figuras a partir de ruído, refinando-as progressivamente até se tornarem claras. Recentemente, modelos de difusão começaram a ser usados para geração de texto também.
18. **Redes Adversárias Generativas (Generative Adversarial Networks, GANs)**: Modelos que usam duas redes neurais em competição (gerador e discriminador) para criar conteúdos melhores.
19. **Aprendizado por Reforço com Feedback Humano (Reinforcement Learning from Human Feedback, RLHF)**: Método de alinhar modelos generativos às preferências humanas, usando reforço baseado em feedback de avaliadores.
20. **Text to Speech (TTS)**: Tecnologia que converte texto escrito em fala natural, permitindo que sistemas de IA “falem” com entonação, ritmo e pronúncia humanas. Usada em assistentes virtuais, audiobooks e acessibilidade.
21. **Speech to Text (STT)**: Tecnologia que converte fala humana em texto escrito, permitindo que sistemas de IA compreendam comandos de voz, transcrevam conversas e facilitem a interação por voz. Base da tecnologia de reconhecimento de fala.
22. **Quantização (Quantization)**: Reduz a precisão numérica de pesos/ativações (ex.: 16→8→4 bits) para economizar **memória, custo e tempo de inferência**. Pequena perda de acurácia é comum.
23. **Destilação (Distillation)**: Técnica que treina de uma IA aluna menor a imitar uma IA professor maior, usando os dados sintéticos gerados pela IA professor. Resultado: modelo mais leve e rápido, com boa parte do desempenho preservado.
24. **Janela de Contexto (Context Window)**: Limite máximo de tokens (palavras, caracteres) que um modelo de linguagem pode processar em uma única interação. Determina quanto texto o modelo “lembra” durante a conversa e afeta diretamente o custo computacional e a qualidade das respostas.
25. **Dados Sintéticos (Synthetic Data)**: Dados gerados artificialmente a partir de modelos de IA, diferentes de dados gerador por humanos.

## Agentes de IA

1. **Agente de IA**: Sistema que percebe o ambiente, decide e age para atingir objetivos. Pode ser puramente digital (bot, assistente) ou físico (robô). Opera no ciclo *perceber → decidir → agir*.
2. **Ambiente**: O contexto onde o agente atua (tudo que está “fora” dele). Pode ser totalmente ou parcialmente observável, determinístico ou estocástico, estático ou dinâmico.
3. **Observações (Percepção)**: Informações que o agente recebe do ambiente: texto de usuários, páginas web, leituras de sensores, respostas de APIs etc. Servem de base para decidir a próxima ação.
4. **Ações**: O que o agente pode fazer: responder em texto, clicar/navegar, chamar ferramentas (APIs), executar código, mover atuadores (no caso de robôs).
5. **Estado**: Representação do que é relevante no momento para decidir. Pode ser o próprio contexto de um LLM, uma memória de curto prazo ou um “estado oculto” quando parte do ambiente não é observável.
6. **Objetivo (Goal)**: Define o que o agente busca e orienta suas decisões e ações. Exemplos: “responder ao cliente”, “navegar até um destino”, “resolver um problema matemático”. O objetivo serve como bússola para o comportamento do agente.
7. **Recompensa**: Sinal (geralmente numérico) que indica sucesso parcial ou total — comum em aprendizado por reforço. Exemplo: nota ao concluir uma tarefa.
8. **Política (Policy)**: Regra que mapeia estados/observações em ações. Pode ser uma rede neural treinada, um conjunto de regras ou um *prompt* que orienta decisões de um LLM‑agente.
9. **Planejamento (Planning)**: Escolha de sequências de ações para alcançar objetivos.
10. **Memória e Contexto**: Mecanismos para manter continuidade: janela de contexto (curto prazo), memória vetorial/BD (longo prazo) e logs. Ajudam a não “recomeçar do zero” a cada passo.
11. **Ferramentas**: Conectores e interfaces externas que um agente pode acionar para ampliar suas capacidades: calculadora, navegação e busca web, RAG/bases de conhecimento, planilhas e bancos de dados, interpretador/execução de código, automações e serviços de terceiros, além de atuadores físicos (robótica).
12. **Sistemas Multiagentes (Multi-Agent Systems, MAS)**: Vários agentes cooperando com papéis complementares (ex.: pesquisador, redator, revisor). Exigem protocolos de comunicação e estratégias de coordenação.
13. **Autonomia e Supervisão Humana (Human-in-the-Loop, HITL)**: Definição de níveis de autonomia, checkpoints de aprovação, trilhas de auditoria e limites de escopo para segurança, conformidade e alinhamento com objetivos humanos.
14. **Persona ou Papel (Role)**: Configuração que define estilo, tom, personalidade e características de um agente. (ex.: “tutor paciente”, “analista jurídico cauteloso”).
15. **Enxame (Swarm)**: Abordagem multiagente inspirada em inteligência de enxame (formigas, abelhas, cardumes), em que muitos agentes simples cooperam com regras locais para resolver problemas. Utiliza mecanismos como votação, leilão de tarefas, sinais/“feromônios” virtuais e difusão de mensagens para convergência.
16. **Agente Autônomo**: Sistema de IA que opera independentemente com mínima supervisão humana, tomando decisões e executando ações baseadas em seus objetivos e percepções do ambiente. Pode aprender e adaptar-se a mudanças, mas mantém limites de segurança e conformidade definidos por humanos.

## Escopo das IAs

1. **Inteligência Artificial Restrita (Artificial Narrow Intelligence, ANI)**: Sistemas especializados em tarefas específicas. Excelentes dentro de um domínio, mas não generalizam sozinhos para outros. Exemplos: filtros de spam, recomendação de conteúdo, reconhecimento de fala, jogar xadrez, condução assistida. Também chamada de *IA estreita* ou *IA fraca*.
2. **Inteligência Artificial Geral (Artificial General Intelligence, AGI)**: Sistema capaz de desempenhar uma ampla variedade de tarefas, em nível comparável ao humano, com raciocínio, planejamento e adaptação a contextos novos. Conceito ainda em debate e sem consenso. Uma definição possível é de um “sistema capaz de executar qualquer tarefa cognitiva em um nível pelo menos equivalente a qualquer ser humano”.
3. **Inteligência Artificial Superior (Artificial Super Intelligence, ASI)**: Sistemas que excedem amplamente o desempenho humano em praticamente todas as tarefas cognitivas relevantes. Podem realizar auto‑aperfeiçoamento e apresentar capacidades estratégicas elevadas. Trazem desafios de **alinhamento**, governança e segurança substanciais.

## Conceitos Éticos, de Segurança e de Governança

1. **IA Responsável (Responsible AI)**: Conjunto de princípios e práticas para projetar, treinar, avaliar e operar sistemas de IA de forma segura, ética e justa.
2. **IA Ética (Ethical AI)**: O mesmo que IA Responsável, mas com um foco maior na ética.
3. **Viés e Equidade (Bias and Fairness)**: Identificação e mitigação de preconceitos nos modelos de IA.
4. **Explicabilidade (XAI)**: Capacidade de um modelo em explicar, de maneira confiável, como chegou a uma determinada decisão.
5. **Interpretabilidade**: Área de pesquisa que busca compreender os mecanismos internos de um modelo que geram uma determinada saída ou decisão.
6. **Observabilidade de IA**: Monitoramento em tempo real sobre o comportamento de sistemas de IA, facilitando a detecção de anomalias.
7. **Responsabilização Algorítmica (Algorithmic Accountability)**: Responsabilidade pelos impactos das decisões automatizadas.
8. **Guard Rails**: Conjuntos de restrições e limites implementados em sistemas de IA para possibilitar operações seguras e éticas.
9. **Guidelines**: Diretrizes que orientam o desenvolvimento e uso responsável de sistemas de IA.
10. **Segurança de IA (AI Safety)**: Área focada em garantir que sistemas de IA funcionem de forma segura e alinhada aos valores humanos. Aborda riscos como alinhamento incorreto, comportamento inesperado e capacidades emergentes que podem representar ameaças à humanidade.
11. **Segurança de IA (AI Security)**: Proteção de sistemas de IA contra ataques, vazamentos de dados e uso malicioso. Inclui resistência a *ataques de prompt*, defesa contra *exemplos adversariais* e prevenção de *jailbreaks*.
12. **Alinhamento em IA (AI Alignment)**: Área de pesquisa que busca garantir que sistemas de IA, mesmo aqueles super inteligentes, ajam de forma alinhada aos valores e objetivos humanos.
13. **Robustez**: Capacidade de um sistema de IA de manter desempenho confiável mesmo em condições adversas ou com dados inesperados. Inclui resistência a falhas, ruído e variações nos dados de entrada.
14. **Ataques de Prompt (Prompt Injection)**: Técnicas que induzem o modelo a ignorar instruções/limites e gerar saídas indesejadas.
15. **Jailbreak**: Técnica que induz o modelo a ignorar restrições e agir de maneira livre, possivelmente de maneira maliciosa ou insegura.
16. **Exemplos Adversariais**: Perturbações sutis em entradas (texto/imagem/áudio) que levam o modelo ao erro. Defesas: *adversarial training*, regularização e detecção de anomalias.
17. **Deriva de Dados (Data Drift)**: Mudanças na distribuição dos dados de entrada ao longo do tempo que podem degradar o desempenho do modelo. Exige *monitoramento contínuo* e estratégias de adaptação.
18. **Deriva de Modelo (Model Drift)**: Degradação do desempenho do modelo devido a mudanças no ambiente ou nos dados. Requer treinamento periódico para manutenção da qualidade.
19. **Monitoramento Contínuo**: Acompanhamento sistemático do desempenho e comportamento dos sistemas de IA em produção, incluindo métricas, alertas e detecção de anomalias.
20. **Conformidade**: Adequação de sistemas de IA a marcos regulatórios, normas técnicas e políticas internas. Inclui avaliações de risco/impacto, classificações por nível de risco e requisitos de documentação.
21. **Regulação de IA**: Marco legal e normativo que estabelece regras para o desenvolvimento, uso e comercialização de sistemas de IA. Define responsabilidades, limites e requisitos de segurança para diferentes aplicações.
22. **Marcação de Conteúdo Sintético**: Técnicas para indicar que um conteúdo foi gerado por IA (ex.: marcas d’água, assinaturas criptográficas, padrões de proveniência). Auxiliam no combate a *deepfakes* e desinformação.
23. **Aprendizado Federado (Federated Learning)**: Treinamento de IA descentralizado para preservar a privacidade dos dados (ver *anonimização*).
24. **Anonimização**: Processo de remover ou alterar informações identificáveis de dados para proteger a identidade dos indivíduos, garantindo a privacidade em tráfego, análises e armazenamento.

## Habilidades e Práticas para Profissionais de IA

1. **Vibe Coding**: Prática exploratória de criar programas / apps / sites / etc. utilizando IA sem a necessidade de conhecimentos prévios de programação, priorizando fluxo criativo, testes curtos e aprendizagem pelo fazer.
2. **Fusion Skills**: Combinação integrada de expertise humana (domínio, criatividade, julgamento) com capacidades da IA (geração, busca, análise). O profissional alterna papéis: orquestra, critica e complementa a IA.
3. **Engenharia de Prompt (Prompt Engineering)**: Técnicas para criar *prompts* eficazes: papéis, objetivos claros, restrições, exemplos, formatos de saída e critérios de qualidade/teste.
4. **Engenharia de Contexto (Context Engineering)**: Gestão da *janela de contexto*: seleção, compressão e ordenação de informações relevantes; uso de *prefill*, *system prompts*, anotações e memória para aumentar qualidade e consistência.
5. **Zero‑shot e Few‑shot**: Formas de orientar o modelo sem (zero) ou com poucos exemplos (few). Úteis para ensinar formato, estilo e critérios de avaliação dentro do próprio *prompt*.
6. **Cadeia‑de‑Pensamento (Chain‑of‑Thought, CoT)**: Estratégias para induzir passos de raciocínio durante a geração de texto, com o objetivo de melhorar a qualidade das respostas.
7. **Auto‑reflexão (Self‑Reflection)**: Padrão onde o modelo analisa a própria saída, identifica falhas e propõe correções (“critique, depois revise”). Aumenta qualidade em tarefas complexas.
8. **Auto‑consistência (Self‑Consistency)**: Executar múltiplas amostragens e escolher a resposta mais consistente/consensual. Reduz alucinações e melhora raciocínio em problemas abertos.
9. **Decomposição de Tarefas (Task Decomposition)**: Quebrar objetivos amplos em etapas mensuráveis (ex.: checklists).
10. **Design de Guard‑Rails**: Arquitetar políticas, classificadores e checagens (pré/pós) para manter o sistema dentro de limites de segurança, conformidade e tom de marca.
11. **Red Teaming de IA**: Testes ofensivos controlados para descobrir falhas, vieses e caminhos de abuso (*ataques de prompt*, *jailbreak*, fuga de dados), com planos de correção.

## Infraestrutura e Processos

1. **API (Application Programming Interface)**: Interface que permite que programas se comuniquem entre si.
2. **MCP (Model Context Protocol)**: Protocolo que permite que agentes de IA se comuniquem entre si e com outros sistemas.
3. **Pipeline de Dados (Data Pipeline)**: Cadeia de processamento e preparação de dados para alimentar modelos de IA.
4. **Computação de Borda (Edge Computing)**: Execução de modelos **próxima da fonte dos dados** (celulares, IoT, veículos), reduzindo latência e custo de tráfego e aumentando privacidade.
5. **Orquestração de IA**: Criar sinergia entre humanos e automação inteligente, coordenando sistemas, modelos e fluxos de trabalho de forma colaborativa. Vai além de tarefas — integra processos com inteligência adaptativa.
6. **Workflows Inteligentes**: Processos automatizados com IA que tomam decisões, aprendem com dados e se adaptam ao contexto, integrando sistemas de forma dinâmica e responsiva entre sistemas de IA e humanos.