Glossário de Inteligência Artificial

**Total de termos: 112**Data: 27/08/2025

# Conceitos Fundamentais

## Algoritmo (Algorithm)

Conjunto de instruções ou regras que o computador segue para resolver um problema ou executar uma tarefa.

## Aprendizado de Máquina (Machine Learning)

Subárea da IA em que os computadores "aprendem" a partir de dados, sem serem explicitamente programados para cada tarefa.

## Aprendizado Não Supervisionado (Unsupervised Learning)

Busca padrões, estruturas ou agrupamentos em dados sem rótulos. Exemplos: agrupar clientes similares.

## Aprendizado por Reforço (Reinforcement Learning)

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.

## 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.

## Aprendizado Supervisionado (Supervised Learning)

Usa dados rotulados (já identificados previamente) para ensinar o modelo a realizar uma tarefa.

## Classificação (Classification)

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.

## Dados (Data)

A matéria-prima da IA. Quanto mais numerosos, variados e de qualidade forem os dados, melhor tende a ser o desempenho do modelo.

## Generalização (Generalization)

Capacidade do modelo de aplicar o que aprendeu no treinamento para situações novas, diferentes dos exemplos originais.

## Inferência (Inference)

Etapa em que o modelo já treinado aplica o que aprendeu para analisar novos dados e dar respostas ou previsões.

## Inteligência Artificial (Artificial Intelligence - IA, AI)

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.

## Modelo (Model)

Uma Inteligência Artificial específica, implementada/configurada para uso.

## Parâmetro (Parameter)

Variável ajustável em um modelo de IA que é otimizada durante o treinamento para melhorar o desempenho e a precisão das previsões.

## 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.

## Processamento de Linguagem Natural (Natural Language Processing - PLN, 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.

## Redes Neurais Artificiais (Artificial Neural Networks)

Modelos inspirados no cérebro humano, compostos por "neurônios artificiais" conectados, capazes de reconhecer padrões complexos em dados.

## Regressão (Regression)

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.

## Sobreajuste (Overfitting)

Quando o modelo "decorou" demais os dados de treino e não consegue lidar bem com exemplos novos.

## Subajuste (Underfitting)

Quando o modelo é simples ou pouco treinado e não consegue nem aprender direito os dados de treino.

## Treinamento (Training)

Processo de ajustar os parâmetros de um modelo de IA até que ele aprenda a realizar bem determinada tarefa.

## Viés e Variância (Bias and Variance)

Duas fontes de erro em modelos: viés é simplificação exagerada, variância é excesso de sensibilidade aos dados. O ideal é equilibrar ambos.

## 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.

# IA Generativa

## Ajuste Fino (Fine-Tuning)

Etapa em que o modelo é refinado com dados específicos, adaptando-o a tarefas ou contextos particulares.

## Ajuste por Instrução (Instruction Tuning)

Técnica para treinar modelos a seguirem melhor instruções dadas em linguagem natural.

## Alucinação (Hallucination)

Quando um modelo generativo cria informações falsas ou inventadas que parecem plausíveis.

## 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.

## Dados Sintéticos (Synthetic Data)

Dados gerados artificialmente a partir de modelos de IA, diferentes de dados gerados por humanos.

## Destilação (Distillation)

Técnica que treina 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.

## Vetorização Semântica (Embeddings)

Representação matemática de palavras, frases ou documentos em vetores, permitindo que modelos entendam relações semânticas entre eles.

## Redes Adversárias Generativas (Generative Adversarial Networks - GANs)

Modelos que usam duas redes neurais em competição (gerador e discriminador) para criar conteúdos melhores.

## IA Generativa (Generative AI)

Área da IA voltada para a criação de novos conteúdos — textos, imagens, sons, vídeos — a partir de padrões aprendidos em grandes volumes de dados.

## 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.

## 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.

## 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.

## 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.

## 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.

## Modelos de Linguagem (Language Models - LMs)

Modelos capazes de compreender e gerar texto em linguagem natural.

## 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.

## 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.

## Prompt (Prompt)

A entrada (pergunta, instrução ou comando) que o usuário fornece para que o modelo generativo produza uma resposta.

## 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.

## 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.

## Speech to Text (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.

## Text to Speech (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.

## Token (Token)

Unidade mínima em que o texto é dividido para processamento. Pode ser uma palavra, parte dela ou até um caractere.

## Transformer (Transformer)

Arquitetura de rede neural que revolucionou a IA generativa, permitindo o processamento eficiente de sequências longas de dados.

## Treinamento Prévio (Pretraining)

Fase inicial em que o modelo aprende padrões gerais da linguagem ou de imagens a partir de enormes conjuntos de dados.

# Agentes de IA

## Ações (Actions)

O que o agente pode fazer: responder em texto, clicar/navegar, chamar ferramentas (APIs), executar código, mover atuadores (no caso de robôs).

## Agente Autônomo (Autonomous Agent)

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.

## Agente de IA (AI Agent)

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.

## Agente (Agent)

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.

## Ambiente (Environment)

O contexto onde o agente atua (tudo que está "fora" dele). Pode ser totalmente ou parcialmente observável, determinístico ou estocástico, estático ou dinâmico.

## 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.

## 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.

## Estado (State)

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.

## Exploração vs. Aproveitamento (Exploration vs. Exploitation)

Dilema fundamental em aprendizado por reforço: explorar novas ações para descobrir melhores recompensas ou aproveitar o conhecimento atual para maximizar o retorno conhecido.

## Ferramentas (Tools)

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).

## Memória e Contexto (Memory and Context)

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.

## 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.

## Observações (Observations - 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.

## Orquestração (Orchestration - Planner-Executor, Críticos, Supervisores)

Coordenação de múltiplos agentes ou componentes com papéis especializados: planejadores que definem estratégias, executores que realizam ações, críticos que avaliam resultados e supervisores que mantêm controle geral.

## Persona ou Papel (Role)

Configuração que define estilo, tom, personalidade e características de um agente. (ex.: "tutor paciente", "analista jurídico cauteloso").

## Planejamento (Planning)

Escolha de sequências de ações para alcançar objetivos.

## 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.

## Recompensa (Reward)

Sinal (geralmente numérico) que indica sucesso parcial ou total — comum em aprendizado por reforço. Exemplo: nota ao concluir uma tarefa.

## Sistemas Multiagentes (Multi-Agent Systems - MAS)

Vários agentes cooperando com papéis complementares (ex.: pesquisador, redator, revisor). Exigem protocolos de comunicação e estratégias de coordenação.

# Escopo das IAs

## Inteligência Artificial Geral (Artificial General Intelligence - IAG, 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".

## Inteligência Artificial Restrita (Artificial Narrow Intelligence - IAR, ANI, IA estreita, IA fraca)

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.

## Inteligência Artificial Superior (Artificial Super Intelligence - IAS, 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.

# Ética, Segurança e Governança

## 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.

## Anonimização (Anonymization)

Processo de remover ou alterar informações identificáveis de dados para proteger a identidade dos indivíduos, garantindo a privacidade em tráfego, análises e armazenamento.

## Aprendizado Federado (Federated Learning)

Treinamento de IA descentralizado para preservar a privacidade dos dados (ver anonimização).

## Ataques de Prompt (Prompt Injection)

Técnicas que induzem o modelo a ignorar instruções/limites e gerar saídas indesejadas.

## Conformidade (Compliance)

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.

## 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.

## 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.

## Exemplos Adversariais (Adversarial Examples)

Perturbações sutis em entradas (texto/imagem/áudio) que levam o modelo ao erro. Defesas: adversarial training, regularização e detecção de anomalias.

## Explicabilidade (Explainability - XAI)

Capacidade de um modelo em explicar, de maneira confiável, como chegou a uma determinada decisão.

## Gestão de Incidentes (Incident Management)

Processos estruturados para identificar, responder e resolver problemas em sistemas de IA, incluindo protocolos de escalação, comunicação e recuperação.

## Guard Rails (Guardrails)

Conjuntos de restrições e limites implementados em sistemas de IA para possibilitar operações seguras e éticas.

## Guidelines (Guidelines)

Diretrizes que orientam o desenvolvimento e uso responsável de sistemas de IA.

## IA Ética (Ethical AI)

O mesmo que IA Responsável, mas com um foco maior na ética.

## IA Responsável (Responsible AI)

Conjunto de princípios e práticas para projetar, treinar, avaliar e operar sistemas de IA de forma segura, ética e justa.

## Interpretabilidade (Interpretability)

Área de pesquisa que busca compreender os mecanismos internos de um modelo que geram uma determinada saída ou decisão.

## Jailbreak (Jailbreak)

Técnica que induz o modelo a ignorar restrições e agir livremente, possivelmente de maneira maliciosa ou insegura.

## Marcação de Conteúdo Sintético (Synthetic Content Marking)

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.

## Moderação de Conteúdo (Content Moderation)

Sistemas e processos para identificar, filtrar e remover conteúdo inadequado, nocivo ou que violam políticas de uso.

## Monitoramento Contínuo (Continuous Monitoring)

Acompanhamento sistemático do desempenho e comportamento dos sistemas de IA em produção, incluindo métricas, alertas e detecção de anomalias.

## Observabilidade de IA (AI Observability)

Monitoramento em tempo real sobre o comportamento de sistemas de IA, facilitando a detecção de anomalias.

## Regulação de IA (AI Regulation)

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.

## Responsabilização Algorítmica (Algorithmic Accountability)

Responsabilidade pelos impactos das decisões automatizadas.

## Robustez (Robustness)

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.

## 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.

## 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.

## Viés e Equidade (Bias and Fairness)

Identificação e mitigação de preconceitos nos modelos de IA.

# Habilidades e Práticas

## Auto-consistência (Self-Consistency)

Executar múltiplas amostragens e escolher a resposta mais consistente/consensual. Reduz alucinações e melhora raciocínio em problemas abertos.

## 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.

## 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.

## Decomposição de Tarefas (Task Decomposition)

Quebrar objetivos amplos em etapas mensuráveis (ex.: checklists).

## Design de Guard-Rails (Guard-Rails Design)

Arquitetar políticas, classificadores e checagens (pré/pós) para manter o sistema dentro de limites de segurança, conformidade e tom de marca.

## 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.

## Engenharia de Prompt (Prompt Engineering)

Técnicas para criar prompts eficazes: papéis, objetivos claros, restrições, exemplos, formatos de saída e critérios de qualidade/teste.

## Fusion Skills (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.

## Red Teaming de IA (AI Red Teaming)

Testes ofensivos controlados para descobrir falhas, vieses e caminhos de abuso (ataques de prompt, jailbreak, fuga de dados), com planos de correção.

## Vibe Coding (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.

## Zero-shot e Few-shot (Zero-shot and 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.

# Infraestrutura e Processos

## API (Application Programming Interface)

Interface que permite que programas se comuniquem entre si.

## 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. Impõe restrições de memória/energia; por isso, usa-se quantização, destilação e operadores otimizados (NPUs, DSPs).

## MCP (Model Context Protocol)

Protocolo que permite que agentes de IA se comuniquem entre si e com outros sistemas.

## Orquestração de IA (AI Orchestration)

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.

## Pipeline de Dados (Data Pipeline)

Cadeia de processamento e preparação de dados para alimentar modelos de IA.

## Workflows Inteligentes (Intelligent Workflows)

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.