MANUAL DE USO RESPONSÁVEL DE FERRAMENTAS DE INTELIGÊNCIA ARTIFICIAL GENERATIVA

SUMÁRIO

1. Introdução
2. Âmbito de Aplicação e Abrangência
3. Definições e Conceitos Fundamentais
4. Riscos e Limitações Técnicas
5. Princípios Éticos, Jurídicos e de Governança
6. Regras para o Uso Responsável
7. Supervisão Humana Efetiva, Responsabilidade e Revisão
8. Diretrizes Específicas para Auditoria e Controle Externo
9. Diretrizes Específicas para Contratação e Homologação de Ferramentas de Inteligência Artificial
10. Disposições Finais
11. Glossário

1. INTRODUÇÃO

A rápida evolução das tecnologias digitais, especialmente da Inteligência Artificial Generativa (IAGen), está transformando significativamente a sociedade e a administração pública. Ferramentas como grandes modelos de linguagem (LLMs) têm sido incorporadas para aumentar eficiência, agilidade e qualidade do trabalho, além de impulsionar a inovação nas rotinas institucionais.

No Tribunal de Contas do Município de São Paulo (TCMSP), o uso responsável da IAGen é uma oportunidade estratégica para fortalecer o controle externo, ampliar a análise de grandes volumes de dados, automatizar tarefas repetitivas e aprimorar a elaboração de documentos, contribuindo para a boa gestão dos recursos públicos.

Apesar dos benefícios, o uso dessas tecnologias exige atenção aos riscos e desafios éticos, jurídicos e técnicos, como a proteção de dados, transparência dos processos, integridade das informações, prevenção de vieses e respeito aos direitos fundamentais.

Este manual apresenta diretrizes claras e alinhadas às melhores práticas para o uso responsável da IAGen no TCMSP, incluindo conceitos, princípios éticos, regras, responsabilidades, procedimentos de revisão, precauções e orientações específicas para auditoria e controle externo. O objetivo é garantir que a inovação ocorra com ética, legalidade, segurança e confiança, fortalecendo o controle externo e a gestão pública, sempre em sintonia com os valores institucionais.

2. ÂMBITO DE APLICAÇÃO E ABRANGÊNCIA

Este manual é de aplicação obrigatória aos membros, servidores, estagiários e colaboradores do TCMSP, independentemente do vínculo, cargo, função ou local de trabalho. Sua abrangência inclui todas as unidades e áreas do Tribunal, tanto nas atividades-fim (controle externo, auditoria, fiscalização, instrução processual e julgamento) quanto nas atividades-meio (gestão administrativa, orçamentária, financeira, tecnológica, recursos humanos, comunicação, atendimento ao público e demais funções de suporte).

As diretrizes deste manual aplicam-se ao uso de ferramentas de IAGen desenvolvidas internamente, contratadas, adquiridas de terceiros ou disponibilizadas por outros órgãos, seja por meio de plataformas públicas ou privadas, com modelos abertos ou fechados. O manual cobre o uso dessas soluções em equipamentos institucionais ou pessoais, sempre que utilizados para atividades do TCMSP.

Estão incluídas todas as situações em que a IAGen for utilizada para apoiar, automatizar, sugerir, analisar, redigir, revisar, classificar, resumir, traduzir, organizar ou auxiliar na produção e tratamento de informações, documentos, relatórios, minutas, notas técnicas, pareceres, manifestações, comunicações institucionais e demais produtos relacionados às funções do Tribunal.

A abrangência também inclui o uso de IAGen em processos de auditoria, fiscalização, controle externo, gestão administrativa, suporte técnico, elaboração de peças processuais, produção de

conteúdo institucional e atendimento de demandas, inclusive em casos de integração com outros sistemas, uso combinado com outras tecnologias ou aplicação em fluxos de trabalho automatizados.

Este manual se aplica ainda à participação em projetos-piloto, iniciativas experimentais, treinamentos, ações de inovação e projetos de pesquisa e desenvolvimento relacionados ao uso de IAGen no TCMSP, mesmo em parcerias com outras instituições, públicas ou privadas.

O cumprimento das normas aqui estabelecidas é obrigatório, cabendo a cada usuário, gestor ou unidade garantir sua observância e comunicar ao Núcleo de Tecnologia da Informação (NTI) qualquer dúvida, omissão ou necessidade de adaptação no uso da IA Generativa nas rotinas institucionais.

3. DEFINIÇÕES E CONCEITOS FUNDAMENTAIS

Para garantir o uso responsável da IAGen no TCMSP, é essencial que membros, servidores, colaboradores e parceiros conheçam os conceitos fundamentais. Esta seção apresenta os principais termos, facilitando a compreensão das diretrizes, limitações e potencialidades das ferramentas adotadas.

Inteligência Artificial Generativa (IAGen) são sistemas capazes de criar, modificar ou sintetizar conteúdos originais – a exemplo de textos, imagens, áudios, vídeos e códigos – a partir de comandos fornecidos por pessoas. Diferentemente de soluções tradicionais baseadas em regras fixas, a IAGen utiliza aprendizado de máquina e análise de grandes volumes de dados para identificar padrões, interpretar linguagem natural e gerar respostas contextualizadas e autônomas.

Destacam-se os **Modelos de Linguagem de Grande Escala (Large Language Models – LLMs),** como ChatGPT, Gemini e Claude. Esses modelos, treinados com grandes conjuntos de textos, geram respostas complexas e podem ser classificados como abertos (código-fonte auditável) ou fechados (proprietários, com menor transparência).

Prompt é a instrução detalhada dada pelo usuário à IAGen, indicando o que deve ser feito e como executar a tarefa. Resultados adequados dependem da clareza, do contexto e da definição precisa dos critérios de sucesso, objetivo, formato de entrega e fontes autorizadas.

Exemplo: "Redija ofício administrativo (1 página) para o Secretário X, com introdução, fundamento e pedido; baseie-se apenas no Relatório Y e nas normas Z; entrega em texto corrido."

Exemplo: "Resuma o Relatório de Auditoria 2024 em 300 palavras, apresentando achados, riscos e recomendações; não utilize fontes externas; inclua tabela com prazos e responsáveis."

Exemplo: "Elabore checklist para análise de contratos com IA: finalidade, dados tratados, segurança, critérios de explicabilidade e plano de contingência."

Exemplo: "Gere script em Python para ler CSV (colunas A,B,C), validar nulos, calcular média e desvio-padrão e criar gráfico de barras; explique o método adotado."

É fundamental atentar para as chamadas **alucinações**, quando a IA gera respostas aparentemente corretas, mas incorretas, incompletas ou fictícias. Essas limitações exigem revisão humana criteriosa e validação prévia das informações.

O **viés algorítmico** é outro risco relevante, podendo surgir de dados de treinamento que refletem desigualdades ou estereótipos, resultando em respostas distorcidas ou discriminatórias. A **mitigação de vieses** é pilar da governança ética da IA.

Outros conceitos essenciais incluem: **explicabilidade** (compreensão de como a IA chegou ao resultado), **auditabilidade** (rastreabilidade e revisão das decisões), **contestabilidade** (direito de questionar decisões baseadas em IA) e **proteção de dados** (cumprimento rigoroso da legislação de privacidade e segurança).

Para consulta de termos técnicos, este manual dispõe de glossário detalhado ao final, que deve ser utilizado em caso de dúvidas quanto à terminologia ou conceitos abordados.

4. RISCOS E LIMITAÇÕES TÉCNICAS

Embora a IAGen ofereça múltiplos benefícios ao TCMSP, é essencial reconhecer e gerenciar seus riscos e limitações de forma consciente. Usuários e gestores devem manter atenção constante às fragilidades que podem afetar a qualidade, confiabilidade e segurança das informações processadas por essas ferramentas.

Além dos riscos relevantes trazidos pelas “alucinações” e vieses algorítmicos, as limitações técnicas também devem ser consideradas. Os modelos de IAGen, apesar de avançados, não compreendem o contexto institucional nem possuem discernimento jurídico ou administrativo, devendo ser usados como apoio, nunca como substitutos do julgamento humano.

É fundamental evitar automação excessiva ou dependência acrítica das respostas da IA. A adoção precipitada de conteúdos gerados sem análise e validação fragiliza decisões e relatórios institucionais.

Portanto, reconhecer e agir sobre esses riscos é indispensável para o uso ético, transparente, seguro e alinhado ao interesse público das ferramentas de IAGen no Tribunal.

5. PRINCÍPIOS ÉTICOS, JURÍDICOS E DE GOVERNANÇA

O uso da Inteligência Artificial Generativa (IAGen) no TCMSP deve seguir princípios éticos, jurídicos e de governança, garantindo adoção responsável, segura e conforme a legislação brasileira e a missão institucional.

O princípio da **legalidade** exige estrita observância às normas vigentes, como Constituição Federal, Lei de Acesso à Informação - LAI, Lei Geral de Proteção de Dados - LGPD, Marco Civil da Internet, Lei de Governo Digital, Lei de Licitações e demais regulamentações aplicáveis. O respeito ao arcabouço legal legitima o uso da IAGen no setor público.

A **moralidade** determina padrões elevados de ética e integridade, com transparência, justiça e combate à corrupção, favorecimento ou discriminação.

A **impessoalidade** assegura que as interações da IAGen sejam neutras, sem influência de interesses pessoais, políticos ou econômicos, garantindo tratamento igualitário e credibilidade institucional.

A **publicidade** exige transparência sobre o uso da IAGen, tornando claro para todos quando há participação de ferramentas automatizadas, conforme os princípios de *accountability* e controle social.

A **eficiência** orienta a utilização da IAGen para aprimorar recursos públicos e resultados, sem jamais prescindir da supervisão humana. A tecnologia auxilia, mas não substitui o julgamento humano dos agentes públicos.

A **revisão e a supervisão humanas** são obrigatórias: todo conteúdo gerado por IAGen deve ser validado por servidor responsável, garantindo precisão e adequação.

O princípio da **responsabilidade** estabelece que o usuário responde integralmente pelo uso e resultados da IAGen, vedando a transferência de culpa à ferramenta.

A **segurança da informação e proteção de dados** devem ser rigorosamente observadas, conforme políticas internas e legislação vigente.

A **governança ativa** é essencial para a sustentabilidade da IAGen, recomendando-se criar ou fortalecer instâncias para monitoramento, revisão e aprimoramento contínuo.

Esses princípios reforçam o compromisso do Tribunal com ética, transparência, responsabilidade, legalidade e inovação, assegurando o uso da IAGen em prol do interesse público e dos direitos fundamentais.

6. REGRAS PARA O USO RESPONSÁVEL

Quanto às regras para uso responsável, é obrigatório observar este manual, resoluções internas e legislação aplicável, visando segurança, ética, legalidade e efetividade institucional.

Todo conteúdo gerado com auxílio de IAGen, mesmo parcial, deve ser revisado e validado por servidor responsável, especialmente em atos oficiais e decisões administrativas, pois a responsabilidade é sempre do agente público.

É proibido inserir ou compartilhar informações pessoais, sensíveis, sigilosas ou protegidas por segredo de justiça em plataformas públicas, abertas ou não autorizadas; priorize soluções internas homologadas e ambientes controlados.

A revisão crítica do conteúdo gerado por IAGen é responsabilidade direta do usuário, que deve analisar exatidão, pertinência, clareza e conformidade, avaliando riscos de erros, omissões, alucinações ou vieses, realizando pesquisas complementares se necessário.

Salvo justificativa técnica formal, é desejável manter registros das principais interações, decisões e produções com IAGen para rastreabilidade, auditoria e prestação de contas, conforme políticas internas e legislação (Marco Civil da Internet).

Deve ser garantida transparência ativa sobre o uso institucional de IA, com divulgação pública conforme regulamentação.

É vedada a geração ou difusão de *deepfakes* ou simulações indevidas de identidade, devendo adotar controles técnicos de autenticação e marca d'água, com guarda dos registros seguindo normas de segurança e proteção de dados.

É vedado o uso de quaisquer sistemas, modelos, ferramentas ou recursos de inteligência artificial — internos ou externos — para a prática de atos ilícitos, antiéticos, abusivos ou que violem direitos fundamentais, incluídas, mas não limitadas a: manipulação, distorção, criação, adulteração ou divulgação de conteúdos sintéticos (imagens, vídeos, áudios, textos ou dados) capazes de causar danos a terceiros; exposição não consentida de pessoas; reprodução de conteúdo discriminatório; assédio; violação de privacidade; ou qualquer conduta que contrarie a legislação vigente, as normas internas deste Tribunal, os princípios da Administração Pública e as políticas de segurança da informação e proteção de dados.

O uso responsável da IAGen exige atualização contínua dos servidores, participação em treinamentos e acompanhamento das melhores práticas nacionais e internacionais. O compromisso com ética, transparência, diligência e inovação é indispensável para um ambiente institucional seguro, produtivo e alinhado aos princípios do serviço público.

7. SUPERVISÃO HUMANA EFETIVA, RESPONSABILIDADE E REVISÃO

O uso de Inteligência Artificial Generativa (IAGen) no TCMSP não exime servidores, colaboradores ou agentes institucionais de sua responsabilidade funcional e ética sobre atos e conteúdos produzidos. Ao contrário, a adoção de IA exige controle, acompanhamento e validação humana rigorosa em todas as etapas do trabalho, tornando indispensável a atuação criteriosa dos responsáveis.

Todo produto, documento, análise, parecer ou comunicação gerado com apoio de IAGen deve ser obrigatoriamente revisado, validado e chancelado por agente público formalmente competente, com análise minuciosa de exatidão, pertinência, clareza, aderência normativa e alinhamento aos objetivos institucionais.

É dever do servidor validar a procedência das informações, corrigir eventuais equívocos, omissões, erros factuais, vieses ou inconsistências identificadas nas respostas da IAGen, realizando ajustes, complementações e pesquisas em fontes confiáveis sempre que necessário. É vedada a tramitação ou uso de conteúdo gerado por IA sem conferência e validação funcional do agente humano.

A responsabilidade pelo uso da IAGen e pelos resultados é exclusiva do servidor, colaborador ou gestor que utilizou a tecnologia. Não se admite alegar falha, limitação ou erro da ferramenta para justificar descumprimento de deveres, infrações disciplinares ou prejuízos.

A revisão, validação e explicação do uso de IAGen são obrigações pessoais e intransferíveis do responsável, que deve estar apto a justificar decisões e procedimentos perante controles internos e externos, demonstrando transparência e fundamentação técnica.

A cultura de responsabilidade e supervisão humana é essencial para que a IAGen seja instrumento de aprimoramento institucional, evitando riscos e vulnerabilidades nos processos do TCMSP. Cabe a todos os usuários adotar postura ética, criteriosa e alinhada às melhores práticas do serviço público, zelando pela confiança social e pelos valores institucionais do Tribunal.

8. DIRETRIZES ESPECÍFICAS PARA AUDITORIA E CONTROLE EXTERNO

No âmbito da auditoria e do controle externo, o uso de ferramentas de IAGen deve ser regido pelos princípios de responsabilidade, segurança, eficiência e transparência, com foco no suporte às atividades finalísticas do TCMSP, sem substituir a análise e a decisão dos auditores.

A IAGen pode ser utilizada para organizar, triar e analisar grandes volumes de dados e documentos, bem como para elaborar minutas, resumos e estruturas de relatórios, além de identificar padrões e indícios relevantes que otimizem a fiscalização e o uso de recursos.

Todos os conteúdos, relatórios ou recomendações geradas com apoio de IAGen devem ser obrigatoriamente revisadas e validadas por auditor ou equipe técnica competente. A decisão final sobre achados e conclusões permanece exclusiva do agente humano, sendo vedada a delegação da análise de mérito ou emissão de manifestações oficiais à ferramenta tecnológica.

Na manipulação de dados sensíveis, sigilosos ou estratégicos, especialmente em processos em andamento ou investigações reservadas, o uso de IAGen exige cautela redobrada, sendo obrigatório empregar apenas soluções internas homologadas, ambientes seguros e respeito às normas institucionais de segurança da informação. É proibido inserir dados protegidos em ferramentas externas, públicas ou não autorizadas.

A capacitação constante das equipes e a atualização dos protocolos de uso de IAGen são essenciais. O compartilhamento de boas práticas entre setores será incentivado para fortalecer a governança, a inovação responsável e a excelência no controle externo.

Essas diretrizes asseguram que a IAGen seja um instrumento estratégico para aprimorar a auditoria, ampliar a capacidade analítica e promover a efetividade das fiscalizações, sempre com responsabilidade técnica e segurança nas decisões.

9. DIRETRIZES PARA A CONTRATAÇÃO, ADESÃO, UTILIZAÇÃO E HOMOLOGAÇÃO DE FERRAMENTAS DE INTELIGÊNCIA ARTIFICIAL

A aquisição, contratação, adesão ou utilização de qualquer ferramenta de IAGen no âmbito do TCMSP, seja ela desenvolvida internamente, fornecida por terceiros ou de acesso público para fins institucionais, deverá observar o Sistema de Gestão de IA do TCMSP e a governança definida pela Resolução nº 08/2026, incluindo requisitos de aquisição, segurança, compatibilidade e revogação contratual por descumprimento (art. 3º e art. 8º).

9.1. Conformidade Legal e Normativa

As ferramentas de IAGen devem atender integralmente, pelo menos, às seguintes normas constitucionais, legais e institucionais:

- **CF/88, arts. 5º, X, LIV, LV; 37, caput**, observando aos princípios de publicidade, eficiência, devido processo e revisão humana;

- **Lei nº 12.527/2011** (**Lei de Acesso à Informação - LAI)** e **Resolução TCMSP nº 29/2019**, controlando o acesso e divulgação de informações sigilosas;

- **Lei nº 12.965/2014 (Marco Civil da Internet)**, no que tange à guarda e acesso a registros, e cadeia de custódia;

- **Lei nº 13.709/2018 (Lei Geral de Proteção de Dados - LGPD)** e **Resolução TCMSP nº 10/2025**, incluindo cláusulas específicas sobre finalidade, base legal, responsabilidades entre controlador e operador, medidas de segurança e reporte de incidentes;

- **Lei nº 14.129/2021 (Governo Digital)**, no que se refere à governança, gestão de riscos e transparência tecnológica;

- **Lei nº 14.133/2021 (Licitações)**, no que dispõe sobre o planejamento, gestão/fiscalização de riscos, e cláusulas técnicas contratuais.

- **Política de Segurança da Informação** (**Resolução TCMSP nº 01/2020** e **Instrução TCMSP nº 01/2020**), preservando confidencialidade, integridade, disponibilidade, autenticidade e não-repúdio.

9.2. Segurança da Informação

Os contratos devem estabelecer:

- cláusulas expressas de confidencialidade para informações institucionais e processuais;
- garantias técnicas e administrativas contra acesso não autorizado e vazamento de dados;

- responsabilização do fornecedor por descumprimento das medidas de segurança.
- controles contra vazamento de dados, jailbreaks e outputs nocivos, bem como, quando cabível, a realização de testes adversariais (*red teaming*) proporcionais ao risco e a observância do Protocolo de Incidentes de IA.

9.3. Responsabilidade do Fornecedor

O contrato deve definir inequivocamente a responsabilidade civil e administrativa do fornecedor por:

- Falhas sistêmicas e interrupções do serviço;
- Geração de conteúdo ilícito ou prejudicial por culpa ou dolo da plataforma;
- Notificação imediata de incidentes de segurança;
- Cooperação plena para mitigação de danos.

9.4. Transparência e Auditabilidade

Na escolha entre ferramentas equivalentes, será priorizada aquela que ofereça:

- Maior transparência sobre modelos de treinamento e funcionamento;

- Mecanismos de rastreabilidade e auditoria das interações e resultados (quando tecnicamente possível);

- Alinhamento com os princípios de explicabilidade e accountability.

9.5. Ferramentas de Acesso Público Gratuito

Para uso institucional, estas ferramentas devem passar, no mínimo, por análise de risco simplificada, com foco especial na **vedação absoluta** de inserção de dados sensíveis, sigilosos ou pessoais, e sempre observar às regras estabelecidas na Resolução nº 08/2026, nas diretrizes e determinações do CG-IA, bem como neste manual.

10. DISPOSIÇÕES FINAIS

Este manual deve ser interpretado e aplicado como um instrumento norteador para a utilização ética, responsável e eficiente das ferramentas de IAGen no âmbito deste TCMSP, complementando as demais normas internas, políticas de segurança da informação, proteção de dados e governança digital vigentes na instituição. É responsabilidade de todos os membros, servidores, estagiários e colaboradores, independentemente de suas funções, assegurar a observância das diretrizes estabelecidas neste documento, comunicando eventuais dúvidas, dificuldades ou situações não previstas à área competente pela gestão de tecnologia da informação do Tribunal.

Diante do caráter dinâmico da evolução das tecnologias de Inteligência Artificial e do surgimento de novos riscos, desafios, oportunidades e práticas na administração pública, este manual poderá ser revisado, atualizado e aprimorado periodicamente, mediante iniciativa do

NTI ou do CG-IA, com a participação, quando pertinente, dos setores diretamente envolvidos e de especialistas na matéria, cabendo a sua aprovação final à Alta Administração desta Corte.

Compete ao Núcleo de Tecnologia da Informação do Tribunal, em parceria com as demais áreas de apoio institucional, disponibilizar canais de comunicação destinados ao esclarecimento de dúvidas, recebimento de sugestões, reporte de incidentes e acompanhamento de eventuais dificuldades relacionadas à aplicação das normas aqui dispostas.

Por fim, ressalta-se que o uso da IAGen deve estar sempre orientado para o aprimoramento do controle externo, a proteção dos direitos fundamentais, o respeito à legalidade e a busca contínua pela excelência na gestão dos recursos públicos, contribuindo para o fortalecimento de um Tribunal de Contas mais moderno, confiável e alinhado às melhores práticas de governança e fiscalização.

11. GLOSSÁRIO

Este glossário apresenta, em ordem alfabética, definições dos principais termos técnicos e conceituais utilizados neste manual, com o objetivo de promover a compreensão clara e uniforme das orientações sobre o uso de Inteligência Artificial Generativa no âmbito do TCMSP. Recomenda-se a consulta frequente a este glossário, especialmente diante de dúvidas quanto ao significado ou ao alcance dos termos empregados.

A/B *Testing* (teste A/B): Técnica de experimentação controlada para comparar duas versões (A vs. B) medindo impacto em métricas definidas.

ABAC / RBAC (controle de acesso por atributos / por papéis): Modelos de controle de acesso. ABAC usa atributos (usuário, recurso, contexto); RBAC usa papéis/perfis institucionais.

Alinhamento (*AI Alignment*): Grau de aderência do comportamento do modelo a objetivos, princípios éticos e restrições definidas pela instituição.

Alucinação (Incorreções Factuais): Fenômeno pelo qual a IA gera respostas plausíveis, porém factualmente incorretas, fictícias ou sem lastro verificável; requer validação humana rigorosa.

Auditabilidade: Capacidade de rastrear, revisar e avaliar decisões/conteúdos gerados pela IA, garantindo transparência, responsabilidade e conformidade normativa.

Avaliação de Impacto Algorítmico (AIA): Análise estruturada dos riscos e efeitos de um sistema de IA sobre direitos, processos e resultados, com medidas de mitigação e monitoramento.

Base de Conhecimento (KB): Conjunto curado de dados/documentos institucionais usado como fonte preferencial para respostas, explicações e evidências.

***Benchmarks* (métricas/coleções de teste):** Conjuntos e protocolos padronizados para avaliar modelos (ex.: precisão, F1, ROUGE, BLEU, toxicidade).

Cadeia de Custódia (de dados e *logs*): Trilha íntegra da origem, transformação e acesso a evidências (dados, prompts, saídas), assegurando autenticidade e auditoria.

***Canary Release* / *Rollback*:** Implantação gradual para pequena fração de usuários com capacidade de reversão rápida em caso de falhas ou regressões.

Conjunto de Dados: Ver ***Dataset*.**

Conteúdo Sintético: Texto, imagem, áudio, vídeo ou código gerado por IA (não capturado diretamente do mundo real).

Contestabilidade: Direito de questionar, revisar ou contestar resultados/decisões assistidos por IA, especialmente quando impactarem direitos e interesses legítimos.

***Context Window* (Janela de Contexto):** Quantidade máxima de tokens processada por interação (soma de instruções, histórico e documentos).

***Data Drift* / *Concept Drift* (Deriva de dados/conceito):** Mudança nos dados de entrada ou na relação entre variáveis e rótulos ao longo do tempo, degradando o desempenho do modelo.

***Data Governance* (Governança de Dados):** Políticas, papéis e processos para qualidade, segurança, ciclo de vida e uso responsável de dados.

***Data Leakage* (Vazamento de dados):** Uso indevido de informação do conjunto de teste/produção no treinamento, gerando avaliações artificialmente altas.

***Dataset* (Conjunto de Dados):** Coleção organizada de exemplos usada para treinar, validar e testar modelos.

Deepfake: termo da área de tecnologia que se refere à mídia (vídeos, áudios ou imagens) manipulada ou sintética que apresenta a aparência e/ou a voz de uma pessoa de forma extremamente realista, fazendo parecer que ela disse ou fez algo que, na verdade, nunca fez.

Deriva de dados/conceito: Ver ***Data Drift* / *Concept Drift*.**

DPIA / RIPD (Relatório de Impacto à Proteção de Dados): Avaliação, nos termos da LGPD, sobre riscos e salvaguardas no tratamento de dados pessoais.

Engenharia de *Prompt* (*Prompt Engineering*): Práticas para projetar prompts eficazes (objetivo, contexto, dados permitidos, critérios de aceitação, formato e restrições).

Explicabilidade: Capacidade de compreender e descrever de modo transparente como a IA chegou a determinado resultado.

***Few-shot* / *Zero-shot*:** Execução de tarefa com poucos exemplos (few-shot) ou sem exemplos (zero-shot) incluídos no prompt.

Filtro de Segurança (*Safety Filter*): Mecanismos automáticos para bloquear conteúdo ilícito/sensível (ex.: PII, discurso de ódio, malware).

***Fine-tuning* (ajuste fino):** Treinamento adicional de modelo pré-treinado para domínio/tarefa específica com dados curados.

Governança de Dados: Ver ***Data Governance*.**

***Guardrails* (Trilhos de Segurança):** Restrições técnicas e de negócio aplicadas às interações (validações, políticas, formatos obrigatórios).

***Hallucination Control* (controle de alucinação):** Conjunto de técnicas (RAG, verificadores, citações obrigatórias) para reduzir respostas sem lastro.

HITL – *Human in the Loop*: Ver **Supervisão Humana efetiva.**

IAGen (Inteligência Artificial Generativa): Sistemas capazes de criar/modificar conteúdos originais (texto, imagem, áudio, vídeo, código) a partir de comandos; usam aprendizado de máquina para identificar padrões, entender linguagem natural e gerar respostas contextualizadas.

Incorreções Factuais: Ver **Alucinação.**

Inferência (*Inference*): Etapa de execução do modelo para gerar saída a partir de um prompt; envolve latência e custo por token.

Inteligência Artificial Generativa (IAGen): Ver **IAGen.**

Inventário de Sistemas de IA do TCMSP: cadastro obrigatório e contínuo de soluções de IA, com classificação por impacto (baixo, moderado, alto) e matriz de impacto baseada em escala, gravidade, reversibilidade, abrangência e sensibilidade dos dados.

Janela de Contexto: Ver ***Context Window*.**

***Jailbreak* / *Prompt* de Quebra:** Tentativas de contornar políticas do modelo por meio de instruções manipuladas.

Latência / *Throughput*: Tempo de resposta do sistema e volume de requisições processado por unidade de tempo.

Licenciamento de Dados / Direitos Autorais: Regras de permissão de uso, redistribuição e derivação de dados/modelos; observar legislação autoral aplicável.

LIME / SHAP (técnicas de explicabilidade): Métodos para estimar a contribuição de variáveis/trechos de entrada no resultado do modelo.

LLMs (Modelos de Linguagem de Grande Escala): Plataformas treinadas com grandes volumes de dados textuais para compreender e gerar linguagem natural de forma coesa; ex.: GPT, Gemini.

***Logging* (registro) / Telemetria:** Coleta estruturada de eventos (*prompts*, parâmetros, saídas, erros) para observabilidade, auditoria e melhoria contínua.

Memória (contextual/conversacional): Informações persistidas para manter contexto entre interações, conforme políticas de retenção e privacidade.

***Model Card* (cartão do modelo):** Documento com finalidade, dados, métricas, limitações, riscos e usos aceitáveis do modelo.

Modelos Abertos: Modelos cujo código-fonte é acessível/auditável e pode ser adaptado por terceiros, favorecendo transparência e controle.

Modelos de Linguagem de Grande Escala: Ver **LLMs.**

Modelos Fechados: Modelos proprietários com código-fonte restrito, dificultando auditoria e controle direto do usuário final.

Observabilidade de Modelos: Monitoramento contínuo de métricas técnicas e de risco (qualidade, vieses, segurança, custo, disponibilidade).

Parâmetros de Decodificação (*temperature*, *top-k*, *top-p*): Controles de aleatoriedade/diversidade na geração de texto.

Política de Retenção de Dados: Períodos e critérios para guarda e descarte seguro de dados, prompts, logs e conteúdos.

***Prompt*:** Instrução detalhada fornecida pelo usuário à ferramenta de IAGen que descreve o que deve ser feito e como executar a tarefa. Deve explicitar objetivo, contexto, critérios de aceitação, formato de entrega e dados/fontes autorizados. A qualidade e adequação da resposta dependem diretamente da clareza, contextualização e objetividade do prompt.

***Prompt* de sistema (*System Prompt*):** Instruções internas fixas que configuram o comportamento do modelo numa aplicação.

Proveniência (*provenance*): Rastreio verificável da origem de conteúdo (dados, modelo, versão, transformações), base para confiança e auditoria.

RAG – Recuperação Aumentada por Geração: Arquitetura em que documentos confiáveis são recuperados e injetados no prompt para fundamentar a resposta.

***Red Teaming* (testes adversariais):** Avaliações proativas para descobrir falhas, vieses e vetores de abuso em condições realistas.

Risco (mapa e apetite de risco): Probabilidade e impacto de eventos adversos; diretrizes definem níveis aceitáveis e respostas (evitar, mitigar, transferir, aceitar).

***Sandbox* (ambiente controlado):** Ambiente isolado para testes/pilotos com dados sintéticos/anonimizados, limites de escopo e monitoramento.

SLA / SLO (acordo/objetivo de nível de serviço): Compromissos de disponibilidade, latência, qualidade e suporte.

Supervisão Humana efetiva (HITL): Participação obrigatória de usuários na revisão/validação das saídas da IA, assegurando controle, responsabilidade e aderência a princípios éticos e legais.

***System Prompt*:** Ver ***Prompt* de sistema.**

***Temperature*:** Ver **Parâmetros de Decodificação.**

***Token* / *Tokenização*:** Unidades mínimas de processamento (pedaços de palavras); tokenização é a conversão do texto para esses *tokens*.

TCO – Custo Total de Propriedade: Custo completo do ciclo de vida (licenças, infraestrutura, dados, operação, monitoramento, segurança, suporte).

Transparência Ativa (em IA): Divulgação clara do uso de IA, limites, fontes, grau de revisão humana e canais de contestação.

Vazamento de dados: Ver ***Data Leakage*.**

Versionamento de Modelos / Artefatos: Controle de versões de modelos, dados, prompts e pipelines, com trilhas de auditoria e reprodutibilidade.

Viés Algorítmico: Tendência de modelos a reproduzir/reforçar distorções ou preconceitos dos dados de treinamento, gerando respostas parciais/inadequadas.

***Watermarking* / Detecção de IA:** Marcação ou sinais estatísticos para identificar conteúdo gerado por IA, usados para transparência e mitigação de abuso.