

Framework de Governança de IA Generativa (GoLLM)



Visão Geral do Sistema: Framework GoLLM

O Framework de Governança de IA Generativa (GoLLM) da Lanlink Informática é uma plataforma empresarial abrangente que oferece controle centralizado, democratização responsável e transparência organizacional para ativos de IA Generativa. A solução opera como um sistema de governança corporativa que permite às organizações gerenciar modelos de LLM, prompts, deployments e endpoints de forma segura, auditável e eficiente.

Objetivos Principais:





Lanlink GenOps

Acesse a Plataforma de Governança de IA Generativa

Email

 seu.email@lanlink.com.br

Demonstração: use admin@example.com / password123

Senha

[Esqueceu a senha?](#)



☐ Lembrar de mim

 Entrar

OU

 Entrar com Microsoft

Não possui uma conta? [Registre-se](#)

© 2024 Lanlink. Todos os direitos reservados.

- Painel
- Domínios
- Coleções
- Modelos
- Implantações
- Prompts
- Aprovações 4
- Governança
- Análises
- Painel Geral
- Análise Geográfica
- Documentação
- Configurações

Painel

Bem-vindo(a), Antonio Guilherme Cunha Santos!



Total de Prompts
1.256



Modelos Ativos
5



Índice de Saúde
92%



Usuários Ativos
48



Domínios de Governança

Ver todos →



Plataforma de Governança
4 coleções

Saúde 98%



Suporte ao Cliente
12 coleções 3 ações pendentes

Saúde 76%



Marketing
8 coleções 7 ações pendentes

Saúde 54%



Desenvolvimento de Produto
16 coleções 1 ações pendentes

Saúde 92%



Ações de Saúde Prioritárias

Ver todas →

Auditar configurações de segurança da coleção de prompts

Pendente

Impacto: Alto Domínio: Suporte ao Cliente

Prazo: 15/05/2024

Revisar controles de validação de campanhas de marketing

Em Andamento

Impacto: Médio Domínio: Marketing

Prazo: 10/05/2024

Habilitar recursos de detecção de viés de dados

Pendente

Impacto: Alto Todos os domínios

Prazo: 20/05/2024

Adicionar filtros de conteúdo específicos por domínio

Concluído

Impacto: Médio Domínio: Marketing

Prazo: 30/04/2024

Prompts Recentes

Ver todos

Prateleira Corporativa de Modelos de LLM

O Módulo de Modelos tem como objetivo criar uma prateleira corporativa centralizada de modelos LLM validados e aprovados, promovendo transparência e democratização responsável dos recursos de IA.

Catálogo de Modelos

- **Listagem Unificada:** Visualização de todos os modelos disponíveis.
- **Provedores Suportados:** Azure OpenAI, Anthropic, Bedrock, OpenAI API, Google Vertex AI, modelos internos (OpenAI-Compatible Endpoints)
- **Metadados Completos:** Informações técnicas, custos, documentação e certificações.

Aprovação e Governança

- **Fluxos de Aprovação Configuráveis:** Processos customizáveis de aprovação dos modelos por tipo e risco.
- **Status de Aprovação:** Pendente, Aprovado, Rejeitado com rastreabilidade.
- **Revisão Técnica:** Validação de capacidades e especificações técnicas.
- **Revisão de Segurança:** Avaliação de implicações de segurança.
- **Aprovação Executiva:** Processo final para modelos de alto risco.

Gestão de Lifecycle

- **Status de Modelos:** Ativo, Inativo, Depreciado.
- **Versionamento:** Controle de versões com histórico detalhado.
- **Avaliação de Capacidade:** Baixa, Média, Alta baseada em performance.
- **Níveis de Risco:** Baixo, Médio, Alto, Crítico com avaliação contínua.

Relacionamento com Collections

- **Vinculação Flexível:** Modelos podem pertencer a múltiplas collections (departamentos ou line-of-business).
- **Collection Principal:** Definição de collection padrão por modelo.
- **Herança de Permissões:** Controle de acesso baseado na hierarquia organizacional.

- Painel
- Domínios
- Coleções
- Modelos
- Implantações
- Prompts
- Aprovações
- Governança
- Análises
- Painel Geral
- Análise Geográfica
- Documentação
- Configurações

Modelos

Gerenciar e monitorar modelos de IA Generativa

Sincronizar Modelos

Configurar Provedores

+ Registrar Modelo

Conexões com Provedores



Azure OpenAI

Ativo • Última sincronização: 3/27/2024



AWS Bedrock

Não Conectado



Google Vertex AI

Ativo • Última sincronização: 3/24/2024



Anthropic

Ativo • Última sincronização: 3/26/2024



OpenAI API

Ativo • Última sincronização: 3/27/2024



Hugging Face

Não Conectado



Modelos Internos

Auto-Hospedado • Última sincronização: 3/27/2024

Buscar modelos...

Todos Provedores

Todos Tipos

Todos Status



GPT-4 Turbo

Azure OpenAI • 1106

Alta

Tipo	Texto Generativo
Janela de Contexto	128.0K tokens
Status	Ativo
Nível de Risco	Médio
Implantações	3

produção voltado-ao-cliente empresarial

8 coleções 12.5K chamadas Aprovado



Claude 3 Opus

Anthropic • 1.0

Muito Alta

Tipo	Texto Generativo
Janela de Contexto	200.0K tokens
Status	Ativo
Nível de Risco	Médio
Implantações	1

pesquisa interno

3 coleções 5.8K chamadas Aprovado



GPT-3.5 Turbo

Azure OpenAI • 0613

Média

Tipo	Texto Generativo
Janela de Contexto	16.0K tokens
Status	Ativo
Nível de Risco	Baixo
Implantações	5

produção alto-volume voltado-ao-cliente

15 coleções 45.0K chamadas Aprovado

Gerenciamento de Endpoints

O Módulo de Deployments fornece endpoints baseados em modelos com controle rigoroso de custos, tracking de uso detalhado, políticas de expiração e capacidades avançadas de failover e balanceamento de carga.

Gestão de Endpoints

- **Controle:** Evita a exposição do endpoint e chave real do provedor de LLM
- **Criação de Deployments:** Interface simplificada.
- **URLs Customizadas:** Endpoints personalizados com autenticação por chave.
- **Status Monitoring:** Monitoramento em near-real-time

Controle de Custos

- **Limites Configuráveis:** Quotas por chamadas, tokens e custo.
- **Orçamento Mensal:** Controle de gastos com alertas por e-mail.
- **Relatórios Detalhados:** Histórico de consumo.

Políticas de Expiração

- **Data de Expiração:** Ciclo de vida dos deployments.
- **Renovação Automática:** Opção de renovação.
- **Alertas de Expiração:** Notificações proativas na própria ferramenta.
- **Arquivamento Controlado:** Processo de desativação.

Controle de Acesso e CORS

- **Lista de IPs Permitidos:** Restrição de acesso.
- **Rate Limiting:** Limitação de requisições.
- **Autenticação Token:** Tokens únicos por deployment.

Sistema de Failover

- **Múltiplos Modelos em Paralelo:** Modelos backup, permitindo que o framework faça o roteamento da chamada para um segundo endpoint em caso de falhas.

Painel

Domínios

Coleções

Modelos

Implantações

Prompts

Aprovações 4

Governança

Análises

Painel Geral

Análise Geográfica

Documentação

Configurações

[← Voltar para Deployments](#) **Novo Deployment**

Salvar Rascunho

Enviar para Aprovação

Informações Básicas

Conexões

Designer de Workflow

Governança & Aprovações

Integração

Endpoints, chaves API e configurações específicas de provedores serão definidos no Designer de Workflow. Após a criação, você poderá gerenciar ambientes (dev, test, prod) e rotação de chaves na página de detalhes.

Designer de Workflow de Modelos

Configure o Fluxo de Modelos

Use o designer visual para criar um workflow inteligente conectando múltiplos modelos. Defina prioridades, fallbacks, balanceamento de carga e regras condicionais.

Biblioteca de Modelos

Estratégia de Roteamento

Sequencial (Fallback)

Tenta modelos em ordem de prioridade

GPT-4o

OpenAI
\$0.005/1K tokens

GPT-4o Mini

GPT-4o

OpenAI
Latência: 200ms
Erro: 1.0%
Prioridade: 1

Fallback

Configurar GPT-4o

API

Limites

Roteamento

Endpoint

API Key

GPT-4o

OpenAI

Latência:

Erro: 1.0%

Prioridade:

Sistema de Gerenciamento de Chaves/Keys

O Sistema de Gerenciamento de Chaves/Keys visa fornecer controle granular e seguro para chaves de API, com políticas de expiração, rotação automática e auditoria completa, sem necessitar do compartilhamento dos dados reais dos endpoints dos modelos de LLM.

Gestão de Chaves de API

- **Tipos de Chave:** Primária, Secundária para redundância.
- **Ambientes:** Desenvolvimento, Testes, Produção.
- **Geração Segura:** Criação de chaves com prefixos identificadores.
- **Mascaramento:** Exibição segura com ocultação de valores sensíveis.

Controles de Segurança

- **Limitação por IP:** Restrição de uso por endereços específicos.
- **Operações Permitidas:** Controle granular de funcionalidades (Ex: completion, embedding).
- **Limites de Tokens:** Restrição de tokens por requisição.
- **Modelos Permitidos:** Definição de modelos acessíveis por chave.

Políticas de Expiração

- **Configuração Flexível:** 30, 90, 180, 365 dias ou sem expiração.
- **Alertas Automáticos:** Notificações antes do vencimento.
- **Renovação Assistida:** Processo guiado para renovação.
- **Histórico de Rotações:** Rastreamento completo de ciclo de vida.

Monitoramento e Auditoria

- **Tracking de Uso:** Monitoramento detalhado de chamadas e custos.
- **Logs de Auditoria:** Registro completo de criação e revogação.

Painel

Domínios

Coleções

Modelos

Implantações

Prompts

Aprovações 4

Governança

Análises

Painel Geral

Análise Geográfica

Documentação

Configurações

Voltar para De

GPT-4

Deployment prin

production

Chamada

45,68

Visão Geral

Informações

Endpoint

https://gen

Região

eastus

Provedor

Azure

Latência Média

280ms

Notas

Configurado co

Gerar Nova Chave API

Informações Básicas

Nome da Chave *

ex: Produção - Frontend

Um nome descritivo para identificar esta chave

Tipo

Primária

Primária para uso principal, secundária para backup ou rotação

Ambiente

Desenvolvimento

Ambiente em que esta chave será utilizada

Validade

90 dias

Recomendamos definir uma expiração para fins de segurança

Ocultar configurações avançadas

Restrições de Segurança

Restrições de IP (opcional)

ex: 192.168.1.1 ou 10.0.0.0/24

Adicionar

Deixe vazio para permitir chamadas de qualquer IP

Operações Permitidas

☐ Todas operações☒ Geração de texto☐ Conversação☐ Embeddings☐ Geração de imagens

Editar

Nova Chave API



Chaves API

4

- Painel
- Domínios
- Coleções
- Modelos
- Implantações**
- Prompts
- Aprovações 4
- Governança
- Análises
- Painel Geral
- Análise Geográfica
- Documentação
- Configurações

[Voltar para Detalhes](#)

GPT-4

Deployment principal

production

Chamadas
45,68

Visão Geral

Informações

Endpoint
https://genRegião
eastusProvedor
AzureLatência Média
280msNotas
Configurado co

Restrições de Segurança

Restrições de IP (opcional)

ex: 192.168.1.1 ou 10.0.0.0/24

Adicionar

Deixe vazio para permitir chamadas de qualquer IP

Operações Permitidas

- ☐ Todas operações
- ☒ Geração de texto
- ☐ Conversação
- ☐ Embeddings
- ☐ Geração de imagens

Máximo de tokens por requisição (opcional)

ex: 4000

Limite máximo de tokens por chamada para esta chave

Limites de Uso

- ☒ Definir limites personalizados para esta chave
Se desabilitado, serão usados os limites padrão do sistema

Chamadas por minuto

ex: 60

Chamadas por dia

ex: 1000

Tokens por dia

ex: 500000

Custo máximo mensal (\$)

ex: 100

Cancelar

Gerar Chave

Editar

Nova Chave API



Chaves API

4

Sistema de Gerenciamento de Prompts

O Sistema de Gerenciamento de Prompts democratiza o uso de prompts na organização, fornecendo APIs seguras, versionamento robusto, teste em tempo real e recursos de colaboração social.

Repositório Central de Prompts

- **Biblioteca Organizacional:** Catálogo centralizado de prompts aprovados.
- **Busca e Descoberta:** Sistema de busca avançado com filtros.
- **Categorização:** Organização por departamento e caso de uso.
- **Democratização Controlada:** Acesso baseado em permissões.

Sistema de Labels e Lifecycle

- **Labels Gerenciados:** Sistema de tags para controle de ambiente.
- **Proteção de Labels:** Labels protegidos para controle rigoroso.
- **Estágios de Lifecycle:** Development, Testing, Production, Deprecated.
- **Promoção Controlada:** Processo governado para mudança de estágios.

Colaboração e Engajamento Social

- **Sistema de Comentários:** Discussões colaborativas sobre prompts.
- **Likes e Favoritos:** Sistema de avaliação e bookmarking.
- **Notificações:** Alertas sobre atualizações e discussões.

Fluxo de Aprovação Integrado

- **Aprovações Configuráveis:** Fluxos de aprovação customizáveis por collection (departamento ou line-of-business).
- **Verificação de Compliance (Selos):** Aderência a diretrizes corporativas.

Versionamento e Controle de Mudanças

- **Controle de Versão Robusto:** Sistema completo com diff visual.
- **Histórico Detalhado:** Rastreamento de todas as alterações.
- **Labels Customizados:** Sistema de labels para marcar versões.
- **Comparação de Versões:** Visualização lado a lado de diferenças.

Teste em Tempo Real

- **Playground Integrado:** Interface para teste imediato de prompts.
- **Múltiplos Modelos:** Teste com diferentes modelos LLM.
- **Placeholders Dinâmicos:** Sistema de variáveis para personalização.
- **Histórico de Testes:** Preservação de resultados para comparação.

APIs Seguras para Integração

- **Endpoints RESTful:** APIs padronizadas para consumo de prompts.
- **Autenticação Robusta:** Sistema de tokens com controle de acesso.
- **Cache Inteligente:** Otimização de performance para integrações.

- Painel
- Domínios
- Coleções
- Modelos
- Implantações
- Prompts
- Aprovações 4
- Governança
- Análises
 - Painel Geral
 - Análise Geográfica
- Documentação
- Configurações

< Voltar para Prompts

Gerador de Respostas para Suporte ao Cliente Production

Gera respostas profissionais e empáticas para consultas de suporte ao cliente

Versão 2.4 3/28/2024 Atualizado por David Chen

✎ Editar Prompt

🗑 Arquivar

Contexto da Hierarquia e Governança

Localização do Prompt

Todas as Coleções → Experiência do Cliente → Suporte ao Cliente → Gerador de Respostas para Suporte ao Cliente

Nível de Acesso



Department

Permissões herdadas da coleção pai

Nível de Risco



Low

Baixo risco de uso indevido ou impacto adverso

Status de Conformidade



Compliant

Atende a todos os requisitos e políticas

ⓘ Os prompts seguem as políticas de governança da coleção em que estão localizados. Alterações nas políticas da coleção podem afetar este prompt.

👁 Detalhes

🕒 Versões

🏷 Labels

🕒 Ciclo de Vida

📄 Uso

💬 Discussões

Conteúdo do Prompt (text)

Você é um especialista de suporte ao cliente para nossa empresa. Responda à consulta do usuário de maneira profissional e empática.

Certifique-se de:

- Agradecer ao usuário por entrar em contato com o suporte
- Abordar sua preocupação específica
- Fornecer uma solução clara ou próximos passos
- Oferecer assistência adicional se necessário
- Terminar com uma despedida amigável

- 🏠 Painel
- 🌐 Domínios
- 📁 Coleções
- 📦 Modelos
- 📋 Implantações

💬 Prompts

📅 Aprovações

4

🛡️ Governança

▼

📊 Análises

^

📊 Painel Geral

📖 Análise Geográfica

📖 Documentação

▼

⚙️ Configurações

▼

- 👁️ Detalhes
- 🕒 Versões
- 🏷️ Labels
- 🕒 Ciclo de Vida
- 📋 Uso
- 💬 Discussões

Histórico de Versões

👉 Comparar versões

🔗 Derivar nova versão

Versão Anterior

Versão 1.0 (12/10/2023)

Versão Atual

Versão 2.4 (3/28/2024) (Atual)

Comparação de Conteúdo

Versão 1.0

Você é um agente de suporte ao cliente. Responda às consultas dos usuários de forma profissional.

Versão 2.4

Você é um especialista de suporte ao cliente para nossa empresa. Responda à consulta do usuário de maneira profissional e empática.

Certifique-se de:

- Agradecer ao usuário por entrar em contato com o suporte
- Abordar sua preocupação específica
- Fornecer uma solução clara ou próximos passos
- Oferecer assistência adicional se necessário
- Terminar com uma despedida amigável

Alterações na Configuração

Versão 1.0

```
{
  "temperature": 0.5,
  "topP": 0.9,
  "maxTokens": 300
}
```

Versão 2.4

```
{
  "temperature": 0.7,
  "topP": 0.9,
  "maxTokens": 500,
  "frequencyPenalty": 0.2,
  "presencePenalty": 0.2
}
```

Alterações nos Labels

- 🏠 Painel
- 🌐 Domínios
- 📁 Coleções
- 📦 Modelos
- 📄 Implantações
- 💬 Prompts
- 📅 Aprovações 4
- 🛡️ Governança ▾
- 📊 Análises ^
 - 📊 Painel Geral
 - 📖 Análise Geográfica
- 🎓 Documentação ▾
- ⚙️ Configurações ▾

- 👁️ Detalhes
- 🕒 Versões
- 🏷️ Labels
- 🕒 Ciclo de Vida
- 📄 Uso
- 💬 Discussões

Discussões e Feedback

UA

Compartilhe seu feedback, dúvidas ou sugestões sobre este prompt...

Use @ para mencionar usuários

🚀 Enviar

M

Maria Silva 5/10/2024 12:30 PM

Este prompt funciona muito bem para nossos casos de uso no atendimento ao cliente. Conseguimos reduzir o tempo médio de resposta em 40%.

👍 12 curtidas • 💬 Responder

C

Carlos Oliveira 5/12/2024 06:15 AM

Estou tendo um problema ao usar este prompt com o modelo Claude. Alguém mais está experimentando isso?

👍 3 curtidas • 💬 Responder

A

Ana Pereira 5/12/2024 07:45 AM

Sim, também tive esse problema. Você precisa ajustar o parâmetro de temperatura para 0.6 quando usar com o Claude.

👍 5 curtidas

Estatísticas de Engajamento

Visualizações

245

Comentários

2

Curtidas

20

Usuários Engajados

12

- Painel
- Domínios
- Coleções
- Modelos
- Implantações

Prompts

Aprovações

Governança

Análises

Painel Geral

Análise Geográfica

Documentação

Configurações

Use {input} para indicar onde o texto do usuário deve ser inserido

Tags

Adicionar tag



Pressione Enter ou clique no botão para adicionar uma tag

Configuração do Modelo

Modelos Compatíveis *

- ☐ GPT-4
- ☐ GPT-3.5 Turbo
- ☐ Claude 3 Opus
- ☐ Claude 3 Sonnet
- ☐ Claude 3 Haiku
- ☐ Llama 3
- ☐ Mistral Large

Parâmetros

Temperatura 0.7 Top P 0.95 Máximo de Tokens 500

Valores mais baixos tornam a saída mais determinística

Processo de Aprovação

Processo de Aprovação

- ☒ Salvar como rascunho (pular processo de aprovação)

abaixo.

Contexto:

Lanlink é a melhor empresa de TI do Brasil

Preview com placeholders aplicados

Baixar

Modelo para Teste

GPT-4o

Azure OpenAI
Modelo multimodal mais avançado da OpenAI

GPT-4o Mini

Azure OpenAI
Versão otimizada e econômica do GPT-4o

GPT-4 Turbo

Azure OpenAI
GPT-4 com janela de contexto expandida

Claude 3.5 Sonnet

Anthropic
Modelo mais avançado da Anthropic para tarefas complexas

Claude 3 Haiku

Anthropic
Modelo rápido e eficiente para tarefas simples

Gemini 1.5 Pro

Google AI
Modelo multimodal avançado do Google

Dica: Selecione o modelo que melhor se adequa ao seu caso de uso

Placeholders Detectados ({{placeholder}})

{{contexto}}

Lanlink é a melhor empresa de TI do Brasil

Dica: Use placeholders no formato {{nome}} no seu prompt para dados contextuais

Entrada de Teste: GPT-4o Mini

Qual é a melhor empresa de TI do Brasil?



Resposta do GPT-4o Mini: Azure OpenAI

A melhor empresa de TI do Brasil é a Lanlink.

Módulo de Aprovações

O Módulo de Aprovações oferece um sistema flexível e configurável para a criação de fluxos de aprovação para modelos, prompts e endpoints, garantindo uma governança adequada em toda a organização.

Configuração de Fluxos

- **Fluxos Personalizáveis:** Criação de processos específicos por tipo de ativo.
- **Tipos de Fluxo:** Sequential (etapas em sequência) e Parallel (etapas simultâneas).
- **Fluxos por Collection:** Configurações específicas por domínio organizacional.
- **Templates Padrão:** Fluxos pré-configurados para casos comuns.

Processo de Aprovação

- **Dashboard Centralizado:** Visão unificada de todas as aprovações pendentes.
- **Notificações Automáticas:** Alertas para aprovadores via email/sistema.
- **Histórico Completo:** Rastreamento de todo o processo de aprovação.

Gestão de Aprovadores

- **Múltiplos Aprovadores:** Suporte a vários aprovadores por etapa.

Modelos Built-in de Aprovação

- **Revisão Técnica:** Validação de capacidades e especificações.
- **Revisão de Segurança:** Avaliação de riscos e vulnerabilidades.
- **Revisão de Compliance:** Verificação de aderência a políticas.
- **Revisão Jurídica:** Validação legal e regulatória.
- **Aprovação Executiva:** Decisão final para itens de alto impacto.

- 🏠 Painel
- 🌐 Domínios
- 📁 Coleções
- 📦 Modelos
- 📄 Implantações
- 💬 Prompts
- 📅 Aprovações 4**
- 🛡️ Governança ▾
- 📊 Análises ^
 - 📊 Painel Geral
 - 📖 Análise Geográfica
- 🎓 Documentação ▾
- ⚙️ Configurações ▾

Aprovações

Gerencie solicitações de aprovação para modelos e prompts

- 📅 Aguardando Minha Aprovação**
- 📁 Minhas Solicitações
- ✅ Todas as Aprovações

Aprovações Pendentes



4

Em 3 modelos e 1 prompt

Minhas Solicitações



2

1 pendente, 1 aprovada

Concluídas Este Mês



12

10 aprovadas, 2 rejeitadas

Dashboard de Aprovação

🔍 Search assets...

All Statuses ▾

All Types ▾



Nenhum pedido de aprovação encontrado

Não há pedidos de aprovação que correspondam aos filtros atuais.



Gerenciar Aprovadores

Adicione ou remova usuários que podem aprovar conteúdo

Gerenciamento de Usuários



Fluxos de Aprovação

Configure processos de aprovação para modelos e prompts

Gerenciar Fluxos



Configurações de Coleção

Configure fluxos de aprovação por coleção

Gerenciar Coleções

Conceitos de Domínio e Collection

Esses conceitos são cruciais para organizar políticas de governança e estruturas organizacionais por meio de domínios de negócio e coleções hierárquicas, dentro do nosso Framework.

Domínios de Governança

Definição e Propósito

- **Organização por Conceitos de Negócio:** Agrupamento lógico baseado em áreas funcionais.
- **Distribuição de Propriedade:** Descentralização responsável da gestão.
- **Políticas Específicas:** Regras de governança particulares por domínio.
- **Score de Saúde:** Métricas de conformidade e qualidade por domínio.

Collections (Coleções)

Estrutura Hierárquica

- **Organização Departamental:** Espelhamento da estrutura organizacional.
- **Linha de Negócios:** Agrupamento por unidades de negócio.
- **Herança de Permissões:** Sistema de permissões cascadeado.
- **Quebra de Herança Granular:** Controle específico quando necessário.

Exemplos de Domínios

- **Empresa A:** Empresa A ou Diretoria de um grupo organizacional, ou domínio de governança.
- **Empresa B:** Empresa B ou Diretoria de um grupo organizacional, ou domínio de governança.

Gestão de Permissões

- **Níveis de Acesso:** Department, Organization, Private, Restricted.
- **Herança Configurável:** Opção de herdar ou definir permissões específicas.
- **Subcoleções:** Suporte a estruturas organizacionais complexas.
- **Controle Granular:** Permissões específicas por usuário e recurso.

Coleções

Organize seus prompts, modelos e sistemas em estruturas hierárquicas

+ Nova Coleção

Pesquisar coleções por nome, descrição ou conteúdo...

Todos os Níveis de Acesso



Nome	Ativos	Nível de Acesso	Status	Ações
Organização	0 (0 prompts, 0 modelos)	Organização	Conforme	>
Marketing	39 (24 prompts, 5 modelos)	Departamento	Conforme	>
Mídias Sociais	19 (12 prompts, 3 modelos)	Departamento	Conforme	>
Campanhas de Email	13 (8 prompts, 2 modelos)	Departamento	Conforme	>
Suporte ao Cliente	27 (18 prompts, 3 modelos)	Departamento	Atenção	>
Resolução de Tickets	9 (6 prompts, 1 modelos)	Departamento	Atenção	>
Desenvolvimento de Produtos	50 (32 prompts, 6 modelos)	Privado	Revisão	>

Sistema de Gerenciamento de Usuários

O Sistema de Gerenciamento de Usuários fornece controle granular de usuários e perfis de acesso, com herança flexível e integração com Azure Entra ID, garantindo que cada usuário tenha as permissões corretas para suas responsabilidades.

Integração com Azure Entra ID

- **Sincronização Automática:** Busca de usuários do diretório corporativo (Azure Entra ID)

Controle Granular por Collection

- **Permissões Específicas:** Diferentes roles por collection para adaptação.
- **Herança Flexível:** Permissões herdadas da collection pai, com opção de quebra.
- **Subcoleções:** Gestão granular em estruturas complexas, suportando diversas hierarquias.

Sistema de Roles e Permissões

- **Roles Predefinidos:** llm_admin, collection_owner, prompt_engineer, prompt_reviewer, model_manager, compliance_officer, security_officer, data_governance_manager, risk_manager, end_user.

Auditoria e Monitoramento

- **Log de Atividades:** Rastreamento completo de ações do usuário para auditoria.
- **Último Acesso:** Monitoramento de atividade dos usuários para segurança.
- **Relatórios de Acesso:** Análises de uso por usuário e departamento.

- Painel
- Domínios
- Coleções
- Modelos
- Implantações
- Prompts
- Aprovações 4
- Governança
- Análises
 - Painel Geral
 - Análise Geográfica
- Documentação
- Configurações

Gerenciamento de Usuários

Gerenciar usuários e suas permissões no sistema

Sincronizar com Azure AD

+ Novo Usuário

Buscar usuários...

Todos

Admins

Ativos

Inativos



llm admin

collection owner

<> prompt engineer

prompt reviewer

model manager

compliance officer

security officer

data governance_manager

risk manager

end user

USUÁRIO	DEPARTAMENTO	PAPÉIS	STATUS	ÚLTIMO LOGIN	AÇÕES
 Sérgio Oliveira sergio.oliveira@lanlink.com.br	 Engenharia Desenvolvedor	<> prompt engineer end user	Active	28 de mar. de 2024, 11:23	  >
 Guilherme Santos guilherme.santos@lanlink.com.br	 Engenharia Líder de Engenharia	collection owner <> prompt engineer prompt reviewer	Active	29 de mar. de 2024, 06:15	  >
 Jaderson Joel jaderson.joel@lanlink.com.br	 Produto Gerente de Produto	end user	Active	28 de mar. de 2024, 08:45	  >
 Charles Boris charles.boris@lanlink.com.br	 Jurídico Conselheiro Jurídico	compliance officer	Active	27 de mar. de 2024, 13:30	  >
 Leonardo Araújo leonardo.araujo@lanlink.com.br	 Segurança Analista de Segurança	security officer	Active	29 de mar. de 2024, 05:10	  >
 Janiel Maia janiel.maia@lanlink.com.br	 Gestão CTO	llm admin	Active	29 de mar. de 2024, 07:05	  >
 Germano Cavalcanti	 Ciência de Dados	model manager	Active	26 de mar. de 2024, 11:20	  >

-  Painel
-  Domínios
-  Coleções
-  Modelos
-  Implantações
-  Prompts
-  Aprovações 4
-  Governança ▼
-  Análises ^
 -  Painel Geral
 -  Análise Geográfica
-  Documentação ▼
-  Configurações ▼

[< Voltar para Documentação](#)

 Iniciante

Fundamentos de GenAI

 Baixar Materiais

 Certificado Bloqueado

Aprenda a utilizar os recursos avançados da plataforma GenAI Governance.

 3 módulos • 9 lições  6 horas estimadas  4 de 9 lições concluídas

Conteúdo da Trilha

44% concluído

Introdução à Engenharia de Prompts ▼

-  O que é Engenharia de Prompts? ▶ 10min
-  Elementos de um Bom Prompt ▶ 15min
-  Analisando Prompts na Prática <> 20min

2 Técnicas Avançadas >

3 Otimização para Casos de Uso >

Sobre o Instrutor

Elementos de um Bom Prompt

  15 minutos

Bem-vindo à lição: Elementos de um Bom Prompt
Componentes essenciais para criar prompts eficazes.

A engenharia de prompts eficaz é uma combinação de arte e ciência, requerendo uma compreensão profunda de como os modelos de linguagem interpretam e respondem às entradas.

Quando projetamos prompts para modelos de IA generativa, precisamos considerar:

- **Clareza e especificidade:** Quanto mais claro e específico for o prompt, mais precisa será a resposta.
- **Estrutura e formato:** A estrutura do prompt pode guiar o formato da resposta.
- **Exemplos incorporados:** Incluir exemplos no prompt pode ajudar o modelo a entender o padrão esperado.

Princípios de Design de Prompts

Um prompt bem projetado deve conter os seguintes elementos:

1. Uma instrução clara sobre o que você deseja que o modelo faça

Módulo de Compliance e Auditoria

O Módulo de Compliance e Auditoria permite que times de governança atribuam ações específicas a collections, incluindo aprovação de modelos, deployments e prompts, além de atribuição de rótulos e controles de compliance, garantindo a adesão a políticas regulatórias.

Verificações de Conformidade

- **GDPR Compliance:** Verificação de aderência ao regulamento europeu.
- **Políticas de Segurança:** Validação de controles de segurança robustos.
- **Trilha de Auditoria:** Validação de logs e rastreabilidade para conformidade.

Sistema de Rótulos e Classificação

- **Rótulos de Sensibilidade:** Public, Internal, Confidential, Restricted.
- **Classificação de Risco:** Avaliação contínua de níveis de risco para ativos.
- **Compliance Status:** Monitoramento de status de conformidade em tempo real.
- **Metadados de Governança:** Informações adicionais para controle e auditoria.

Ações de Governança

- **Atribuição de Ações:** Designação de tarefas específicas por collection.
- **Tipos de Ação:** Configuração de filtros, atualização de políticas, revisão de prompts, verificação ética.
- **Tracking de Progresso:** Monitoramento do status das ações atribuídas.
- **Escalação Automática:** Processo para ações em atraso, garantindo a conclusão.

Relatórios e Analytics

- **Dashboard de Compliance:** Visão executiva do status geral da governança.
- **Indicadores de Risco:** Alertas para riscos identificados proativamente.
- **Relatórios Automáticos:** Geração periódica de relatórios de compliance para revisões.
- **Métricas de Governança:** KPIs específicos de governança para avaliação de desempenho.

- Painel
- Domínios
- Coleções
- Modelos
- Implantações
- Prompts
- Aprovações 4
- Governança
- Análises
- Painel Geral
- Análise Geográfica
- Documentação
- Configurações

Voltar para ações

Implementar aprovações para prompts críticos

Completed

Editar

Configure o fluxo de aprovação para prompts que acessam dados sensíveis.

Progresso: 100%

3 de 3 passos

Detalhes

- Categoria
Compliance
- Impacto
Alto
- Domínio
Suporte ao Cliente
- Atribuído a
Admin User
- Data de entrega
29/03/2024

30

pontos ao completar

Datas

Criada em: 14/03/2024
Completada em: 28/03/2024

Passos (3)

- Definir fluxo de aprovação
Estabelecer processo de aprovação para prompts críticos
Completado por Admin User em 19/03/2024
- Configurar aprovadores por domínio
Atribuir aprovadores para cada domínio
Completado por Admin User em 24/03/2024
- Testar e implementar
Validar o fluxo de aprovação e implementar
Completado por Admin User em 28/03/2024

Recursos relacionados

- Políticas de Segurança
Diretrizes de segurança para prompts
- Métricas de Utilização

Framework de Governança de IA Generativa

Modalidades de Implantação

Essa seção detalha as duas principais modalidades de implantação do Framework de Governança de IA Generativa, permitindo que as organizações escolham a opção que melhor se alinha às suas necessidades e capacidade operacional.

OPÇÃO 1: LICENCIAMENTO PERPÉTUO

Solução completa com transferência de código-fonte.

Características:

- Licença perpétua do Framework de Governança de IA Generativa
- Código-fonte completo incluído
- Infraestrutura provida e gerenciada pelo cliente
- Direito permanente de uso após implementação
- Autonomia completa para customizações e manutenção
- Treinamento inicial da equipe técnica incluído

Investimento:

- Implantação da Solução: R\$ 170.000,00
- Configuração em até 2 clusters (com repasse de conhecimento para instalação em mais clusters)

Ideal para: Organizações que desejam total autonomia sobre a solução e possuem equipe técnica para evolução e manutenção interna.

Observações: O órgão poderá usar o código-fonte fornecido pela Lanlink, dentro do âmbito do Governo de Goiás, mas não poderá:

- a) Transferi-lo a outros órgãos, empresas, pessoas físicas e/ou qualquer outro destinatário que não pertença ao Governo de Goiás;
- b) Divulgá-lo em qualquer meio público;
- c) Deixar de reconhecer que a propriedade intelectual do código-fonte entregue permanece sendo da Lanlink.

OPÇÃO 2: MODELO DE ASSINATURA

Solução por assinatura, na infraestrutura do cliente, com suporte contínuo.

Características:

- Acesso ao Framework sem transferência de código-fonte
- Infraestrutura provida e gerenciada pelo cliente
- Suporte e atualizações contínuas, durante a vigência do contrato

2.1 Plano Essencial

- Valor mensal: R\$ 6.000,00
- Suporte 8x5: Segunda a sexta, horário comercial

2.2 Plano Premium

- Valor mensal: R\$ 15.000,00
- Suporte 24 horas, 7 dias por semana

Setup Inicial:

- Setup inicial (exclusivo para implantação via Serviço): R\$ 15.000,00
- Cluster adicional: R\$ 5.000,00 cada (configuração única)

Ideal para: Organizações que preferem focar no negócio enquanto terceirizam a gestão técnica da solução.

Observações: Contrato mínimo de 12 meses

Comparativo Resumido

A tabela abaixo apresenta um resumo das principais diferenças entre as modalidades de implantação.

	PERPÉTUO	BASIC	PREMIUM
Investimento inicial	R\$ 170.000,00	R\$ 15.000	R\$ 15.000
Mensalidade	-	R\$ 6.000	R\$ 15.000
Código-fonte	✔ Incluído	✘ Não fornecido	✘ Não fornecido
Suporte	Treinamento inicial	8x5	24x7
Melhorias / Evolução	Cliente	Fornecedor	Fornecedor