



🔗 Documentação

Dados bons, resultados ótimos

Termos do GOgenier

📄 Termos de uso

O poder do **GOgenier** sobre dados não *estruturados*.

O **GOgenier** é um poderoso modelo de linguagem que pode ser usado para gerar conteúdo para uma variedade de formatos, incluindo texto, código e dados.

O **GOgenier** é particularmente eficaz para gerar conteúdo para dados não estruturados. Dados não estruturados são dados que não estão organizados e podem gerar conteúdo que é relevante e informativo.

Passo a passo para brincar!

1 Configure a API do LLM

- Primeiro você precisa ter acesso a umas das opções de API de LLM, seja OpenAI ou Google Gemini (Vertex)

2 Crie uma fonte de dados

- Após ter configurado a API em *Configurações->LLM*, crie uma fonte de dados com algum conteúdo em *Fonte de dados*

3 Crie um modelo

- Com a fonte de dados criada, é hora de criar um modelo com esse conteúdo, basta ir em *Modelos*, cadastrar o modelo e logo abaixo, processar.

🎮 Brincar

- Com o modelo criado, agora vá para *Playground* e utilize uma das 3 opções Chat, Copilot e Serach para realizar consultas.

🌟 Brincadeira Avançada

- A brincadeira fica muito melhor quando você cria agentes inteligentes, eles funcionam como agentes humanos e tem por definição, início, meio e fim.

💡 Boas práticas

Dicas para gerar conteúdo para RAG

Aqui estão algumas dicas para gerar conteúdo para o Genier:

Seja **claro** e **conciso** em suas instruções. O Genier é capaz de entender instruções complexas, mas é sempre melhor ser claro e direto.

Textos **simples** como *descrição de produto*, *instruções de uso* e *características* são ótimos para gerar conteúdos em **perguntas e respostas**.

Exemplo de conteúdo:

Fechadura eletrônica de embutir, 20cm x 3cm de espessura, suporta Tag e Biometria. A fechadura smart de embutir com maçaneta IFR 7000 faz por luzes, visualizar imagens de câmeras, ligar e desligar equipamentos e muito mais. A IFR 7000 conta com abertura por até 9 senhas, até 100 tags de acesso (considerando 10 acessos diários). A fechadura smart IFR 7000 conta com funções como travamento automático, função não perturbe, gerenciamento inteligente e app - item vendido separadamente). Com mortise padrão ABNT 14913 a IFR 7000 possui uma das instalações mais fáceis do mercado.

Responde pergunta do tipo

Em uma porta de entrada de 2cm qual fechadura digital posso utilizar?

💡 Relevância de conteúdos

A recuperação de conteúdo para construção de contextos é limitada em **N** (pode ser configurado na criação do modelo) pedaços (chunks). Esses

Se o conteúdo buscado é genérico (ex: Tem WiFi?) e consta em muitos pedaços de conteúdo, pode ser que a resposta esperada estará em um p
Nesses casos, é interessante incrementar conteúdo específico para esse contexto, exemplo: em um catalogo de muitos produtos que tem inter

💡 Complemento de conteúdos

Ainda por questões de quantidade máxima de chunks (pedaços) de conteúdo, as vezes é necessário criar conteúdos complementares para resq

Por exemplo: Imagine que você tenha uma lista grande de um tipo de produto, voltamos aqui para fechaduras, se cada documento fala sobre u
possivelmente não terá a lista completa de fechaduras, pois apenas uma parte dos documentos serão utilizados.

Para resolver casos como esse, adicione um novo documento com a lista de todas as fechaduras, sem muitos detalhes, pois os detalhes estão r

💡 Conteúdos da WEB

Sabemos que boa parte dos conteúdos podem já estar disponível na WEB, seja no seu website ou intranet, por isso o **GOgenier** disponibiliza d
conteúdos dinâmicos, entretanto, existem outros fatores que podem atrapalhar o scrap dos sites, como firewall e complexidade na estrutura d

QuickScan é ótimo para sites estáticos ou com pouca complexidade de JS.

DeepDive vai melhor com sites complexos, mas não faz recursividade, ou seja, é importante passar a lista de URLs que pode ser adquirida via s

❓ Meu agente está respondendo conteúdo que eu não quero

O agente segue um script, se você der a tarefa de responder sobre algo, mas não forneceu o conteúdo, **ele irá criar**.

❓ Insights

O conteúdo transcrito em insights é submetido aos LLMs para realizar análises e geração de conteúdos, portanto fica limitado aos tamanhos d

❓ Recebendo as entidades capturadas nos agentes

Para receber as entidades (propriedades) capturadas na conversa do agente na resposta da API, basta ter alguma instrução do tipo: *Gere um json*

❓ Erro no copilot/chat

A mensagem de erro recebida quase sempre está relacionada ao modelo criado vazio, isso ocorre pelos seguintes motivos, 1. a chave de api do
utilize o search no playground.

❓ Posso vender via agente

Com certeza, agentes de vendas é um dos propósitos do GOgenier, você pode criar um agente que atue apenas nas relações e descobertas de i
disponibilidade de estoque, valor de vendas, descontos ou qualquer outro dado muito dinâmico precisa ser usado com cautela, pois em uma
com a condição de ressalvas como disponibilidade e preços.

❓ Recuperação de documentos (modelos)

Quando um ou mais documentos é importado em fonte de dados, faz parte do processo dividir esses dados em "pedaços", a divisão é necessári
de janelas (quantidade de tokens) permitidos durante uma interação.

O GOgenier permite manipular essa quebra de documentos de diversos tamanhos e formas. Para entender melhor, vamos exemplificar com un
ou +/- 1600 tokens. Se não existisse particionamento, cada busca no documento iria gerar 16 mil tokens. Uma vez particionado, o GOgenier vai
conteúdo relevante para a resposta, nesse caso o sistema irá gastar apenas 400 tokens.

O "N" no exemplo também é definido, onde o mínimo é 1 e o máximo são 10. Esse "N" pode ser alterado conforme conhecemos o tipo dos dac

❓ Quando devo mudar de TIER?

Alguns fatores para mudança de tier são bem claros, como limites de espaço de armazenamento, limites de agentes e número de requisições p
sistema e assim necessitar a auteração do tier. Como não temos controle do tamanho e quantidade de integrações, o termômetro fica por sua

A maioria das vezes, o tempo de resposta está muito mais limitante no LLM, mas grandes ou complexas integrações podem impactar e nesse ca

💡 Integração GOgenier Flow

O GOgenier disponibiliza uma plataforma de integração nativa, isso facilita muito o intercambio da IA Generativa, mais especificamente os age
criar as integrações e aciona-las via linguagem natural.

Como acionar o flow:

1. Crie uma instrução, por exemplo: "Crie um json com a chave name" e abra um chamado
2. Adicione um ação do tipo Genier Flow

- 3. Na descrição, coloque um texto que remeta a instrução, por exemplo "Abertura de chamados"
- 4. Em caminho, coloque o path da integração do flow
- 5. Em rota, coloque o caminho configurado no componente "Agent Executor"

Como o flow recebe os dados:

Quando uma instrução aciona o flow, os dados passados serão recebidos no componente "Agent Executor", normalmente em JSON, mas depe

Recomendável utilizar autenticação na integração.

💡 Database

A partir da versão 1.1.0 do GOgenier, foi adicionado um gerenciador de banco de dados. Esse gerenciador tem os seguintes propósitos:

- 1. Disponibilizar uma area de dados dinamicos facilitada para criação de agentes com coleta de dados ou consulta de dados.
- 2. Evitar a necessidade do cliente expor o acesso ao seu banco de dados.
- 3. Disponibilizar logs do sistema, inicialmente alguns focados em auditoria.

O recurso pode ser acessado para cadastro e consultas através dos agentes de forma simples na ações.

Além de permitir acesso aos dados via agentes, esses dados podem ser acessados (cadastro/consulta) via API REST para integrações externas.

O QUE ELE NÃO É

Esse ambiente não é um repositório de dados de larga escala e alta performance, e também não contempla DR nem mesmo alta disponibilidade do tier.

💡 Webchat

A partir da versão 1.2.0 do GOgenier, foi disponibilizado um Webchat de fácil integração, basta colar o código do iframe que se encontra nas cor script, alterar e customizar e hospedar em seu próprio site, utilize os parâmetros conforme o iframe nas configurações do agente.

Para baixar, basta usar o link <https://app.genier.ai/fortics/godb/webchat.zip>

💡 Opções de LLM

Cada LLM pode se comportar diferente, sempre que alterar o LLM do agente é necessário revisar seu funcionamento. A implementação do Gern disponibiliza gratuitamente chaves de API para testes do Gemini, mas atente-se aos limites de uso, como quantidade máxima de tokens e rate l serviço.

*Preços de LLM 22/09/2024

Empresa	LLM	Milhão de Tokens Input	Output
OpenAI	GPT-4o	\$5.00	\$15.00
OpenAI	GPT-4o-mini	\$0.15	\$0.60
Google	Gemini Pro	\$3.50	\$10.50
Google	Gemini Flash	\$0.075	\$0.30

⚙ Changelog

- 1.0.0 Lançamento do GOgenier
- 1.1.0 Banco de dados para auxílio em agentes
- 1.2.0 Webchat com personalização
- 1.3.0 Adicionado opção de recuperação de senhas
- 1.4.0 Adicionado suporte ao Gemini (beta)
- 1.5.0 Adicionado suporte a dados estruturados com CSV.

