

# Configurações do Chat (Busca e Modo de Resposta)

Este documento orienta a escolha das configurações do chat para obter respostas mais adequadas ao contexto institucional, com maior precisão e menor risco de erro. As opções se organizam em dois eixos: o **Tipo de busca** (como o chat procura informação na base documental) e o **Modo de resposta** (quanto o chat aprofunda a busca e estrutura a resposta antes de responder).

## Tipo de busca

O **Tipo de busca** deve ser selecionado conforme a natureza do que se pretende localizar nos documentos.

- **Palavra-Chave:** indicada quando há necessidade de localizar informações objetivas e específicas, como número de processo, protocolo, datas, nomes, itens, cláusulas, referências ou trechos literais. Funciona melhor quando o usuário informa termos exatos. Como limitação, pode falhar quando há sinônimos, abreviações ou variações de escrita.
- **Similaridade:** indicada quando a solicitação depende do sentido do conteúdo e não de palavras exatas, como entender do que trata um documento, identificar justificativas, fundamentos, contexto, encaminhamentos ou explicar uma decisão. Como limitação, pode retornar trechos “parecidos”, mas não necessariamente o ponto exato desejado, exigindo validação.
- **Híbrida:** combina Palavra-Chave e Similaridade e, por isso, é recomendada como opção padrão em cenários comuns, especialmente quando a pergunta mistura identificadores exatos com demanda interpretativa (ex.: “Processo X: qual foi a última determinação?”). Como limitação, pode trazer mais resultados e exigir seleção mais cuidadosa do que é realmente pertinente.

## Modo de resposta

O **Modo de resposta** define o nível de profundidade e o esforço de organização da resposta.

- **Modo instantâneo:** prioriza agilidade, realizando busca mais simples e respondendo rapidamente. É adequado para demandas diretas, rotineiras e de baixo risco (ex.: localizar um dado pontual).

- **Modo pensamento:** realiza busca mais profunda e organiza melhor a resposta. Deve ser priorizado quando houver maior complexidade, múltiplas condições, necessidade de síntese estruturada, comparação entre alternativas, análise de consistência, documentos longos ou risco institucional associado ao erro. Pode demandar mais tempo e gerar respostas mais extensas quando não houver orientação de concisão.
- **Modo direto:** responde sem consultar documentos. É indicado para revisão e padronização de textos, redação institucional e explicações conceituais gerais quando a resposta não depende da base documental. Não deve ser utilizado quando a solicitação exige fidelidade a conteúdo específico de processo ou documento.

## Regras práticas de escolha

Para padronização de uso, recomenda-se a seguinte lógica de decisão:

1. **A resposta precisa estar ancorada em documento/processo?** Se sim, evitar **Modo direto**.
2. **Há termo exato (número, data, item, nome, cláusula)?** Se sim, priorizar **Palavra-Chave**; se não, priorizar **Similaridade**; em caso de dúvida, usar **Híbrida**.
3. **A demanda tem impacto institucional relevante ou é complexa?** Se sim, priorizar **Modo pensamento**; se não, **Modo instantâneo** tende a ser suficiente.

## Exemplos de combinações recomendadas

- “Localize onde o documento menciona ‘UST’.” → **Palavra-Chave + Instantâneo** (ou **Pensamento** se o documento for extenso)
- “Resuma a última decisão e encaminhamentos do processo.” → **Híbrida + Pensamento**
- “Reescreva este e-mail com tom institucional e mais objetivo.” → **Modo direto**
- “Compare duas alternativas e indique riscos e limitações.” → **Híbrida + Pensamento**

## Diretriz de operação e risco

Em demandas documentais e de maior impacto (jurídico/processual, licitação, decisões administrativas e temas técnicos críticos), recomenda-se priorizar **Híbrida + Modo pensamento**, salvo quando houver necessidade clara de termos exatos (caso em que **Palavra-Chave** se torna preferencial). Sempre que a resposta apresentar omissões, contradições, extrapolações ou tom inadequado, isso deve ser registrado no feedback, pois influencia diretamente a calibração e a melhoria contínua do sistema.

---

Revisão #1

Criado 2026-02-04 11:49:31 UTC por fbarbosa

Atualizado: 2026-02-04 11:50:01 UTC por fbarbosa