Starting April 29, 2025, Gemini 1.5 Pro and Gemini 1.5 Flash models are not available in projects that have no prior usage of these models, including new projects. For details, see **Model versions and lifecycle** (/vertex-ai/generative-ai/docs/learn/model-versions#legacy-stable).

Retrieval and generation output of Vertex AI RAG Engine

This page explains each field of the output from Vertex AI RAG Engine.

retrieveContexts

This section describes each field defined in the `retrieveContexts` API and uses the fields in sample code.

Fields

Field name	Description
`source_uri`	The original source file before it's imported into RAG. If the file is imported from Cloud Storage or Google Drive, `source_uri` is the original file URI in Cloud Storage or Drive. If the file is uploaded, `source_uri` is the file's display name.
`source_display_name`	The file's display name.
`text`	The text chunk that is relevant to the query.
`score`	The similarity or distance between the query and the text chunk. The similarity or distance depends on the `vectorDB` that you choose. For `ragManagedDB`, the score is the `COSINE_DISTANCE`.

Sample output

This code sample demonstrates the use of the fields to produce sample output.

```
contexts {
    source_uri: "gs://sample_folder/hello_world.txt"
    source_display_name: "hello_world.txt"
    text: "Hello World!"
    score: 0.60545359030757784
  }
```

generateContent

Most of the fields defined for the `generateContent` API can found in the Response body (/vertex-ai/generative-ai/docs/model-reference/inference#response).

Fields

This section describes each field defined in the `grounding_metadata` part of the `generateContent` API and uses the fields in sample code.

Field name	Description
`text`	The response generated by Gemini.
`grounding_chunks`	The chunks returned by Vertex AI RAG Engine.
`retrieved_context`	A repeated field that can have zero or more chunks used to ground the generated content.
• `uri`	• The `source_uri` specifies where the data is originally stored.
• `title`	• The `source_display_name` is the filename or display name of the original file.
• `text`	• The text chunk is used to ground the Gemini response.
`grounding_supports`	The relationship between the generated content and the grounding chunks. This is a repeated field. Each `grounding_supports` field shows the relationship between one text segment of the generated context and one or more text chunks that are RAG retrieved.
`segment`	The grounded text segment of the generated text.

Field name	Description
• `start_index`	• The first index of the grounded text. If the `start_index` is missing, then the `start_index` is `0`.
• `end_index`	• The last index of the grounded text.
• `text`	• The grounded text.
`grounding_chunk_indices`	The chunk that's used to ground the text segment. There can be more than one chunk used to ground the text. The index starts from `0`, which represents the first chunk in the `grounding_chunks` field. The ground is on the entire chunk. The part of the chunk that grounds the response isn't specified.
`confidence_scores`	The score that's used to ground the text on a given chunk. The highest score possible is `1` and the higher the score, the higher the confidence level. Each score matches each `grounding_chunk_indices`. Only the chunks with a confidence score of at least `0.6` are included in the output.

Sample output

This code sample demonstrates the use of the fields to produce sample output.

```
candidates {
  content {
    role: "model"
    parts {
      text: "The rectangle is red and the background is white. The rectangle appears
    }
  }
  grounding_metadata {
    grounding_chunks {
      retrieved_context {
        uri: "a.txt"
        title: "a.txt"
        text: "Okay , I see a red rectangle on a white background . It look
      }
    }
    grounding_chunks {
      retrieved_context {
        uri: "b.txt"
        title: "b.txt"
        text: "The video is identical to the last time I described it . It
      }
    }
```

```
   grounding_chunks {
     retrieved_context {
       uri: "c.txt"
       title: "c.txt"
       text: "Okay ,    I    remember    the    rectangle    was    blue    in    the    past    session
     }
   }
   grounding_supports {
     segment {
       end_index: 49
       text: "The rectangle is red and the background is white."
     }
     grounding_chunk_indices: 2
     grounding_chunk_indices: 0
     confidence_scores: 0.958192229
     confidence_scores: 0.992316723
   }
   grounding_supports {
     segment {
       start_index: 50
       end_index: 120
       text: "The rectangle appears to be on some type of document editing software.
     }
     grounding_chunk_indices: 0
     confidence_scores: 0.98374176
   }
 }
}
```

What's next

- To learn more about RAG context in the API reference, see Context
  (/vertex-ai/generative-ai/docs/reference/rest/v1/projects.locations/retrieveContexts).

- To learn more about RAG, see Vertex AI RAG Engine overview
  (/vertex-ai/generative-ai/docs/rag-overview).