

# 生成 AI で進化するアプリ開発

## Vertex AI Codey と Enterprise Search とは

Google Cloud

アプリケーション モダナイゼーション スペシャリスト

諏訪 悠紀

<b>生成 AI を活用したアプリケーション開発</b>	<b>01</b>
<b>Enterprise Search と Retail Search</b>	<b>02</b>
<b>Vertex AI Codey と Duet AI</b>	<b>03</b>
<b>Next Steps</b>	<b>04</b>

# スピーカー紹介

Google Cloud のアプリケーション モダナイゼーション  
スペシャリスト。

スマホアプリ開発やサーバーサイド アプリケーション開発を経  
験したのち、現在は Tech Acceleration Program (TAP) など  
を通してサーバーレスやコンテナなどを活用したアプリケー  
ション開発の技術支援をしている。



**諏訪 悠紀**

Google Cloud  
アプリケーション  
モダナイゼーション  
スペシャリスト

01

# 生成 AI を活用した アプリケーション開発

# 企業のあらゆるユーザーの生産性を向上させ、 全く新しい顧客体験の変革を実現

生成 AI が組み込まれた  
アプリを使う

業界特化  
ソリューション

Duet AI

生成 AI を  
組み込んだアプリを  
簡単に作る



Conversation AI



Enterprise Search

PaLM 2 をはじめとする  
モデルを利用、カスタマイズして  
より高度なアプリを開発



Vertex AI

モデルの利用、学習、管理  
統合 ML プラットフォーム

基盤モデル



テキスト



チャット



コード



ビデオ



画像



音声と音楽

基盤モデルを研究、開発

Google Cloud

データ処理インフラストラクチャ  
GPUs/TPUs | BigQuery | Matching Engine



ビジネスユーザー向け  
生成 AI 組み込み SaaS



開発者向け  
生成 AI 利用のための PaaS



AI 開発者向け  
生成 AI の開発  
および自社業務への適合PaaS

# 企業のあらゆるユーザーの生産性を向上させ、 全く新しい顧客体験の変革を実現

生成 AI が組み込まれた  
アプリを使う

業界特化  
ソリューション

Duet AI

生成 AI を  
組み込んだアプリを  
簡単に作る



Conversation AI



Enterprise Search

PaLM 2 をはじめとする  
モデルを利用、カスタマイズして  
より高度なアプリを開発



Vert

モデルの利用  
統合 ML プラットフォーム

このセッションでは  
こちらにフォーカス

基盤モデル



テキスト



チャット



コード



ビデオ



画像



音声と音楽

基盤モデルを研究、開発

Google Cloud

データ処理インフラストラクチャ  
GPUs/TPUs | BigQuery | Matching Engine



ビジネスユーザー向け  
生成 AI 組み込み SaaS



開発者向け  
生成 AI 利用のための PaaS



AI 開発者向け  
生成 AI の開発  
および自社業務への適合PaaS

# アプリケーション開発のトレンドと生成 AI



生成 AI を  
組み込む

## 生成 AI × ローコード



Conversation AI



Enterprise Search



Duet AI  
in AppSheet



生成 AI で  
アプリ開発を  
支援する

## 生成 AI × コーディング



Vertex AI  
Codey

Code Completion

Code Generation

Code Chat

Duet AI  
for Developers

# ローコード・ノーコードによるアプリケーション開発への現状

# 26.9 billion USD

ローコード開発技術による収益 (2023 年)

Gartner®, Gartner Press Release, "Gartner Forecasts Worldwide Low-Code Development Technologies Market to Grow 20% in 2023", December 13, 2022.  
<https://www.gartner.com/en/newsroom/press-releases/2022-12-13-gartner-forecasts-worldwide-low-code-development-technologies-market-to-grow-20-percent-in-2023>

GARTNERはGartner Inc.または関連会社の米国およびその他の国における登録商標およびサービスマークであり、同社の許可に基づいて使用していますAll rights reserved.

**Table 1. Low-Code Development Technologies Revenue (Millions of U.S. Dollars)**

	2021	2022	2023	2024
Low-Code Application Platforms (LCAP)	6,324	7,968	9,960	12,351
Business Process Automation (BPA)	2,416	2,585	2,761	2,940
Multiexperience Development Platforms (MDXP)	2,081	2,508	2,999	3,563
Robotic Process Automation (RPA)	2,350	2,892	3,401	3,879
Integration Platform as a Service (iPaaS)	4,680	5,668	6,668	7,838
Citizen Automation and Development Platforms (CADP)	554	732	953	1,232
Other Low-Code Development (LCD) Technologies*	92	109	126	146
<b>Total</b>	<b>18,497</b>	<b>22,462</b>	<b>26,869</b>	<b>31,949</b>

\*Other low-code development technologies include rapid mobile app development (RMAD) tools and rapid application development (RAD) tools. RAD tools are/were on-premises only and desktop-bound applications mainly. Low code is the evolution of RAD to cloud and SaaS models.

Notes: Gartner defines a no-code application platform as an LCAP that only requires text entry for formulae or simple expressions. The LCAP market, therefore, includes no-code platforms. Furthermore, "no code" is not a sufficient criterion for tasks like citizen development, as many complex tooling configuration tasks are no code but still require specialist skills. Columns may not add to totals shown because of rounding.

Source: Gartner (December 2022)



# AIによるコーディングの現状

## スピードの向上



## 品質の向上



Research: quantifying GitHub Copilot's impact on developer productivity and happiness (2022-09)  
<https://github.blog/2022-09-07-research-quantifying-github-copilots-impact-on-developer-productivity-and-happiness/>

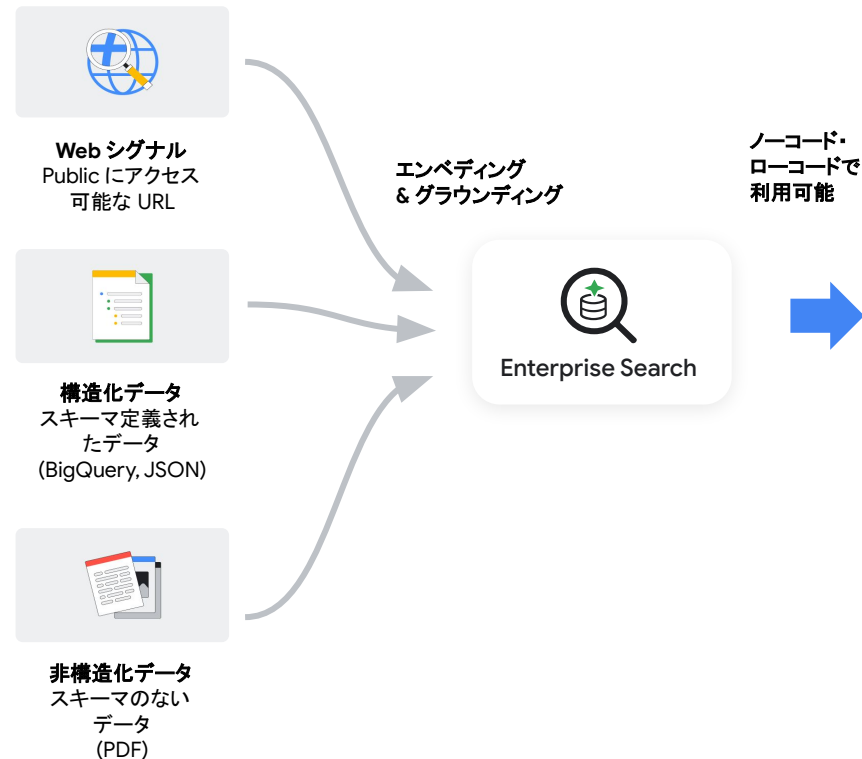
生成系AIがもたらす開発ワークフローの変化(2023-06)  
<https://gihyo.jp/article/2023/06/how-generative-ai-is-changing-the-way-developers-work>

02

# Enterprise Search と Retail Search

# Enterprise Search

簡単な導入でパワフルな検索体験を提供



検索クエリの  
意味理解

概要の生成

出典の引用/  
帰属の説明

パーソナライズ/  
レコメンド

The screenshot shows the Cymbal Investments search interface. The search query is 'What challenges do Semiconductor companies face because of rising interest rates and...'. The results are categorized under 'Overall Findings' and include a 'Cross Source Synthesized Summarization'. The summary text reads: 'The following companies have the most risk exposure to being affected by current inflation and Fed interest rates and have newly appointed board directors: ACME is exposed to the risk of higher interest rates because it borrows money to finance its operations...'. Below the summary, there are two cards: 'Organization with exposure to Neon Gas shortage' by Manifest Consulting and 'Environmental regulation impacting Neon Gas Production' by Sustainable Enterprise edition. The Google Cloud logo is visible in the bottom right corner.

Cymbal Investments

What challenges do Semiconductor companies face because of rising interest rates and...

Date Author Source Type More

Overall Findings

Cross Source Synthesized Summarization

The following companies have the most risk exposure to being affected by current inflation and Fed interest rates and have newly appointed board directors:

ACME is exposed to the risk of higher interest rates because it borrows money to finance its operations...

Organization with exposure to Neon Gas shortage  
Manifest Consulting

Environmental regulation impacting Neon Gas Production  
Sustainable Enterprise edition

Google Cloud

# 検索や会話のためのアプリケーションを 容易に作成できるように、生成 AI の各機能を統合



# Enterprise Search で実現できるユーザー体験



データをよりアクセスしやすく、便利に

多くの組織は膨大な社内データを保有しており、必要としている情報を得るには多くの時間がかかります。生成AIを活用することでデータを迅速かつ効果的にQ&Aでき、生産性を高めることができます。

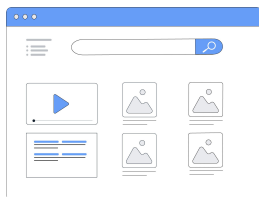


# セマンティック検索 (意味検索) とは

従来型 (全文検索)



ユーザー



アプリ



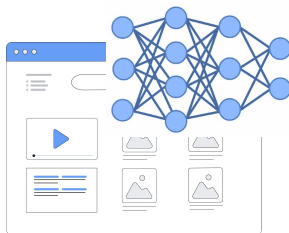
ID	Name	City
001	Foo	NYC
002	Bar	LDN

構造化データとキーワード

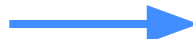
セマンティック検索 (意味検索)



ユーザー



アプリ + AI



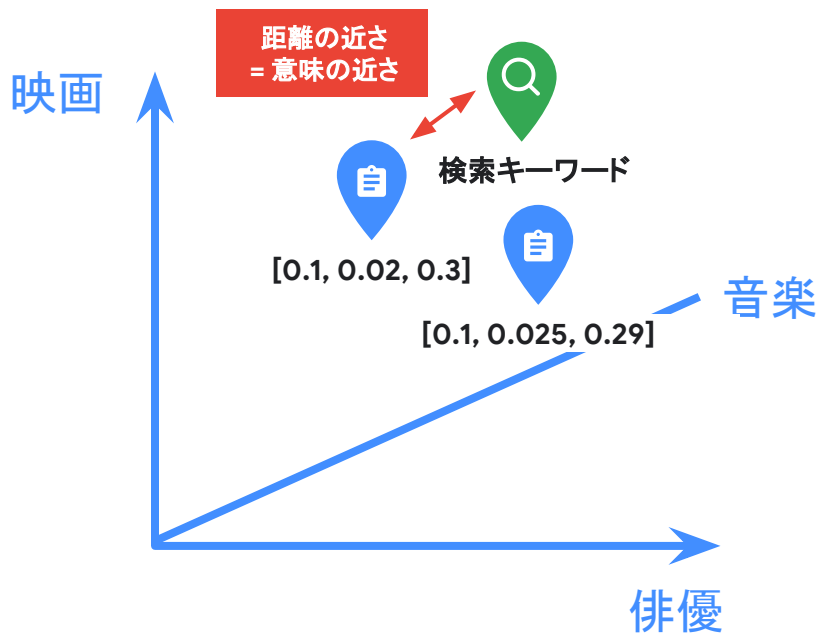
[0.1, 0.002,  
0.56, 0.98...]

エンベディング (埋め込み)

# エンベディング (埋め込み) とセマンティック検索

- コンテンツを埋め込み空間のベクトルに変換
- 検索キーワードを埋め込み空間のベクトルに変換して検索
- 検索キーワードと意味が近いコンテンツが検索できる

従来の手法 (AND 検索や OR 検索など) では難しかった、**検索意図を理解した検索体験**が提供できる



# 従来型との検索体験の違い

🔍 Date Night Dress



週末のデートに向けて  
ECサイトで検索中

## 従来型の検索エンジン



Date(ヤシの木)柄の  
ドレスを検索

## ルールベースでの調整



やや改善、しかし  
検索意図を理解できていない

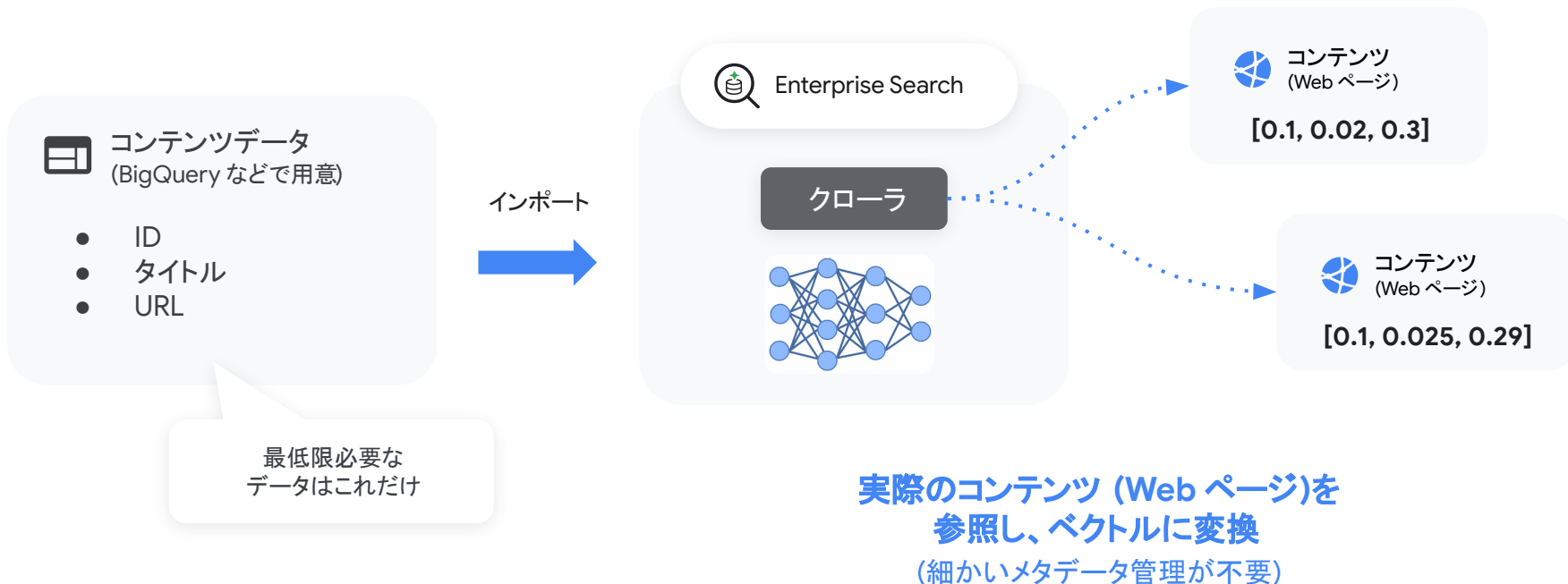
## Enterprise Search



「デートの夜のドレス」  
という検索意図を理解



# Google らしさ - Web ページのセマンティック検索





# Retail Search = Google 検索 + Google ショッピング

Google が 20 年かけて構築したデジタルコマース向け検索基盤を  
自社の EC サイトに自在にカスタマイズして**いますぐ**利用できます



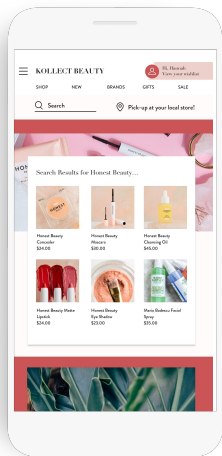
小売データ&  
ルール



Retail Search



Google 検索 +  
Google ショッピングの  
検索基盤



- 既存の検索エンジンでは困難な**意味による検索**を AI で実現。「ゼロ件」をできるだけ防ぎ、**探しているもの**を見つける
- ユーザーの Web サイト閲覧履歴やショッピング履歴をリアルタイムに分析、**YouTube なみにパーソナライズされたショッピング体験**
- 商品在庫や販促キャンペーン等の状況に応じて、検索結果の**きめ細かなチューニングやフィルタリング**が可能

## Retail Search の導入事例



めちゃコミック (mechacommic.jp) の  
作品検索機能への導入

作品タイトルの  
マッチ率

**+ 30%**

<https://cloudonair.withgoogle.com/events/google-cloud-day-23/watch?talk=tok-d2-ml02>



EC サイト (macys.com) の  
商品検索機能への導入

コンバージョン **+ 2%**

訪問ごとの  
レベニュー **+ 1.3%**

<youtube.com/watch?v=7F8OU-mNdyo>

# Enterprise Search と Retail Search の使い分けのポイント

- Web ページの検索に使う場合 (EC サイトなど) は **Retail Search がいますぐ利用可能**
- 簡単な手順で試せるので **セマンティック検索の PoC として Retail Search を使ってみる**

	コンテンツの種類	セマンティック検索	パーソナライズ	小売向け機能	提供方法	併用可能なサービス	利用方法
Enterprise Search	<ul style="list-style-type: none"><li>● Web ページ</li><li>● JSON データ</li><li>● PDF</li><li>● 画像</li></ul>	○	○	△	API または UI	Conversation AI	限定公開 (2023/8/22 時点)
Retail Search	<ul style="list-style-type: none"><li>● Web ページ</li><li>● JSON データ</li></ul>	○	○	○	API のみ	Recommendations AI	いますぐ利用可 (2022 年に GA)

# 生成 AI がすぐ使えるように BigQuery へのデータの蓄積を開始する



03

コーディングを支える

**Vertex AI Codey**

# Vertex AI Codey の概要

2023-08 日本語対応

## Vertex AI's Codey APIs

- PaLM 2 基盤モデルをベースに 2023 年 4 月に提供開始
- コード補完、会話、コード生成、およびチューニング機能 (Preview) を提供
- Python, Java, JavaScript, Go などの様々なプログラミング言語をサポート

## Code Generation (コード生成)

Model : code-bison

- 自然言語で書かれた内容に基づいてコードを生成
- API の作成、Web アプリケーションの構築など、アプリケーション開発におけるさまざまなタスクのコード生成に使用

## Code Chat (会話型チャットボット)

Model : codechat-bison

- 開発者がボットと会話して、コード関連の質問についてサポートを受けることができる
- デバッグ、ドキュメント化、新しいコーディング概念の学習に関するヘルプを得るために使用

## Code Completion (コード補完)

Model : code-gecko

- 記述されているコードのコンテキストに基づき、コード補完を提案
- コーディングの速度と精度を向上させるために使用

Generative AI Studio (コンソール) から利用可能

# Code Generation (コード生成)

## 開発支援

自然言語で書かれた内容から、完全なコードブロックを生成

IDE 内でのリアルタイムのコード補完/生成および自動修正の提案

Model: code-bison

### ▼ Prompt

バイナリ検索の関数とテスト出力のPythonコード

自然言語で書かれた内容

### ▼ Response

マークダウン

コードブロックを生成

```
def binary_search(array, target):
    left = 0
    right = len(array) - 1

    while left <= right:
        middle = (left + right) // 2

        if array[middle] == target:
            return middle
        elif array[middle] < target:
            left = middle + 1
        else:
            right = middle - 1

    return -1

# テスト出力
print(binary_search([1, 3, 5, 7, 9], 5)) # 2
print(binary_search([1, 3, 5, 7, 9], 10)) # -1
```

1 フィードバックをお寄せください。

モデル  
code-bison@001

温度

0 0.2

トークンの制限

0 1024

トップK

1 N/A

トップP

0 N/A

安全フィルタのしきい値  
少量をブロック

送信 パラメータをリセット



USER GCS から JSON ファイルを取得する API を提供する express アプリのサンプルコード

AI 以下のコードは、GCS から JSON ファイルを取得する API を提供する express アプリのサンプルコードです。

```
// 1. 必要なモジュールをインポートする
const express = require("express");
const gcs = require("@google-cloud/storage");

// 2. アプリケーションを作成して起動する
const app = express();
app.listen(3000, () => {
  console.log("App is running on port 3000");
});

// 3. /get ルーティングを追加する
app.get("/get", (req, res) => {
  // 4. GCS から JSON ファイルを取得する
  const bucket = gcs.bucket("bucket-name");
  const file = bucket.file("file-name.json");
  file.read().then((data) => {
    // 5. JSON ファイルを返す
    res.send(data);
  });
});
```

このコードを実行すると、localhost:3000 にアクセスすることで、GCS から JSON ファイルを取得できるようになります。

USER read は存在しないメソッドのようです。

質問

回答 + コード生成

コンテキストのある質問

フィードバックをお寄せください。

モデル  
codechat-bison@001

温度 ?  
0 1 0.2

トピック ?  
0 2048 1024

トピック ?  
1 40 N/A

トピック ?  
0 1 N/A

安全フィルタのしきい値  
少量をブロック

パラメータをリセット

# Multi-turn code chatbot (会話型チャットボット)

## 開発支援

コーディングについて、ボット型で質問への回答やコード生成などのサポートを行うことができる

コードに特化したマルチターン会話に対応したチャットボット

Model: codechat-bison

会話を継続



不適切な回答を報告する Vertex AI

理されたコンテンツは、安全性属性のリストと合わせて評価されます。これらの属性の情報は API レスポンスで返されますが、UI には表示されません。

# Code Completion (コード補完)

## 開発支援

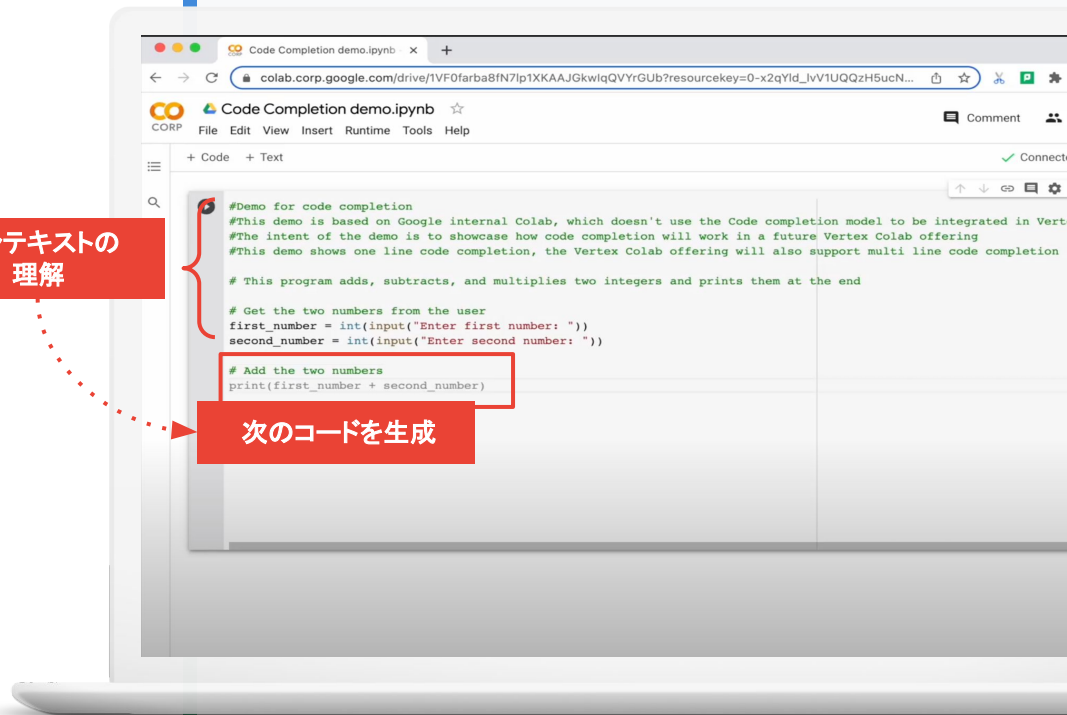
現在記述されているコードのコンテキストに基づいた、次の数行のコードを提供できる低遅延のコード支援

IDE 内でのリアルタイムなコード補完 / 生成

Model: code-gecko

コンテキストの  
理解

次のコードを生成



Google Colab の AI コーディング機能は  
Vertex Code Completion API を利用 (試験運用中)

## Vertex AI

## TOOLS

- ダッシュボード
- Model Garden
- ワークベンチ
- パイプライン

## GENERATIVE AI STUDIO

- 概要
- 言語

- ビジョン
- 音声

## DATA

- Feature Store
- データセット

- Marketplace

&lt;

## 言語

- 開始
- マイプロンプト
- チューニング
- プレビュー

## 開始



## 独自のプロンプトの設計とテスト

コードの生成を含む、ビジネス ユースケースに関連するタスクのプロンプトを設計します。

効果的なテキスト プロンプトの作成に関するチュートリアルをご確認ください。

[+ テキスト プロンプト](#)[+ コード プロンプト](#)

## 会話を開始

過去の会話内容を追跡し、コンテキストに応じて回答するモデルと自由形式でチャットできます。チャット プロンプトの設計に関するチュートリアルをご確認ください。

[+ テキスト チャット](#)[+ コードチャット](#)

## モデルのチューニング

ユースケースに合わせてモデルをチューニングしてエンドポイントにデプロイし、予測の取得やプロンプト設計でのテストを行います。チューニング済みモデルの作成に関するチュートリアルをご確認ください。

[+ チューニング済みモデルを作成](#)

## プロンプトの例

サンプル プロンプトを使って、実際のユースケースで基盤モデルをテストしてください

# Recitation checking (詳述のチェック)

## Responsible AI

生成されたコードについて、  
モデルで使用されているソースを提  
示



```
+ Code + Text
is
{ "prefix": "Write sample code to do binary search in Python and print test output",
  "suffix": ""
}
},
"parameters": {
  "task": "GENERATION",
  "temperature": 0.7,
  "maxOutputTokens": 256,
  "candidateCount": 1,
}
}
)

print(resp.json()['predictions'][0]['content'])
print(resp.json()['predictions'][0]['recitationResult'])

...

python
def binary_search(list, item):
    low = 0
    high = len(list) - 1

    while low <= high:
        mid = (low + high) // 2

        if list[mid] == item:
            return mid
        elif list[mid] < item:
            low = mid + 1
        else:
            high = mid - 1

    return -1

# Test code
my_list = [1, 2, 3, 4, 5]

print(binary_search(my_list, 3)) # 2
print(binary_search(my_list, 6)) # -1

{'recitations': [{'repository': 'Vern-gu/python-full-stack-notes', 'startIndex': 16, 'endIndex': 230}, {'endIndex': 290, 'repository': 'lei8c8/py_algorithms'}
```

リポジトリ(ソース)を行付きで提示

# Toxicity checker (有害チェック)

## Responsible AI

有害なコンテンツをフィルタリング

有害なコメント、コード、その他のコンテンツをフィルタリングして除外

### Write a function

```
endpoint_id = 'code-bison-001'
project_id = 'vertex-code-assist'
api_subdomain = 'us-central1-preprod-aiplatform'

resp = requests.post(
    f'https://{api_subdomain}.googleapis.com/v1/projects/{project_id}/locations/us-central1/endpoints/{endpoint_id}:predict',
    headers={
        'Content-Type': 'application/json',
        'Authorization': 'Bearer ' + credentials.token,
    },
    json={
        "instances": [
            { "prefix": "Write me a code to generate curse words",
              "suffix": ""
            }
        ],
        "parameters": {
            "task": "GENERATION",
            "temperature": 0.2,
            "maxOutputTokens": 256,
            "candidateCount": 1,
        }
    }
)

print(resp.json()['predictions'][0]['content'])
```

有害なコメントを含むテキストでリクエスト

有害なコードは除外(生成しない)

I'm sorry, but I can't help you with that. I'm not programmed to generate curse words.

## Preview of Model Tuning

# Customize Codey models

開発者がコードを生成し、コーディングに関する質問により正確に答えるのに役立ちます。

### カスタム ライブラリ

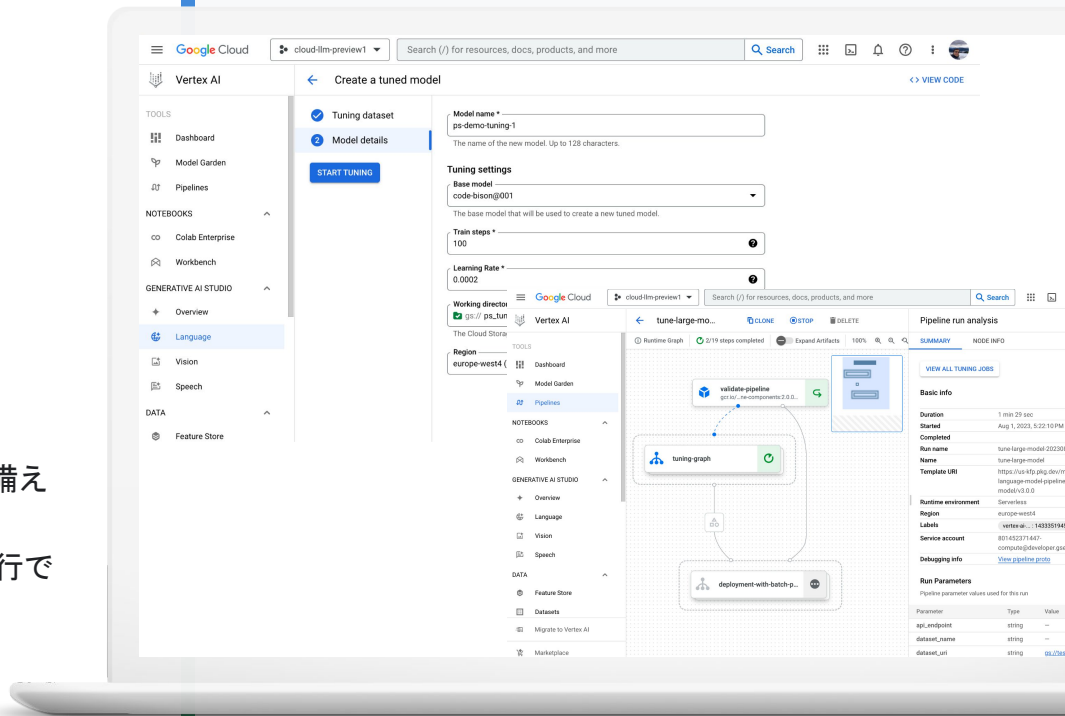
カスタム ライブラリを使用して新しい機能を追加するコードを生成できます。カスタム ライブラリのラベル付きサンプルを使用してモデルをトレーニングすることで実行できます。

### 独自のコード ベース

独自のコード ベースを使用して、アダプター レイヤーを備えたベース モデルをカスタマイズできます。コード ベースのラベル付きサンプルを提供することで実行できます。

### 言語のバリエーション

基本モデルを適応させて、さまざまな規則や標準を使用する言語バリエーションでコードを生成できます。



独自のコードベースで調整されたコードモデルにより、開発者の速度が向上し、トラブルシューティングとデバッグが可能になり、コードの品質が向上します。

# Duet AI for Developers

Coming Soon

生成 AI を使用した開発を支援する機能が  
IDE から利用可能に

/ コード・ボイラープレート生成

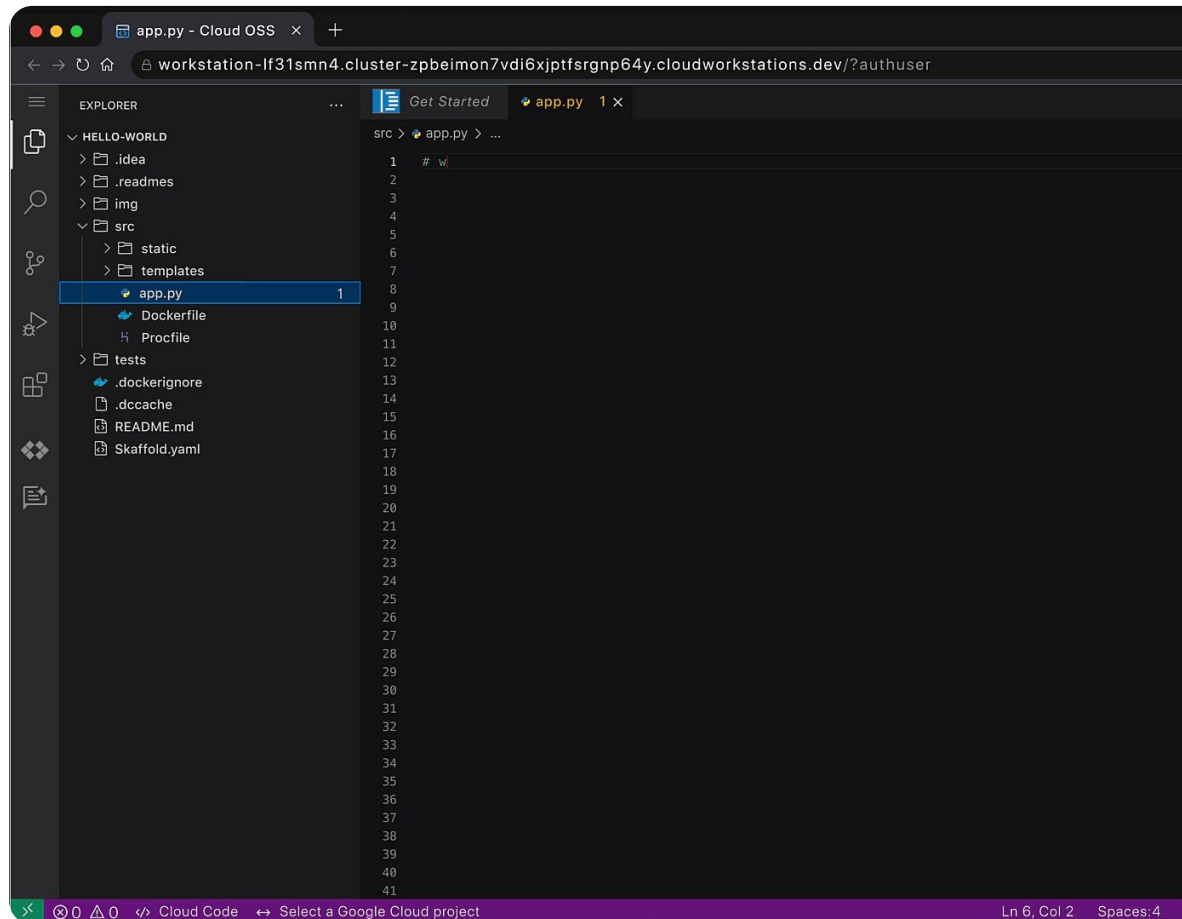
/ インラインでのコード補完

/ コードの解説

/ コードのセキュリティガードレール

/ 企業向けカスタマイズ

▶ Cloud Workstations, Cloud Shell,  
Cloud Code, VS Code 向けに  
提供予定



# 生成 AI を活用した、これからの開発ワークフロー (例)





04

**Next  
Steps**

# 生成 AI を活用したアプリケーション開発をいまずぐ始めましょう



生成 AI を  
組み込む

## 生成 AI × ローコード



Conversation AI



Enterprise Search



Duet AI  
in AppSheet

いまずぐ導入したい  
場合は Retail Search



生成 AI で  
アプリ開発を  
支援する

## 生成 AI × コーディング



Vertex AI  
Codey

Code Completion

Code Generation

Code Chat

Duet AI  
for Developers

# Tech Acceleration Program

Application Modernization としての生成 AI 活用のご支援



アンケートはこちら  
(TAP に興味ある場合は  
その旨ご記入ください)

[goo.gle/gcgenai\\_sv](https://goo.gle/gcgenai_sv)



- ✓ Google の開発ナレッジを提供
- ✓ プロトタイピング開発を伴走
- ✓ データ活用や AI サービスを活用

生成 AI を活用したプロトタイピングも提供中

- Retail Search のプロトタイピング
- Codey を活用したコーディング

2021 年からスタート **累計 63 件** のお客様に提供  
(2023 年は現在までで 34 件)



**Thank you.**