

Ok Gemini, what's for dinner?

Maia Grotepass - @maiatoday
Ivan Pletinskij- @ivan.pletinskij



Outline

- Intro - Who are we, why this talk (Ivan + Maia)
- Problem definition (Maia)
- Possible tech options (Maia)
 - Google ai studio vs Vertex
 - Android vs kmp
 - AS vs Fleet vs IntelliJ
 - Gradle vs Amper
 - The problem with ai studio in europe
 - Will it work? Test on the dashboard
- The app (Ivan)
 - Architecture?
 - Sample code?
 - Repo?
 - Gotchas and tricks?
- The demo (Ivan and Maia)
- The experience of building it (Ivan)
- How much did it cost? (Maia)
- Wrap up, conclusion, next steps (Maia + Ivan)



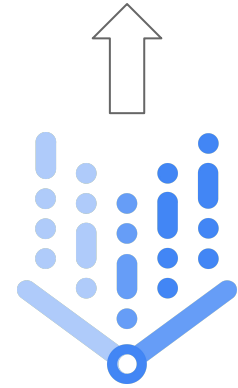


Gemini 1.5 Flash Gemini 1.5 Pro Gemini 1.0 Pro





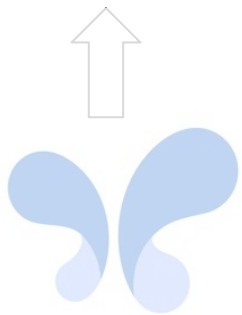
Gemini 1.5 Flash Gemini 1.5 Pro Gemini 1.0 Pro



REST



Gemini 1.5 Flash Gemini 1.5 Pro Gemini 1.0 Pro



easy

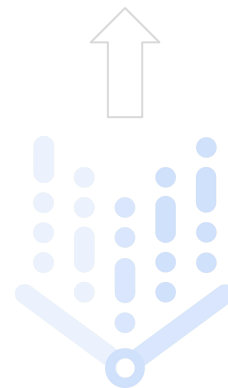
aistudio.google.com

Android Studio
wizard

SDK Android



SDK



REST



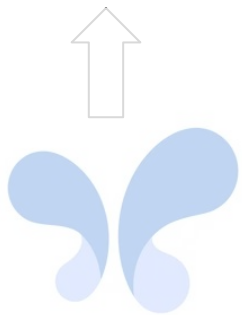
Gemini

Gemini Nano

- Pixel 8 and Samsung \$\$\$
- Currently only for early access



Gemini 1.5 Flash Gemini 1.5 Pro Gemini 1.0 Pro



easy

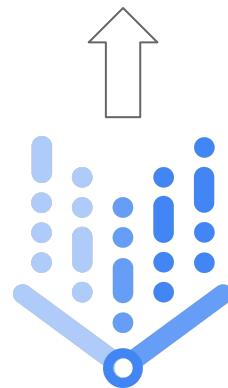
aistudio.google.com

Android Studio
wizard

SDK Android



SDK



REST

Marcel Pintó



Three ways to use AI on Android: The Good, the Bad and the Ugly

Artificial Intelligence is revolutionizing the world, more and more apps are integrating AI, but are we doing it the right way?

In this talk, I will explore practical strategies for using AI responsibly in your Android apps. We will start with the “Ugly”, by showcasing a basic implementation and exploring the consequences. We will move forwards by improving our strategy but exposing the “Bad” pitfalls and leaks we could face. To finish, I will walk you through the “Good” strategy to craft secure prompts, leakproof API keys, and build robust architectures for cost-optimized AI integration in your Android apps.

THINGS

Session

Advanced

Watch the recording!

Will it work? Let's test on the dashboard

The screenshot shows the Google Cloud Vertex AI console interface. The top navigation bar includes the Google Cloud logo, a search bar, and various utility icons. The left sidebar contains a 'TOOLS' section with links to Dashboard, Model Garden, and Pipelines, and a 'NOTEBOOKS' section with links to Colab Enterprise, Workbench, and Vertex AI Studio. The 'MULTIMODAL' section is highlighted, showing options for Language, Vision, and Speech. The main content area is titled 'Untitled prompt' and features a 'System instructions' field, a 'Prompt' field, and a 'Response' field. The 'Prompt' field contains a request for a recipe based on an image of an Italian dish, with a 'SUBMIT' button and a token count of 319. The 'Response' field displays the generated recipe: 'Easy Italian Pasta with Tomato Sauce', including a description and a list of ingredients. The right sidebar contains configuration options for the model (gemini-1.5-flash-001), region (us-central1), temperature, and output token limit, along with a 'SAFETY SETTINGS' section and a 'VIEW SERVICE DISCLAIMERS' link.

Google Cloud console interface showing a Vertex AI prompt and response.

System instructions:

Prompt:

Given the image of an **Italian** takeaway dish, write a recipe that is easy to prepare at home with clear instructions and include the following ingredients: **three ripe tomatoes**. Then provide grocery list of the additional ingredients with their quantities in **metric**, which is needed for the recipe. Exclude the ingredients I listed already.

Response:

Easy Italian Pasta with Tomato Sauce

This recipe is simple and delicious, perfect for a weeknight meal.

Ingredients:

- 3 ripe tomatoes
- 200g dried pasta (e.g., penne, rigatoni, or your favorite)
- 1 tbsp olive oil

Model: gemini-1.5-flash-001

Region: us-central1 (Iowa)

Temperature: 1

Output token limit: 8192

Add stop sequence: ?

Press Enter after each sequence

SAFETY SETTINGS

Advanced

RESET PARAMETERS

We want your [feedback](#).

[VIEW SERVICE DISCLAIMERS](#)

[REPORT INAPPROPRIATE RESPONSES](#)

Demo time









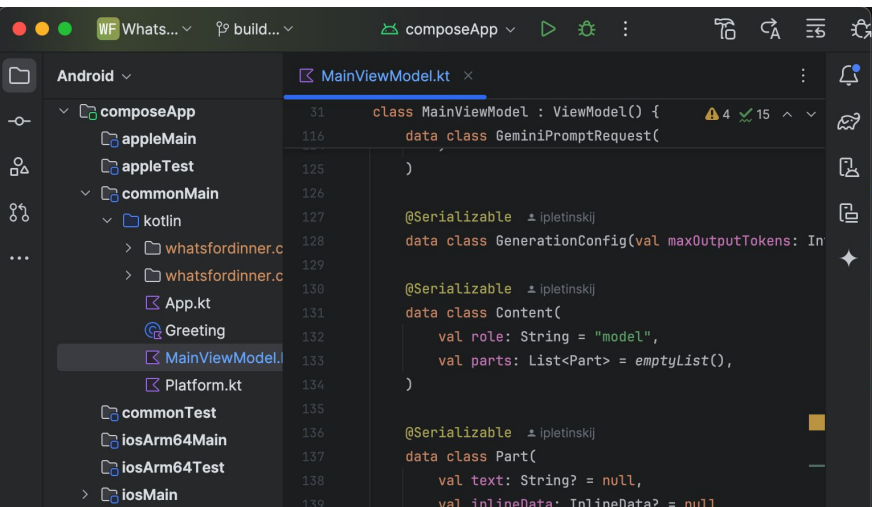




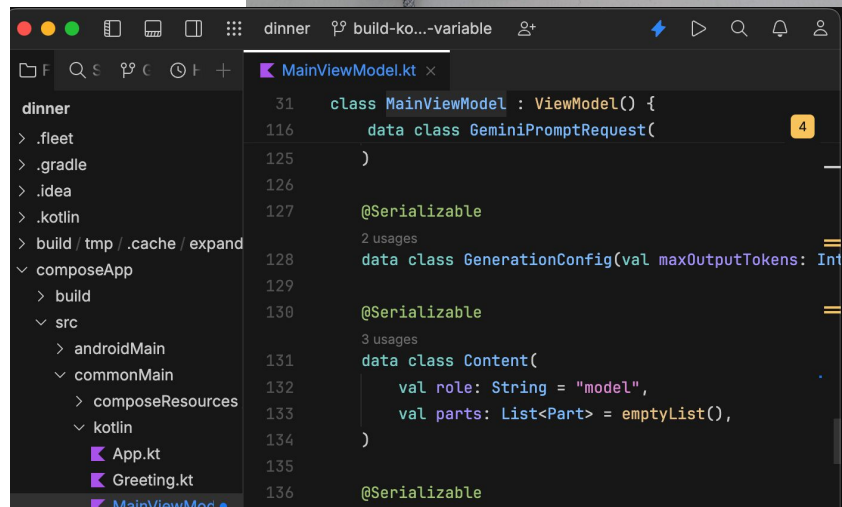
Backup demo video for if the live video fails

Tools and frameworks

Android Studio



Fleet



Architecture

Tech stack

- Kotlin Multiplatform
- Gradle Kotlin DSL
- Jetpack Compose
- Ktor
- Kotlin serialization
- Peekaboo
- BuildKonfig

App architecture

Share a piece of logic

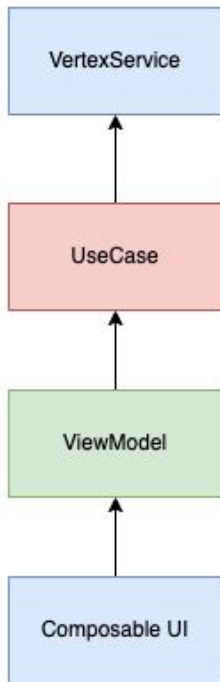
Improve your app's stability by sharing an isolated and critical part of the app. Reuse the Kotlin code you already have to keep the applications in sync.

UI

Logic



App architecture



```
▼ WhatsForDinner ~/Downloads/WhatsForDinner
> .fleet
> .gradle
> .idea
> .kotlin
> build
▼ composeApp
> build
▼ src
  ▼ androidMain [main]
    ▼ kotlin
      ▼ com.handen.whatsfordinner
        MainActivity.kt
      data
        CreateHttpClient.kt
      res
        AndroidManifest.xml
    commonMain
      ▼ kotlin [commonMain] sources root
        data
          dto
            CreateHttpClient.kt
            VertexService
          domain
            GetGeminiRecipeUseCase
          ui
            MainScreen.kt
            MainViewModel
        iosMain
          ▼ kotlin
            data
              CreateHttpClient.kt
              MainViewController.kt
        build.gradle.kts
  gradle
  iosApp
    .gitignore
    build.gradle.kts
    gradle.properties
    gradlew
    gradlew.bat
    local.properties
    README.md
    settings.gradle.kts
```

Code walkthrough

Gotchas and tips and tricks

- Use Vertex dashboard and CLI to debug prompts and requests
- Use online Base64 converters to debug image encoding
- K2 Compiler feels really faster
- To write DTO schema I used “Get Code” function with curl selected on Google Cloud

Tokens and characters

Token

Will it fit into the question?

312

Characters

Billing

In 300

Out 1227

Billing

Billing account
My Billing Account 1

Overview

Cost management

Reports

Cost table

Cost breakdown

Budgets & alerts

Billing export

Cost optimisation

FinOps hub

Committed-use discounts...

CUD analysis

Pricing

Reports

GENERATE QUERY PRINT SHARE SAVE VIEW LEARN

Saved views

Period	Total Cost	Includes Credits
1-15 June 2024 (total cost)	US\$0.00	includes US\$0.00 in credits
June 2024 (forecasted total cost)	US\$0.01	includes US\$0.00 in credits

1-15 June 2024 (total cost) ?
US\$0.00
includes US\$0.00 in credits

June 2024 (forecasted total cost) ?
US\$0.01
includes US\$0.00 in credits

— ?
US\$0.00 over 17-31 May 2024

— ?
US\$0.00 over May 2024

Show cumulative

June 21, 2024

\$1

\$0

1 Jun 2 Jun 3 Jun 4 Jun 5 Jun 6 Jun 7 Jun 8 Jun 9 Jun 11 Jun 13 Jun 15 Jun 17 Jun 19 Jun 21 Jun 23 Jun 25 Jun 27 Jun 29 Jun

Billing account —
My Billing Account 1

Overview

Cost management

Reports

 Cost table

 Cost breakdown

 Budgets & alerts

 Billing export

Cost optimisation

FinOps hub

- Committed-use discounts...

% CUD analysis

 Pricing

Reports

 GENERATE QUERY PRINT SAVE VIEW

Saved views

1-15 June 2024 (total cost) ?

US\$0.00

includes US\$0.00 in credits

— ?

US\$0.00 over 17–31 May 2024

June 2024 (forecasted total cost) ?

US\$0.01

includes US\$0.00 in credits

— ?

US\$0.00 over May 2024

☐ Show cumulative

June 21, 2024

\$1

\$0

1 Jun 2 Jun 3 Jun 4 Jun 5 Jun 6 Jun 7 Jun 8 Jun 9 Jun 11 Jun 13 Jun 15 Jun 17 Jun 19 Jun 21 Jun 23 Jun 25 Jun 27 Jun 29 Jun

Model	Feature	Type	Price (≤ 128K context window)	Price (> 128K context window)
Gemini 1.5 Flash	Multimodal	Image Input	\$0.0001315 / image	\$0.000263 / image
		Video Input	\$0.0001315 / second	\$0.000263 / second
		Text Input	\$0.000125 / 1k characters	\$0.00025 / 1k characters
		Audio Input	\$0.0000125 / second	\$0.000025 / second
		Text Output	\$0.000375 / 1k characters	\$0.00075 / 1k characters

Image + input text + output text = $0.0001315 + 0.000125 + (0.000375 \times 2)$

USD 0.0010065 per call

10USD = 9935 calls

Ivan+Maia one call each every day of the year

10USD \sim 13 years

Summary

- server-side LLMs **dominate**
- Prompt **packaged** in an app
- **easy** to to build multi platform apps with Vertex AI

Next

for us

- Leverage Google **Cloud functions** to make it possible to adjust prompts on the BE side
- Use **Amper** to configure build scripts
- Make a **desktop** version?
- Better **UI**

Next - for you

<https://github.com/RuggedOdyssey/dinner>

Links for next steps

[AI Studio](#)

[Vertex AI for firebase](#)

[Vertex AI pricing](#)

[Vertex models](#)



Maia Grotepass
@maiatoday
Ivan Pletinskij
@ivan.pletinskij

