inertGo

Student User Manual

Your AI-powered study companion

Table of Contents

A Note to Students

Education today faces a paradox. Families invest significant amounts in tuition, coaching centres, and study materials, yet millions of students still struggle to get timely, personalised guidance when they need it most – during that late-night revision session, while solving a difficult problem set, or when a concept just does not click.

The gap is not in effort or intent. It is in **access**. Access to a patient tutor who can explain the same concept five different ways. Access to instant feedback on whether your approach is correct. Access to structured practice that adapts to your pace.

inertGo was built to bridge this gap.

For a fraction of the cost of a single tuition session, inertGo puts an AI-powered study companion in your pocket. It gives you:

- **Instant explanations** – ask any question, any time, and get a step-by-step answer.
- **Complete freedom to learn at your own pace** – no fixed schedule, no waiting for a teacher to be available.
- **Structured daily tests** that build consistency and track your progress over time.
- **A Writing Pad** to show your working, sketch diagrams, and get feedback on your handwritten solutions.
- **On-device AI** – groundbreaking technology that runs AI models directly on your phone, so you can learn even without internet, completely free and private.
- **Choice of AI models** – from fast, lightweight models for quick lookups to powerful reasoning models for complex problems, you pick what works best for you.

This manual will walk you through every feature of the app. Whether you are using AI tools for the first time or switching from another platform, this guide will help you get the most out of inertGo.

Your learning, your pace, your companion.

1. Getting Started

1.1 What This App Does

inertGo lets you chat with an AI tutor, run daily tests, and write or attach content (text, images, and Writing Pad notes). It is designed to feel like a study companion with structured analysis and fast access to learning tools.

1.2 Key Concepts (Plain Language)

- **LLM (Large Language Model):** The technology behind the AI. You do not need to understand how it works – just know that it reads your question and writes a reply.
- **AI Helper:** The part of the app that answers your questions using an LLM.
- **Model:** The type of AI helper you pick. Some are faster, some are more detailed.
- **Provider:** The company or service that hosts a model (for example, OpenAI, Google Gemini, or a local server).
- **On-Device Model:** An AI model that runs directly on your phone – no internet needed, completely free and private.
- **Analysis:** A more detailed explanation mode for tricky questions.

1.3 Main Areas You Will Use

The app has four main tabs in the bottom navigation bar:

Tab	What It Does
Home	Browse subjects, chapters, questions, and chat with AI
Tests	Take daily tests, review results, and get AI analysis
Dashboard	View your learning analytics and conversation history
Settings	Configure your profile, LLM providers, and preferences

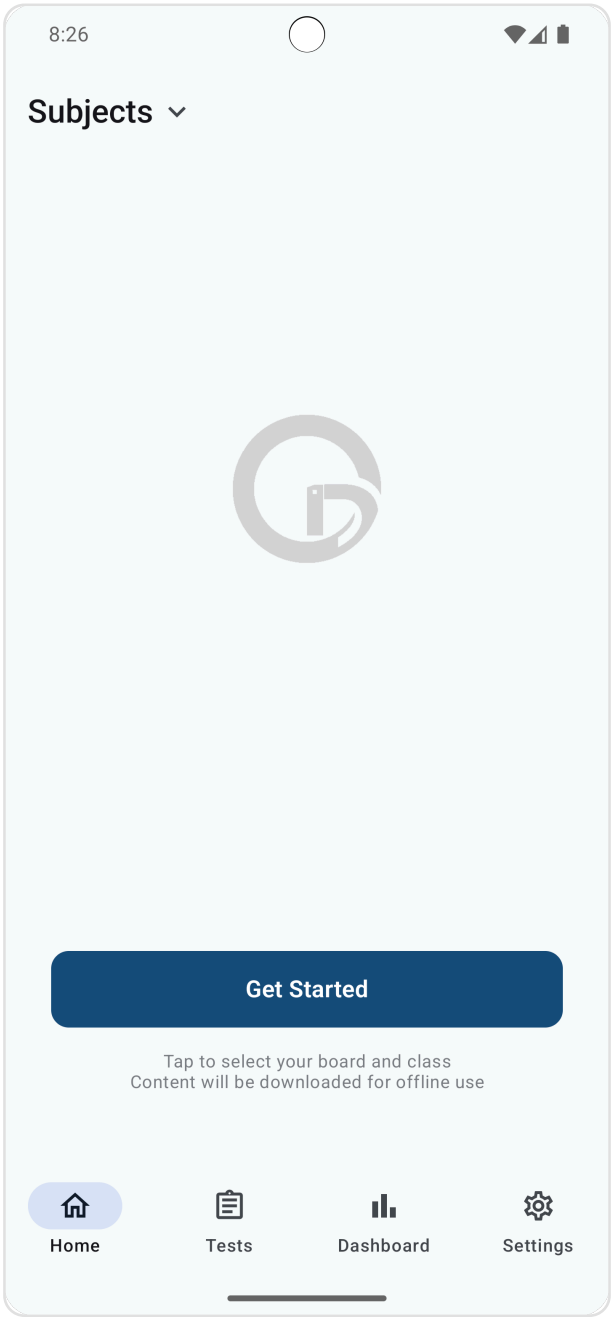

Figure 1: Figure 1.1 – Home screen (initial state with Get Started prompt)

2. Quick Start

Follow these steps the first time you open the app.

1. Open inertGo.
2. Sign in with your Google account.
3. If prompted, grant camera, microphone, and storage permissions.
4. Select your board and class when prompted (for example, CBSE Class 12).
5. Go to **Settings > LLM Providers** and select a provider. Enter your API key if asked (your teacher or admin can give you one). Or download an on-device model for free, offline AI.
6. Return to the home screen and open a chapter to browse questions.
7. Open a question and type a message in the chat area. Tap **Send**. The AI will reply in a few seconds.
8. To take a Daily Test, tap **Tests** in the bottom navigation and choose today's test.

You are ready to go. The sections below explain each feature in detail.

3. First-Time Setup

3.1 Login

When you first open the app, tap **Continue with Google** to sign in.

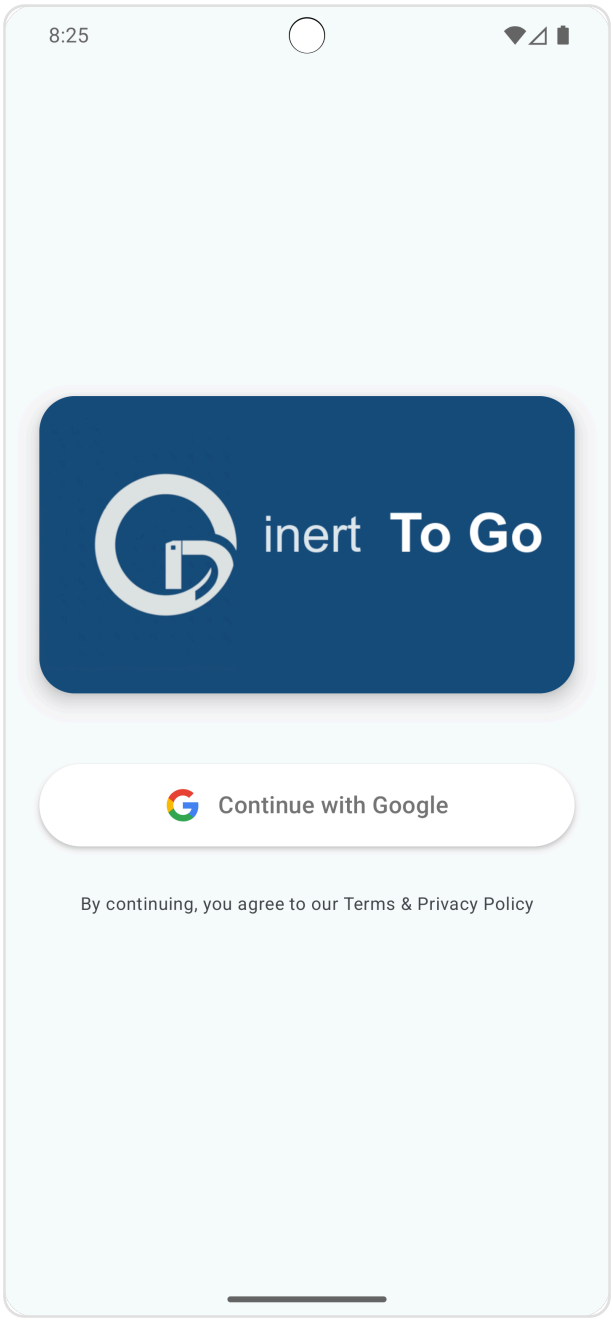

Figure 2: Figure 3.1 – Login screen

3.2 Permissions

The app may ask for the following permissions:

- **Camera:** Required to capture images for attachments.
- **Microphone:** Required for voice input (if available).

- **Storage:** Required to save and attach images.

Grant these when prompted, or enable them later in your device settings.

3.3 Profile Configuration

After login, select your board and class. This allows the app to download relevant content for your subjects.

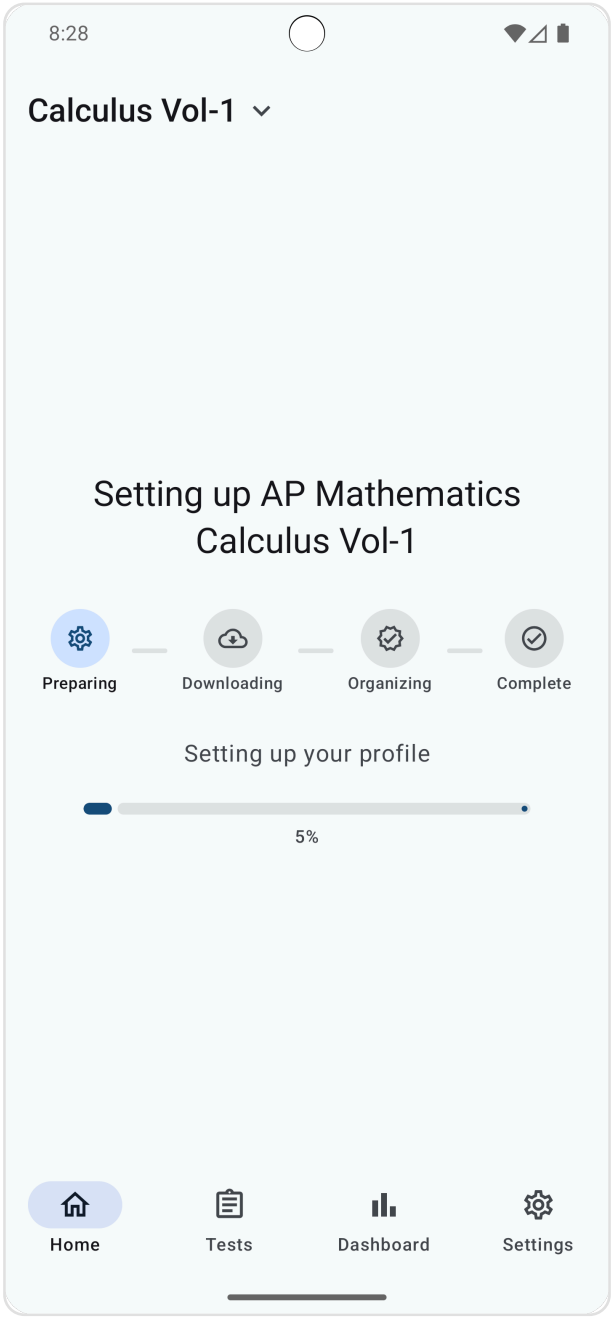

Figure 3: Figure 3.2 – Profile configuration screen

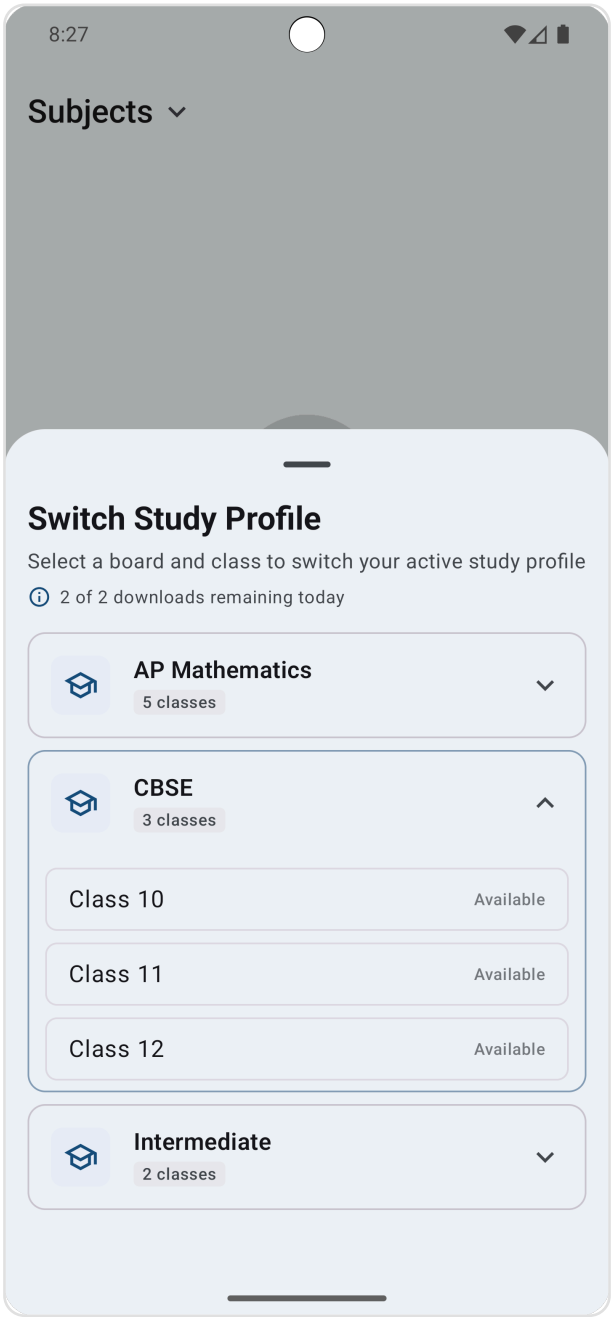

Figure 4: Figure 3.3 – Selecting a board and class from the available options

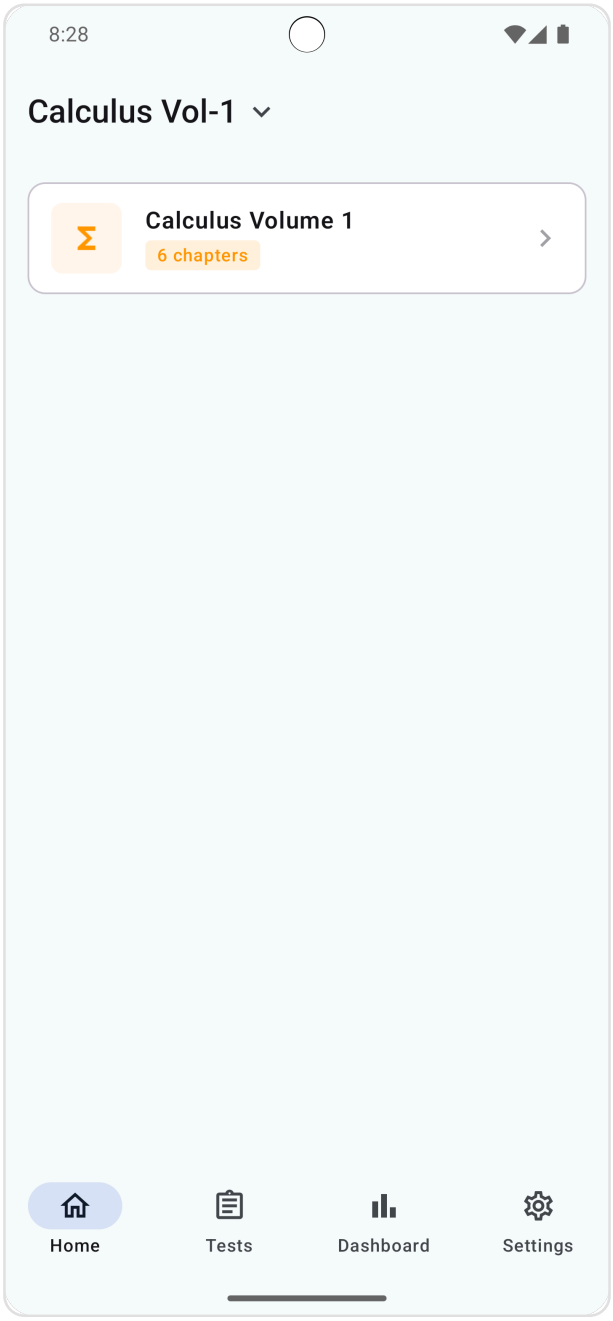

Figure 5: Figure 3.4 – Home screen after profile configuration, showing your subjects

3.4 Switching Profiles

You can switch to a different board or class at any time from the profile selector on the home screen.

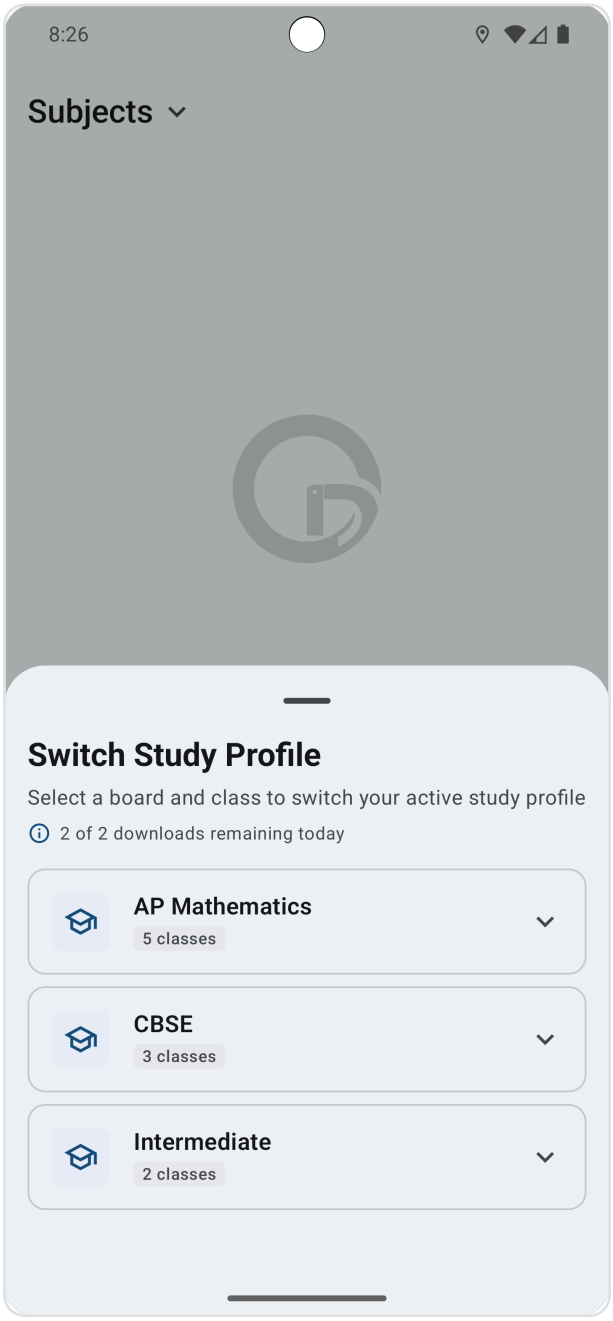

Figure 6: Figure 3.5 – Profile selector (collapsed) on the home screen

4. Home Tab – Browsing and Practicing Questions

4.1 What This Tab Shows

The Home tab is your main study area. It shows your subjects, chapters, and questions. Each question has its own conversation area where you can chat with the AI.

4.2 Browsing Questions

1. From the Home tab, tap a subject card to see its chapters.
2. Tap a chapter to see the list of questions.
3. Tap a question to open it with the full question text and a conversation area below.

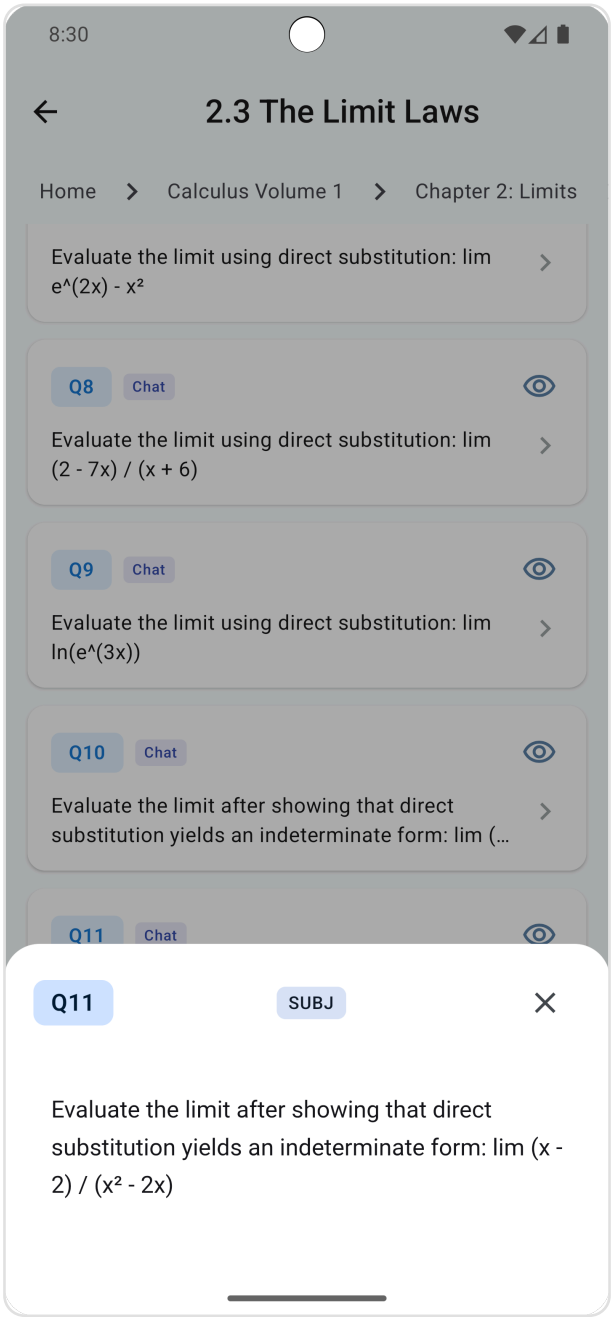

Figure 7: Figure 4.1 – Question list with a question preview at the bottom

4.3 Chatting with the AI on a Question

Once you open a question, you will see the question text at the top and a conversation area below. This is where you interact with the AI.

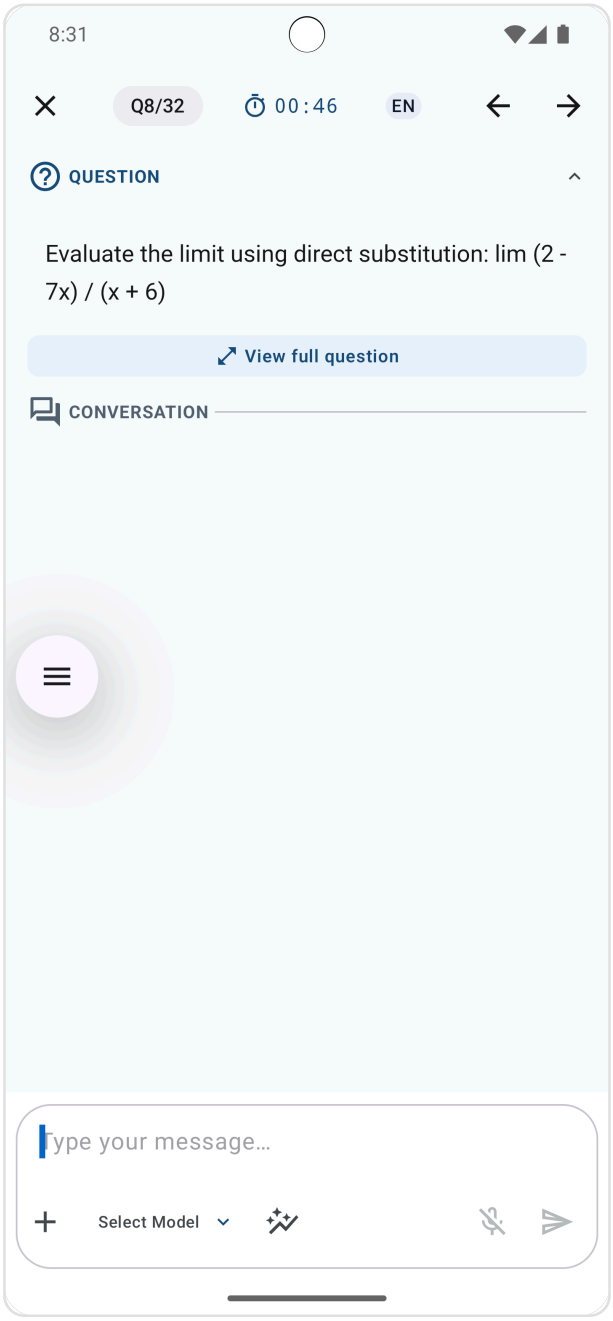

Figure 8: Figure 4.2 – Question screen with conversation area and compose bar

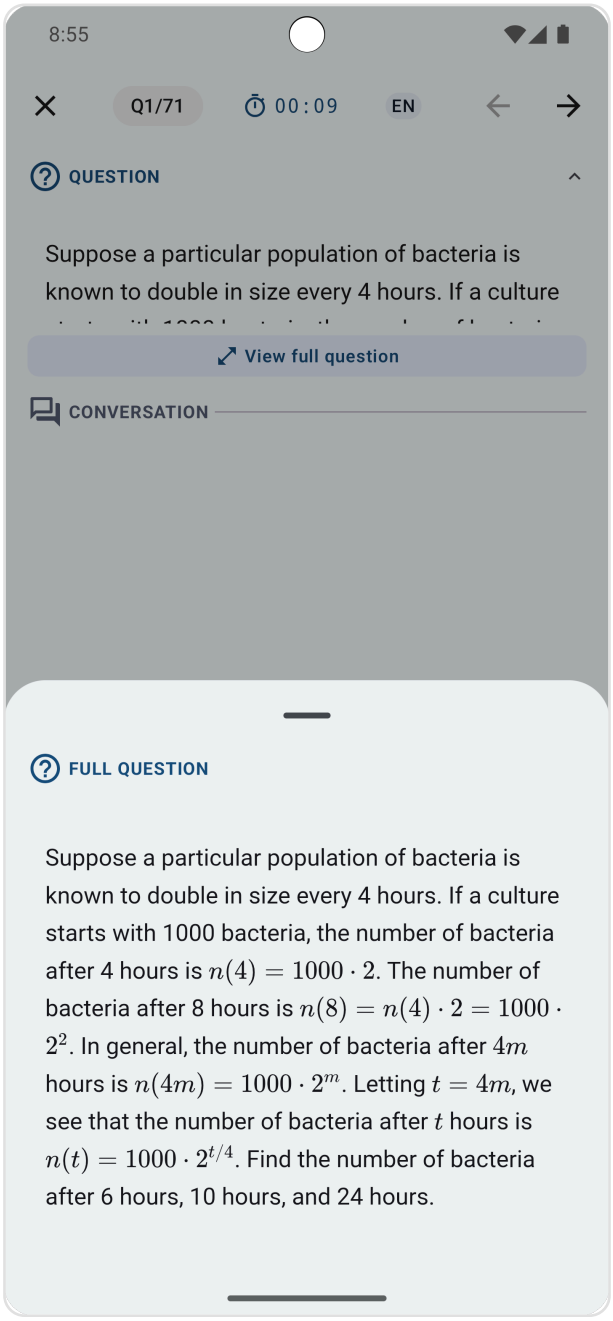

Figure 9: Figure 4.3 – Full question view during a conversation

4.4 How to Ask Good Questions

- Be specific: “Explain photosynthesis step by step” works better than “tell me about plants.”
- If the answer is wrong or unclear, ask a follow-up: “Can you explain step 2 again?”
- For math, ask the AI to explain each step and check the answer.

4.5 Chat Tools

- **Typing indicator:** Shows while the AI is preparing a reply.
- **Model picker:** Lets you switch models from the compose bar (see Section 5).
- **Voice input:** Hold the mic button to speak instead of typing (if available).

4.6 Language Switcher – Get AI Replies in Your Language

inertGo supports 25+ languages across four regions. You can ask your question in English and have the AI reply in your preferred language – or switch languages at any time during a conversation.

How to Switch Languages

1. On the question screen or during a test, tap the **language chip** in the toolbar (it shows your current language, for example **EN** for English).
2. A bottom sheet opens with the title **Select Language**.
3. Use the search bar to find a language quickly, or scroll through the grouped list.
4. Tap a language to select it. The AI will reply in that language from your next message onward.

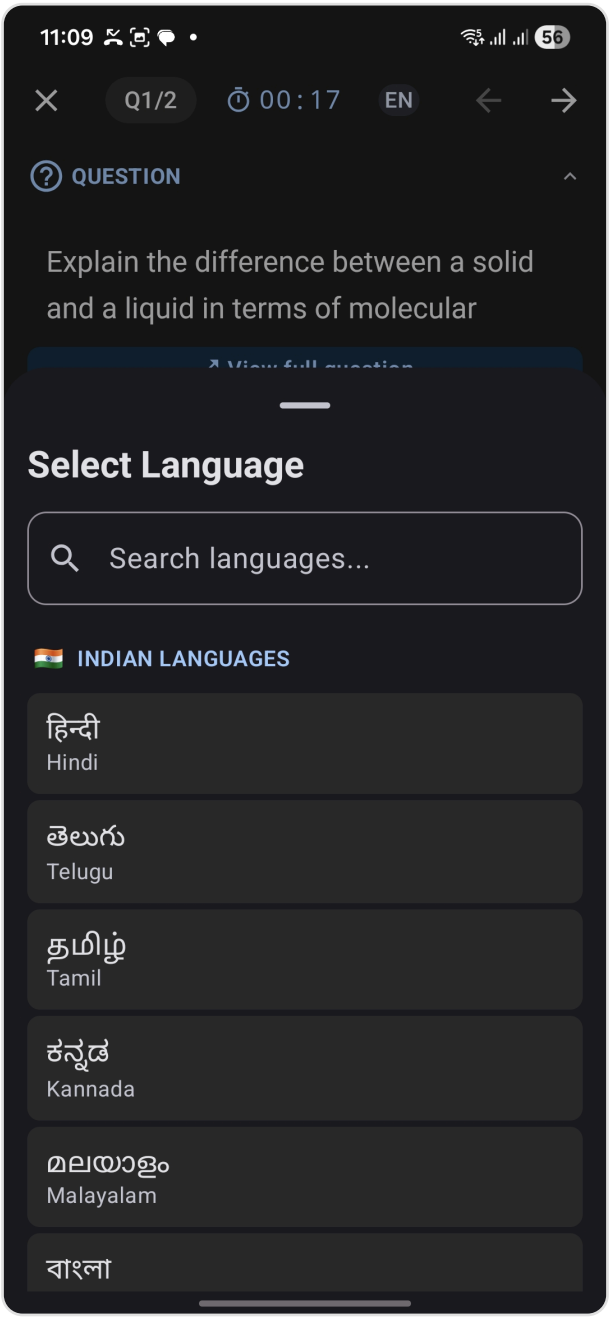

Figure 10: Figure 4.4 – Language switcher showing Indian languages (Hindi, Telugu, Tamil, Kannada, Malayalam, Bengali, and more)

Supported Languages

Region	Languages
Indian	Hindi, Telugu, Tamil, Kannada, Malayalam, Bengali, Marathi, Gujarati, Punjabi, Odia
European	English, Spanish, French, German, Portuguese, Italian, Dutch, Polish, Russian, Ukrainian
Middle East and Africa	Arabic, Turkish, Persian, Hebrew, Swahili
Asia Pacific	Korean, Indonesian, Vietnamese, Thai, Malay

Tips

- The language setting applies per conversation. You can use Hindi in one question and English in another.
- Technical terms (formulas, code, chemical names) are usually kept in English even when the rest of the reply is in your chosen language.
- If you study in a regional-medium school, set your preferred language once and all AI replies will follow.

4.7 Using the Writing Pad During Practice

The Writing Pad lets you sketch, draw, or write handwritten work and attach it to your conversation. This is especially useful for math problems, diagrams, and showing your working.

Opening the Writing Pad

1. In the compose area, tap the **+** button.
2. Choose **Writing Pad**.
3. Write or draw your notes.
4. Tap **Attach** to send your pages to the chat, or tap the save icon to keep them for later.

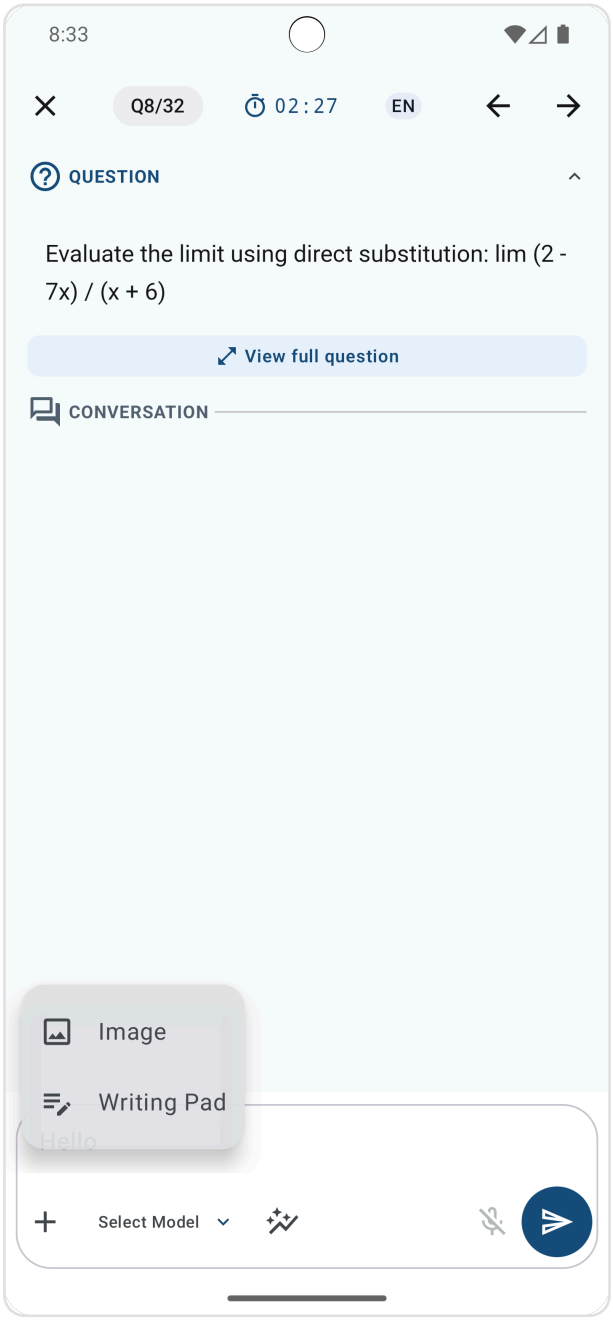

Figure 11: Figure 4.5 – Attachment menu showing Image and Writing Pad options

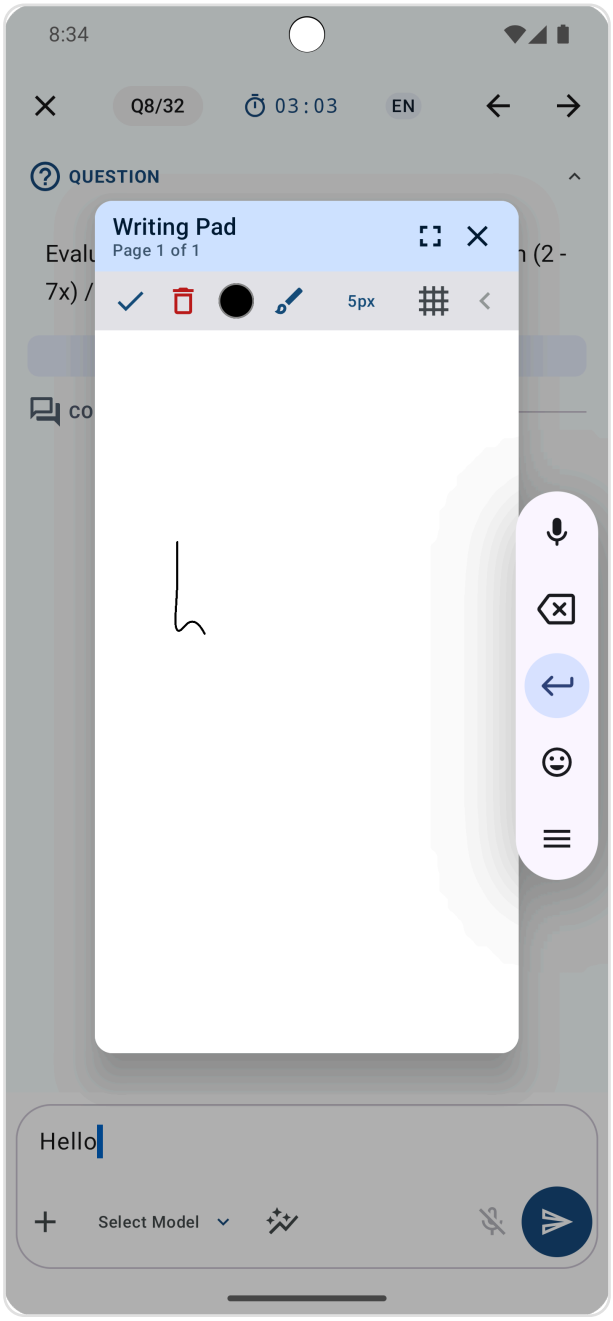

Figure 12: Figure 4.6 – Writing Pad in normal mode with drawing tools

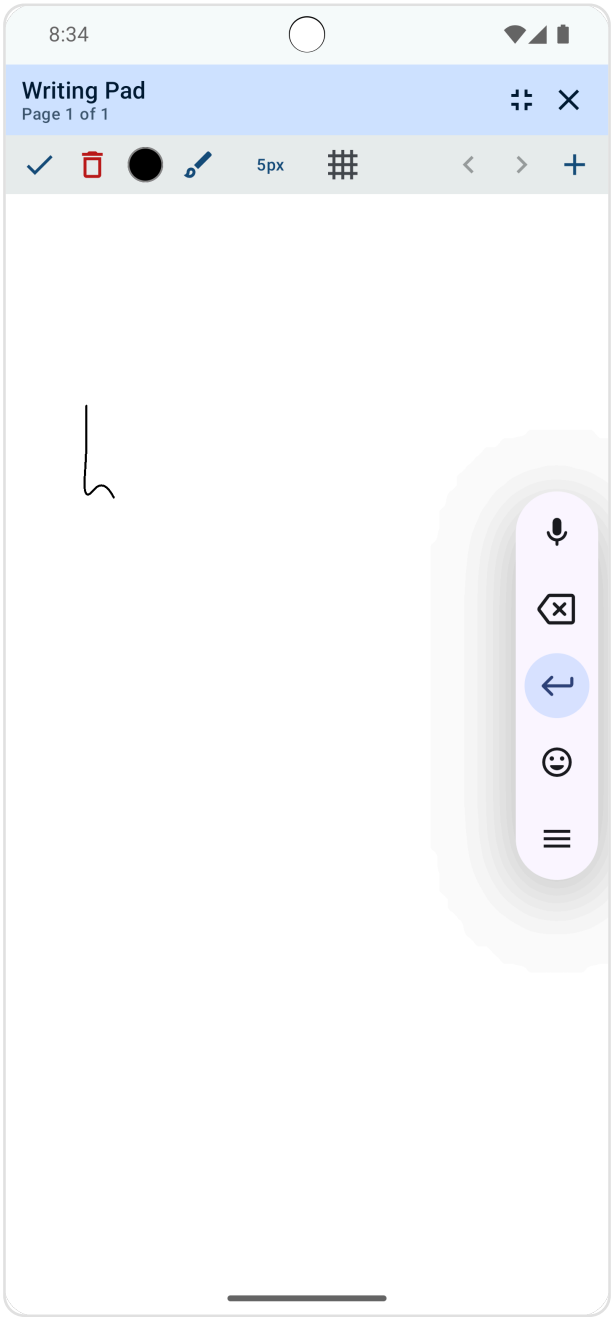

Figure 13: Figure 4.7 – Writing Pad in full-screen mode

Writing Pad Tools

Tool	What It Does
Attach	Sends your pages to the chat
Clear Page	Clears the current page
Color Picker	Choose pen color
Pen/Eraser	Switch between drawing and erasing
Stroke Width	Pick line thickness (for example, 4px or 10px)

Tool	What It Does
Background	Switch between Plain, Grid, or Ruled
Pages	Add up to 3 pages and switch with arrows
Maximize/Minimize	Toggle full-screen mode

Writing Pad Tips

- Keep one topic per page for easier reuse.
- Use multiple pages to organize steps (for example, one page per math problem).
- Your pages are treated as images, so use an image-capable model when attaching them.

4.8 Attaching Images During Practice

You can also attach photos of your notes, textbook pages, or handwritten work.

1. In the compose area, tap the **+** button.
2. Choose **Image**.
3. Pick an image from your gallery or capture a new one with the camera.
4. Confirm and send.

Note: Image attachments require an image-capable model. If images are not working, check your model selection in the compose bar.

5. Choosing the Right AI Model

5.1 Why Model Choice Matters

Different models have different strengths. Some are fast and good for simple questions. Others are more powerful and better for complex problems. Choosing the right model helps you get better answers and saves time.

5.2 Switching Models from the Compose Bar

1. Open a question's conversation area.
2. Tap the **Select Model** chip in the compose bar.
3. Choose a provider, then a model.
4. Type your message and send.

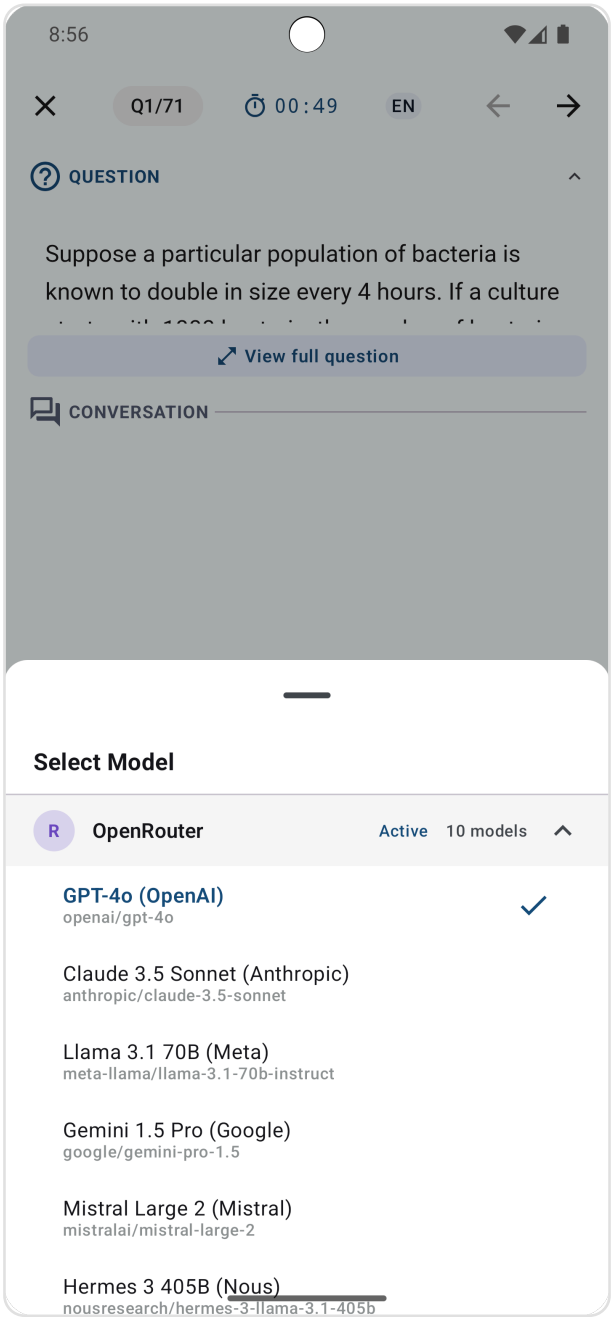

Figure 14: Figure 5.1 – Model selection bottom sheet showing available providers and models

5.3 Model Types

Text Models (Normal Chat)

- Used for regular questions and answers.
- Best for explanations, summaries, and step-by-step help.

Image Models

- Used when you attach images and want the AI to read them.
- Best for diagrams, handwritten work, or screenshots.
- Only shown if the provider supports images.

On-Device Models

- Run directly on your phone – no internet required.
- Completely free and private – your data never leaves your device.
- Best for quick help when you are offline or want privacy.
- See Section 6.6 for setup instructions.

Voice Models (If Available)

- Used for voice conversations or spoken answers.
- Only shown if the provider supports voice.

5.4 What Works With What

What You Are Sending	Model Type Needed
Text-only messages	Any text model (cloud or on-device)
Image attachments	Image-capable models only
Writing Pad pages	Image-capable models (pages are treated as images)
Voice input/output	Voice-capable models only

If a model type is not supported, the app will hide it or prevent selection.

5.5 Popular Models Available Through OpenRouter

OpenRouter gives you access to dozens of models from different providers through a single API key. Here are some of the most popular models available today:

Model	Provider	Best For	Speed	Cost
GPT-5.2	OpenAI	All-round best – writing, coding, reasoning, and image understanding	Medium	Higher
Gemini 3.1 Pro	Google	Complex reasoning, long documents, multimodal (text + images + video)	Medium	Medium
Claude Sonnet 4.6	Anthropic	Deep analysis, multi-step reasoning, long essays, and careful explanations	Medium	Medium
DeepSeek V3.2	DeepSeek AI	Strong reasoning and coding at very low cost – great value	Slower	Very Low
Kimi K2.5	Moonshot AI	Coding with vision, multimodal, agentic tasks – strong open-source option	Medium	Low

Model	Provider	Best For	Speed	Cost
Qwen 3.5	Alibaba	Multilingual support (200+ languages), math, and reasoning	Fast	Low
Llama 4 Maverick	Meta	General chat, coding, multimodal – open-weight, strong value	Fast	Low
Mistral Large 3	Mistral AI	Multilingual conversations, efficient reasoning, open-source (Apache 2.0)	Fast	Low

5.6 Model Selection Guide – Simple to Complex

Use this table to pick the right model depending on what you are trying to do. Models at the top are faster but less detailed. Models at the bottom are more powerful but may take longer.

Complexity	Example Questions	Recommended Model	Why
Simple recall	"What is the formula for area of a circle?" / "Define osmosis."	On-device (Gemma 3n E2B), Llama 4 Maverick, Mistral Large 3, or Qwen 3.5	Quick factual lookups do not need heavy reasoning. A fast or on-device model gives instant answers – even offline.
Short explanation	"Explain the difference between speed and velocity." / "What are the steps of cell division?"	Gemini 3.1 Pro, GPT-5.2, or On-device (Gemma 3n E2B)	These models give clear, structured short explanations.
Step-by-step math	"Solve: integrate x^2 dx" / "Find the derivative of sin(2x)."	GPT-5.2, Claude Sonnet 4.6, DeepSeek V3.2	These models are reliable at showing each step clearly. DeepSeek V3.2 offers similar quality at lower cost.
Complex reasoning	"Prove that √2 is irrational." / "Compare mitosis and meiosis with diagrams."	Claude Sonnet 4.6, GPT-5.2, Gemini 3.1 Pro	Multi-step reasoning needs a more capable cloud model.
Image + handwriting	Photo of handwritten notes, a textbook diagram, or Writing Pad sketch	GPT-5.2, Gemini 3.1 Pro, Kimi K2.5, On-device Gemma 3n (all image-capable)	Only image-capable models can read your photos and sketches.
Long essay or analysis	"Write a 500-word essay on the causes of World War I." / "Analyze this poem line by line."	Claude Sonnet 4.6, GPT-5.2	Longer outputs need models with higher output limits and strong coherence.
Budget-friendly study	Any question where you want good answers at minimal cost	DeepSeek V3.2, Qwen 3.5, Kimi K2.5, or On-device models	These models offer strong performance at a fraction of the price.

Tips for choosing:

- When in doubt, start with the default model. Switch only if the answer is too short, too slow, or not detailed enough.
- For everyday homework help, GPT-5.2 or Gemini 3.1 Pro are excellent all-rounders.
- For budget-conscious studying, DeepSeek V3.2 and Qwen 3.5 deliver strong results at very low cost through OpenRouter.
- For quick definitions or formulas, use an on-device model or a fast model like Llama 4 Maverick to save time.
- If you attach an image or Writing Pad page, always check that your selected model supports images.
- If you have no internet, use an on-device model – it works completely offline.

6. Settings and LLM Providers

6.1 The Settings Tab

The Settings tab lets you manage your profile, privacy, and AI configuration.

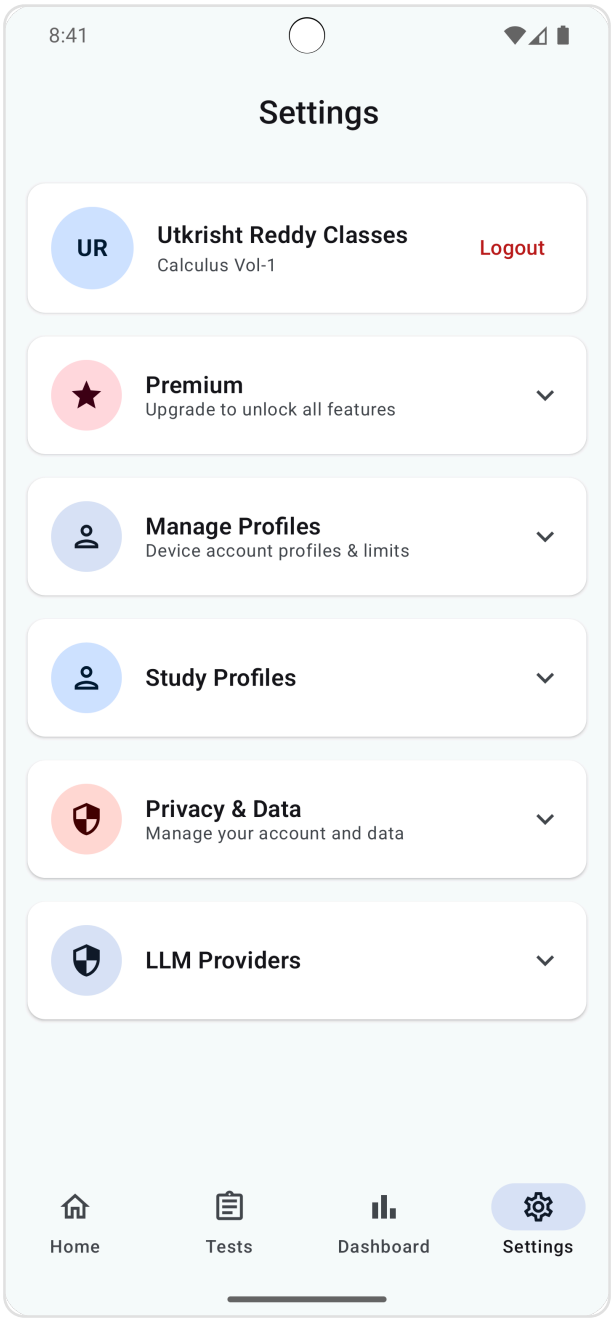

Figure 15: Figure 6.1 – Settings tab showing profile, premium, privacy, and LLM Providers

6.2 Opening LLM Providers

1. Open the app.
2. Go to **Settings**.
3. Tap **LLM Providers**.

AI Settings control which AI helper answers your questions. You must configure at least one provider before using AI features.

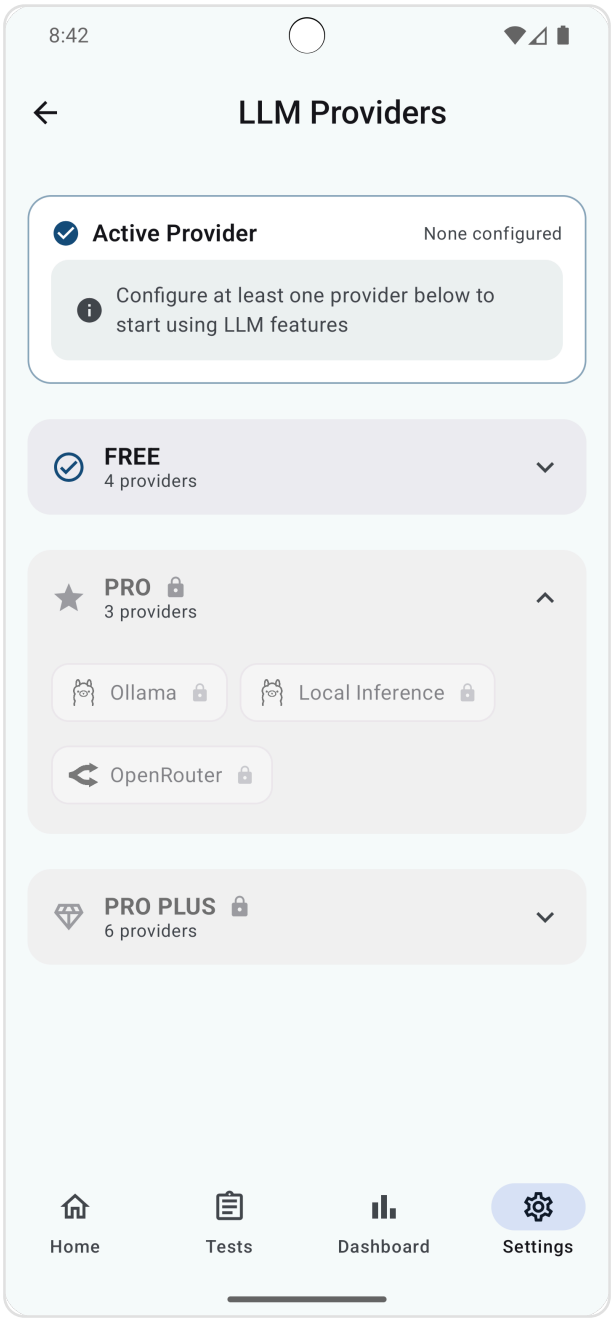

Figure 16: Figure 6.2 – LLM Providers screen showing provider tiers (Free, Pro, Pro Plus)

6.3 Provider Tiers

Providers are organized into three tiers:

Tier	What It Includes	Who It Is For
Free	Built-in providers with no API key needed	Students who want to get started quickly

Tier	What It Includes	Who It Is For
Pro	Ollama, Local Inference, OpenRouter	Students who want more model options or local/private AI
Pro Plus	AWS Bedrock, Azure OpenAI, Google Vertex AI, and more	Advanced users or school-managed deployments

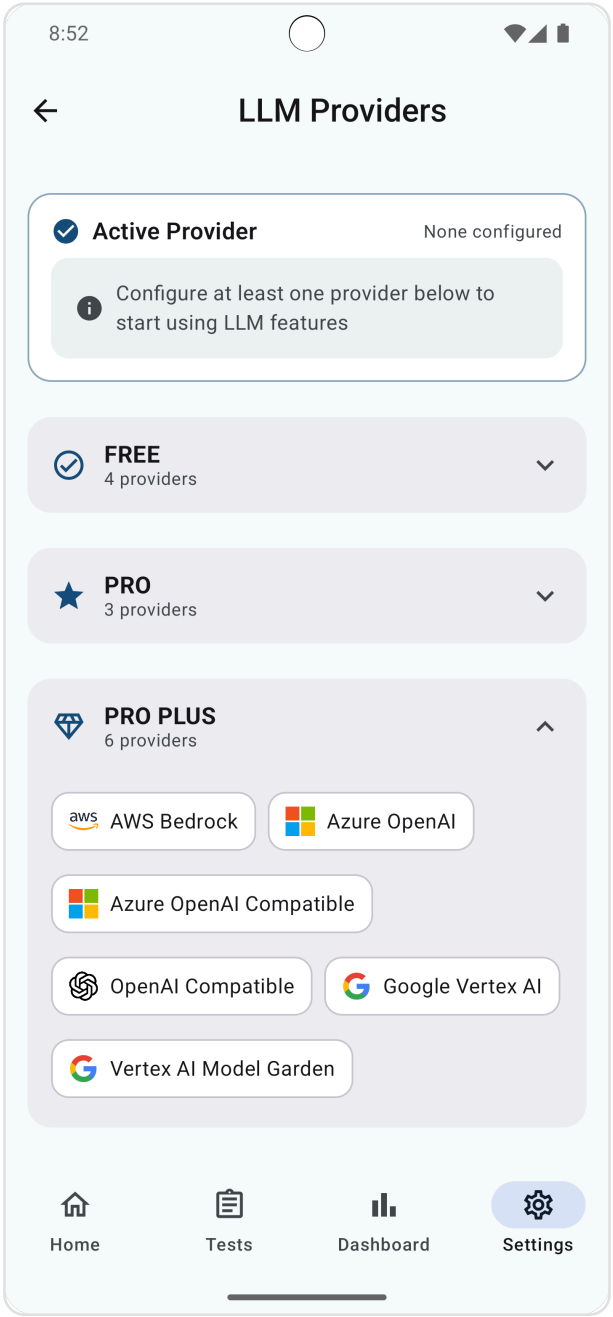

Figure 17: Figure 6.3 – Pro Plus tier showing enterprise-grade provider options

6.4 Understanding Pro and Pro Plus – What You Get and Why It Matters

The three provider tiers – Free, Pro, and Pro Plus – give you increasing levels of choice, privacy, and control. This section explains what each tier unlocks and helps you decide which one is right for you.

6.4.1 Pro Tier – More Models, More Flexibility

The Pro tier adds three providers on top of the Free tier:

Provider	What It Is
Ollama	Run open-source AI models (Llama 4, Mistral, Qwen, etc.) on a local computer. Free, private, unlimited use. Requires a PC or laptop running Ollama server on the same Wi-Fi network.
Local Inference	Auto-discovers any OpenAI-compatible AI server on your local network. Works with Ollama, LM Studio, or any local server.
OpenRouter	A single API key that gives access to 100+ models from OpenAI, Anthropic, Google, Meta, Mistral, DeepSeek, and more. Pay-per-use at low rates. Great for trying different models without separate accounts.

For students:

- OpenRouter is the easiest way to access many models with one key.
- Ollama and Local Inference let you run AI on a home computer for free – no per-message cost and full privacy.
- Pro is ideal if you want model variety or if your school runs a local AI server.

For institutions:

- Schools can set up an Ollama or LM Studio server and let all students connect via Local Inference – zero per-student API cost.
- OpenRouter lets schools use a single managed API key across the institution.

6.4.2 Pro Plus Tier – Enterprise Control, Privacy, and Independence

The Pro Plus tier is designed for schools, coaching centres, and advanced users who need enterprise-grade control over their AI setup. It adds these providers:

Provider	What It Is	Privacy and Independence Benefit
AWS Bedrock	Amazon's managed AI service. Requires AWS Access Key, Secret Key, and Region.	Data processed within your own AWS account. AWS does not use your data to train models. You control the region (data residency). Institutions can set IAM policies, usage quotas, and audit logs.

Provider	What It Is	Privacy and Independence Benefit
Azure OpenAI	Microsoft's hosted OpenAI models. Requires endpoint URL, API key, and deployment name.	Data stays within your Azure tenant. Microsoft guarantees enterprise data protection under your Azure agreement. Your data is not used for model training. Supports private endpoints (no public internet).
Azure OpenAI Compatible	Same as Azure OpenAI but uses the OpenAI-compatible REST format. For custom or third-party deployments on Azure.	Same Azure data protection. Useful when the institution runs its own inference endpoints on Azure.
OpenAI Compatible	Connect to any server that implements the OpenAI chat completions API. Requires a custom endpoint URL.	Full control – the institution decides where the server runs. Can be on-premises, in a private cloud, or in any region. Maximum independence.
Google Vertex AI	Google Cloud's enterprise AI platform. Uses project-level authentication.	Data processed in your Google Cloud project. Google's enterprise terms apply – no training on your data. Regional deployment options for data sovereignty.
Vertex AI Model Garden	Access 200+ models (including Llama, Mistral, Gemma) deployed on Google Cloud infrastructure. Requires project ID and region. Uses OAuth2 auto-authentication.	Models run inside your Google Cloud project with your billing. Same Google Cloud data protection. You choose which open-source models to deploy.

Note: inertGo also supports image-generation variants (Gemini Image and Vertex AI Image) for providers that offer them. These work the same way as their text counterparts but are used when you need AI-generated images.

6.4.3 Comparison Table – Free vs Pro vs Pro Plus

Feature	Free	Pro	Pro Plus
Cloud providers	OpenAI, Anthropic, Google Gemini	1. OpenRouter (100+ models)	1. AWS Bedrock, Azure OpenAI, Google Vertex AI
Local/private AI	On-device models only	1. Ollama, Local Inference	1. OpenAI Compatible (any server)
Data privacy	Data sent to provider's servers	Ollama/Local keeps data on your network	Enterprise agreements, data residency controls, audit logs
Who controls the AI	The provider (OpenAI, Google, etc.)	You (for local) or provider (for OpenRouter)	Your institution's cloud account
Cost control	Per-API-key usage	Free (local) or per-use (OpenRouter)	Institution-managed budgets and quotas
Best for	Individual students getting started	Students wanting variety or privacy	Schools, coaching centres, institutions

6.4.4 Real-World Examples

Example 1 – Individual student on a budget: Riya uses the Free tier with Google Gemini for everyday homework. When she needs to solve complex physics problems, she switches to OpenRouter (Pro) and uses DeepSeek V3.2 – which costs less than 1 paisa per question. For late-night revision without internet, she uses an on-device model.

Example 2 – Student who values privacy: Arjun's family is concerned about his study data being sent to foreign servers. His father sets up Ollama on the family laptop with Llama 4 Maverick. Arjun connects via Local Inference (Pro) – all his questions and answers stay within their home network, completely free.

Example 3 – School deployment (Pro Plus): Delhi Public School configures Azure OpenAI under their Microsoft education agreement. They set up a single deployment and distribute API keys to 500 students. All student data stays within the school's Azure tenant in the Mumbai region. The IT admin monitors usage through Azure dashboards and sets monthly cost caps per student.

Example 4 – Coaching centre with multiple locations (Pro Plus): Vidya Academy runs Google Vertex AI Model Garden and deploys Gemini 3.1 Pro and Llama 4 Maverick on their Google Cloud project. Students at all 12 branches connect through the app. The centre controls which models are

available, monitors costs per branch, and all data stays within their Google Cloud project under Indian data residency.

6.5 Configuring a Provider

1. Select a provider card (for example, OpenAI, OpenRouter, or Local Inference).
2. Enter your **API key** if required. Keep this private – do not share it with classmates.
3. Under **Model**, tap a model chip to select your preferred model, or tap **Probe Models** to fetch the latest available models.
4. Tap **Test & Save** to verify your configuration.

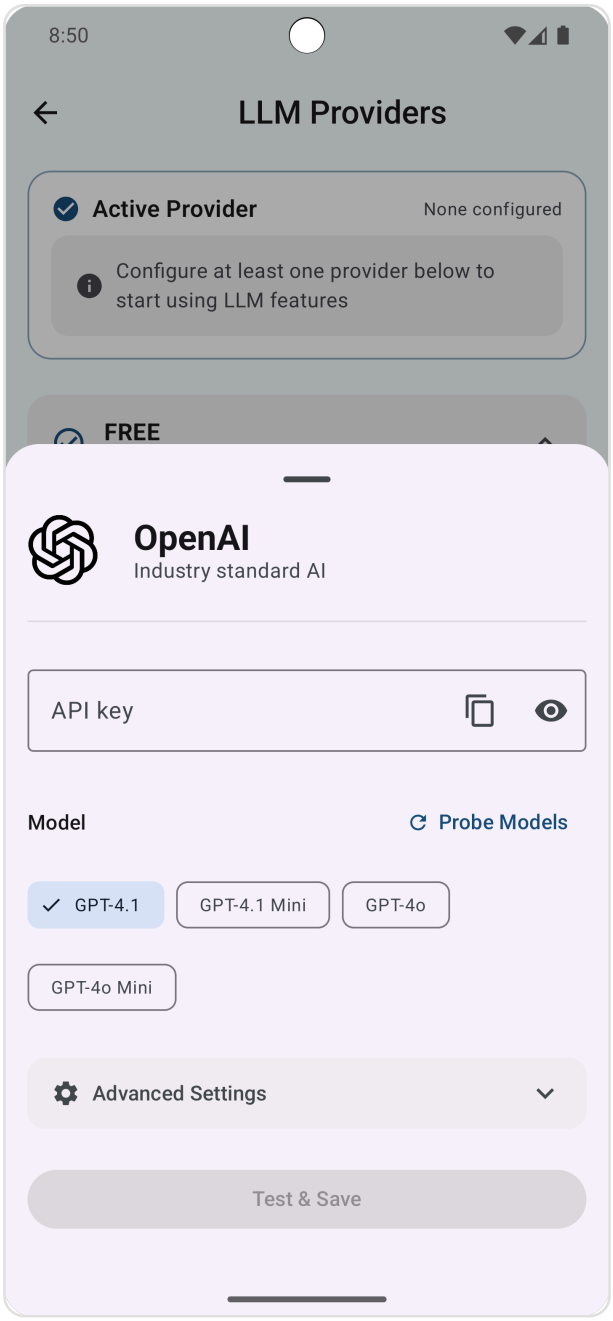

Figure 18: Figure 6.4 – Provider settings for OpenAI showing API key entry, model chips, and Test & Save

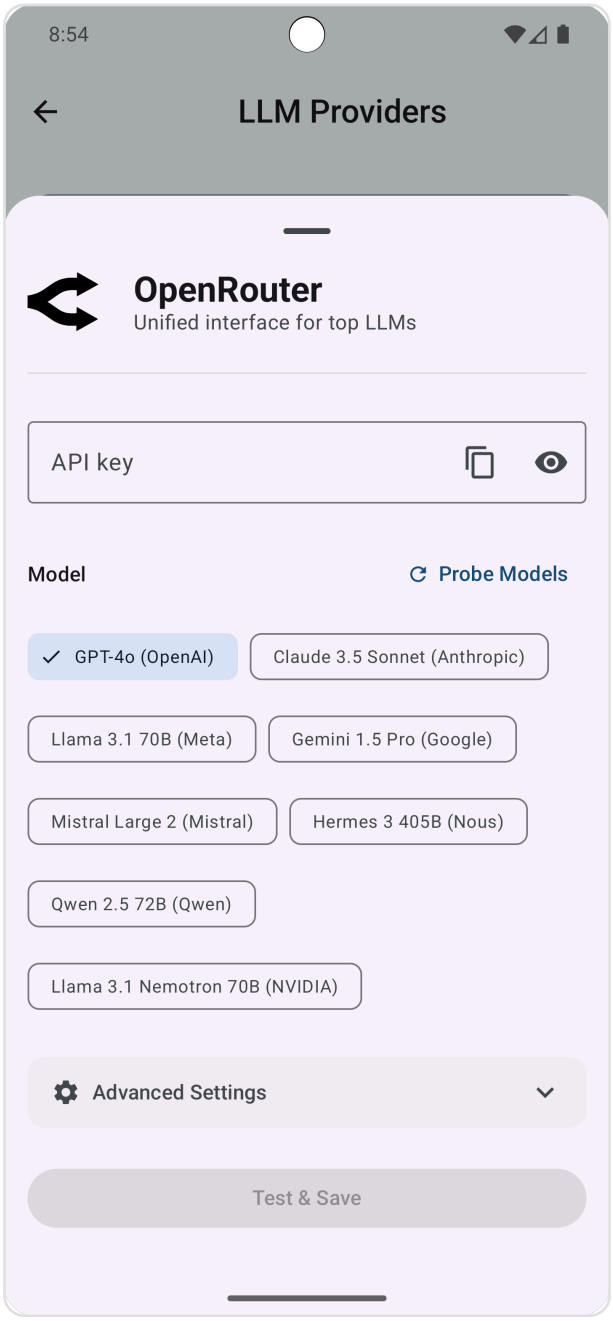

Figure 19: Figure 6.5 – OpenRouter provider configuration with model list

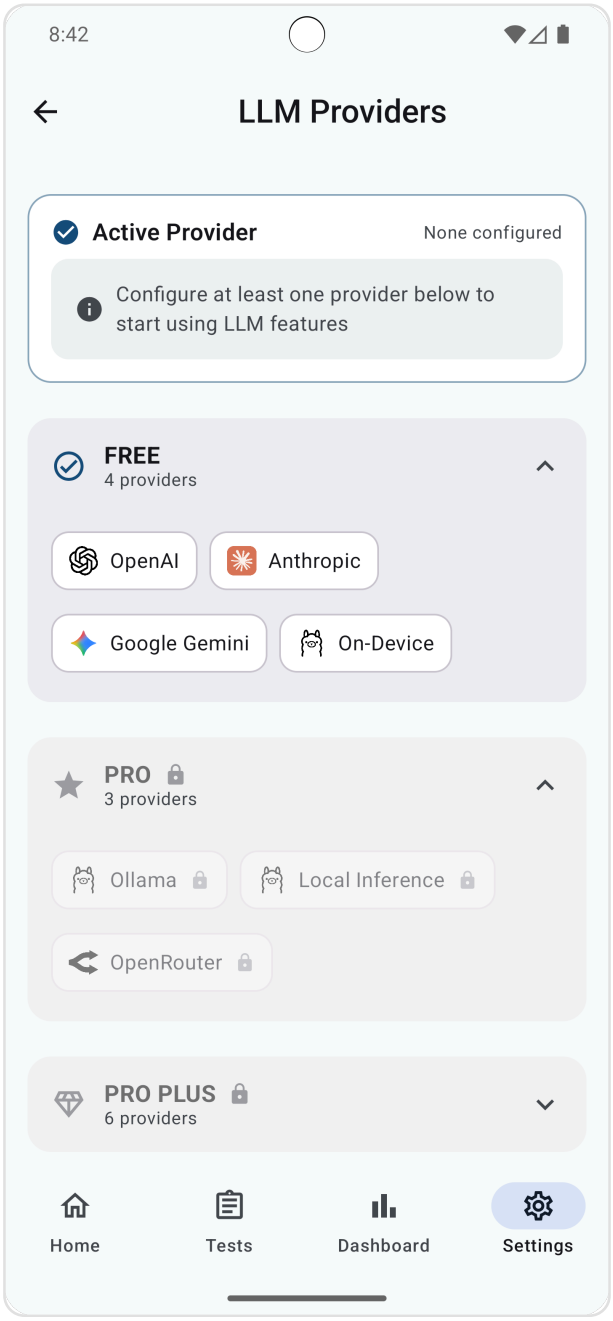

Figure 20: Figure 6.6 – Additional provider configuration options

6.6 On-Device AI Models – Learn Offline, Free, and Private

This is one of inertGo's most powerful features. On-device models run AI directly on your phone, which means:

- **No internet required** – study on a train, in a village, or anywhere without Wi-Fi.
- **Completely free** – no API keys, no subscriptions, no per-message costs.
- **Fully private** – your questions and answers never leave your device.

Available On-Device Models

Model	Size	Best For
LFM2 1.2B (ExecuTorch)	1.1 GB	Fast, lightweight answers. Good for definitions and quick help.
Gemma 2 2B	1.4 GB	Small and fast. Good for quick responses and simple explanations.
Gemma 3n E2B (Multimodal)	2.9 GB	Mobile-optimized. Can understand both text and images.
Gemma 3n E4B (Multimodal)	5.0 GB	More powerful multimodal model. Better for detailed explanations.

How to Download an On-Device Model

1. Go to **Settings > LLM Providers**.
2. Select the on-device models section.
3. Tap **Download** next to the model you want.
4. Wait for the download to complete (requires internet only for this one-time download).
5. Once installed, the model works entirely offline.

Note: Some models (marked with a warning icon) require an access token to download. Your teacher or admin can provide this.

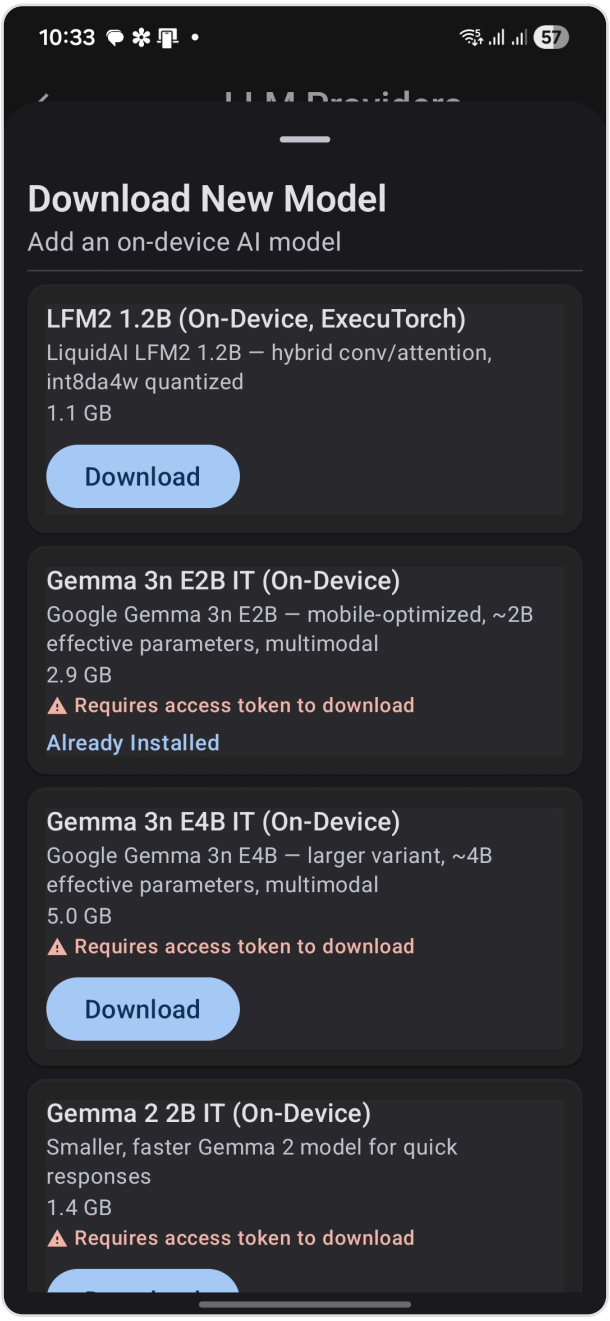

Figure 21: Figure 6.7 – On-device model download screen showing available models and sizes

Tips for On-Device Models

- Start with **Gemma 2 2B** (1.4 GB) if your phone has limited storage.
- Use **Gemma 3n E2B** if you want to attach images or Writing Pad pages offline.
- On-device models are slightly less powerful than large cloud models, but they are perfect for quick help, revision, and offline study.
- Make sure your phone has enough free storage before downloading.

6.7 How On-Device AI Works – Under the Hood

When you use an on-device model, the AI runs directly inside the inertGo app on your phone's processor. No data is sent to any server. This section explains what happens behind the scenes and how to get the best experience.

What Happens When You Send a Message

1. **Model loading:** The first time you send a message, the app loads the model file into your phone's memory (RAM). This takes 5–15 seconds depending on the model size. Once loaded, the model stays in memory so future messages are faster.
2. **Tokenization:** Your message is converted into a sequence of numbers (called tokens) that the AI can understand.
3. **Inference:** The AI processes the tokens and generates a response, one word at a time. You will see the reply appear gradually as each word is produced – this is called *streaming*.
4. **Session caching:** The conversation context is kept in a structure called a *KV cache*, so follow-up questions in the same conversation are faster because the AI remembers what was already discussed.

Available On-Device Models – Detailed Guide

Model	Size	Vision	Best For
Qwen 2.5 3B Instruct	2.1 GB	No	Best all-round on-device tutor. Strong at math, structured explanations, and follow-up questions. 32K context window.
Qwen 2.5 Math 1.5B	990 MB	No	STEM specialist with Chain-of-Thought reasoning. Scores 79.7% on the MATH benchmark. Ideal for step-by-step math problem solving.
Gemma 3n E2B IT	2.9 GB	Yes	Google's mobile-optimized model with GPU acceleration. The only on-device model that can understand images – attach Writing Pad pages or textbook photos. Requires a HuggingFace access token to download.
SmolLM2 1.7B Instruct	1.1 GB	No	Compact and fast. No authentication required to download. Good zero-friction starting point for students with limited storage.

Performance Expectations

On-device models run on your phone's hardware, so performance depends on your device. Here is what to expect:

Factor	What to Expect
First message	5–15 seconds to load the model, then the reply begins streaming. Subsequent messages in the same conversation are faster.
Response speed	Typically 5–15 tokens per second on a mid-range phone (roughly 1–3 words per second). Flagship phones are faster.
RAM usage	Models need 3–4 GB of free RAM. Close other apps if you experience slowness or crashes.
Battery	AI inference uses significant processing power. Expect moderate battery drain during active use – similar to playing a mobile game.
Storage	Models are stored permanently after download. You can delete them from Settings > LLM Providers to free space.

When to Use On-Device vs Cloud Models

Use On-Device When	Use Cloud When
No internet available (travel, rural areas, offline study)	You need the highest quality answers for complex problems
You want complete privacy – nothing leaves your phone	You need image understanding and Gemma 3n is not enough
You want free, unlimited AI with no API costs	You need long, detailed essays or multi-step proofs
Quick lookups: definitions, formulas, short explanations	Speed is critical and you have a fast internet connection
Late-night revision without using mobile data	You are working on creative writing or open-ended analysis

Tips for Best On-Device Performance

- **Close background apps** before using on-device AI. More free RAM means faster responses.
- **Start with SmolLM2 1.7B** if your phone has limited storage or RAM. Upgrade to Qwen 2.5 3B when you are comfortable.
- **Use Qwen 2.5 Math 1.5B** specifically for math and science problems – it is trained to show step-by-step working.
- **Use Gemma 3n E2B** when you need to attach images offline (Writing Pad pages, textbook photos).

- **Keep questions focused.** On-device models work best with specific, clear questions rather than broad prompts.
- **Do not switch models mid-conversation.** The app resets the AI's memory when you switch, so start a new question if you want to try a different model.

6.8 Using Local Inference (Advanced)

If you want to run AI on a computer on your local network (free and private), you can use the Local Inference option.

1. In LLM Providers, select **Local Inference**.
2. Enter your local server URL, or tap **Discover Local Servers** to find one automatically.
3. Tap **Test & Save**.

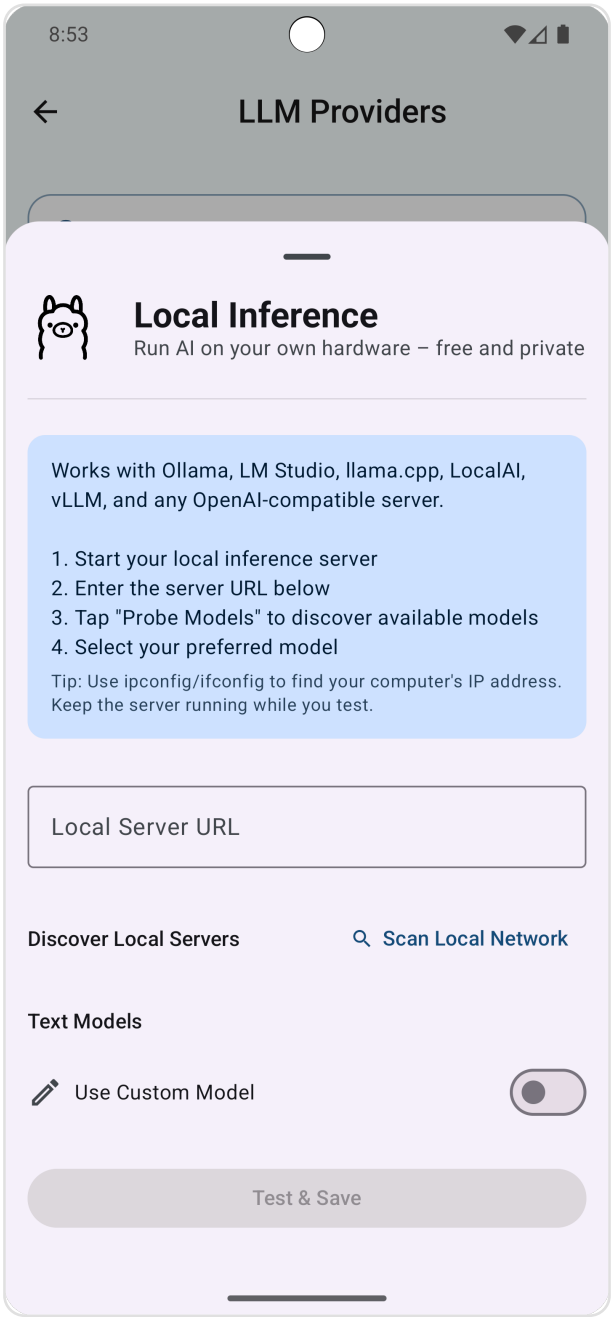

Figure 22: Figure 6.8 – Local Inference setup with server discovery options

6.9 Recommended Settings for Students

- Use a Free-tier provider or an on-device model for best stability and zero cost.
- Only switch providers or models if your teacher recommends it or if you need a specific capability (such as image support).
- If you are unsure, use the default model shown in the app.

7. Daily Tests

7.1 What Daily Tests Are

Daily Tests provide short, focused practice sessions to help you build consistency. You can take tests, review your answers, and use the AI to understand questions you got wrong.

7.2 Opening Daily Tests

1. From the bottom navigation, tap **Tests**.
2. Choose the test for today from the "Due Today" section, or browse by topic.
3. Tap **Start** to begin.

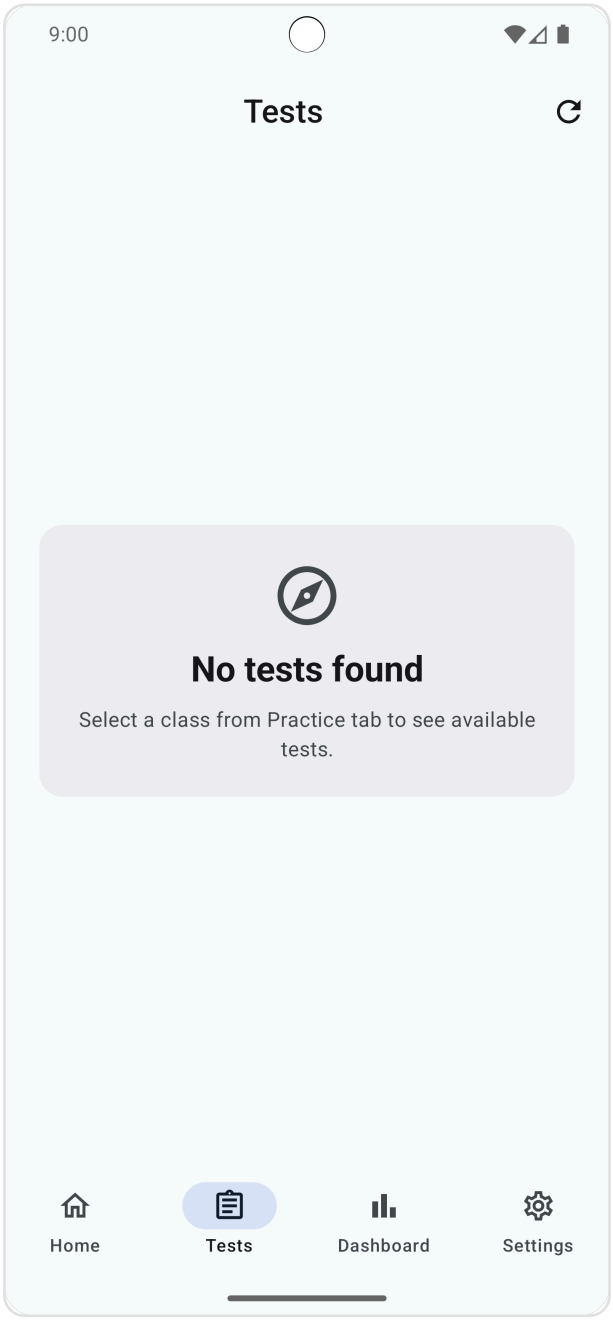

Figure 23: Figure 7.1 – Tests tab (initial empty state before sync)

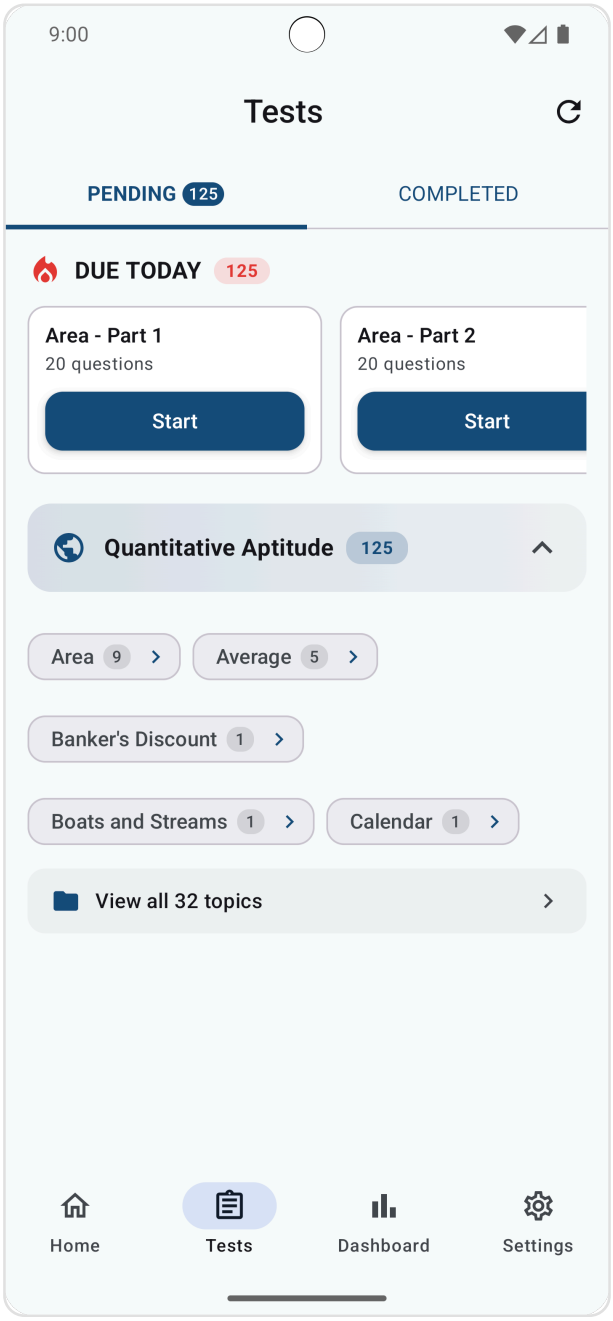

Figure 24: Figure 7.2 – Tests tab after syncing, showing pending tests grouped by topic

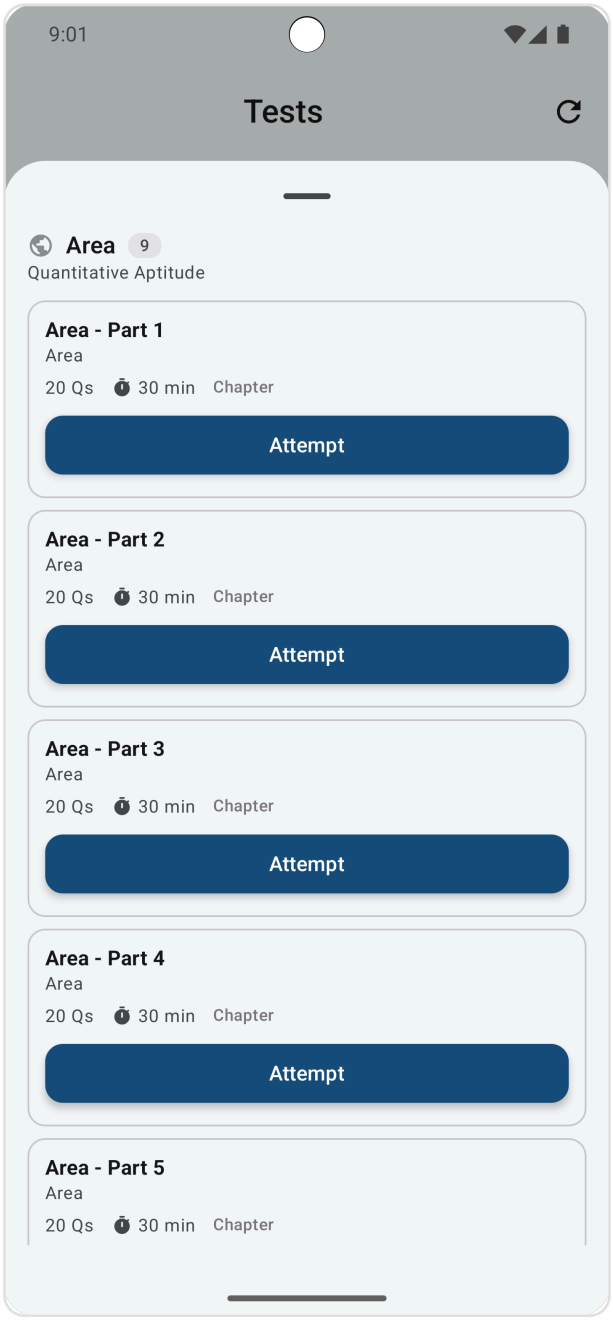

Figure 25: Figure 7.3 – Test list for selecting a specific test

7.3 Taking a Test

- Read each question and select your answer from the choices.
- Use **Next** and **Previous** arrows or the question grid at the bottom to move between questions.
- Optionally, write a reasoning for your answer in the “Why did you choose this?” field.
- When you reach the last question, tap **Submit**.

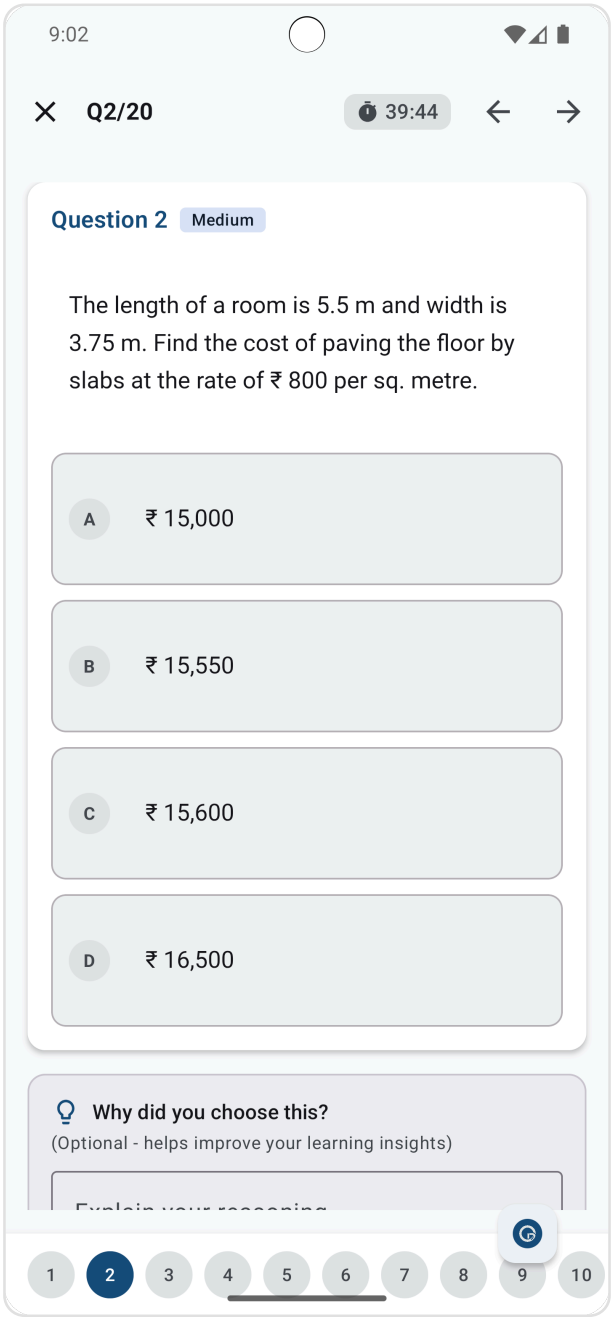

Figure 26: Figure 7.4 – Active test: question text, answer choices, timer, and question grid

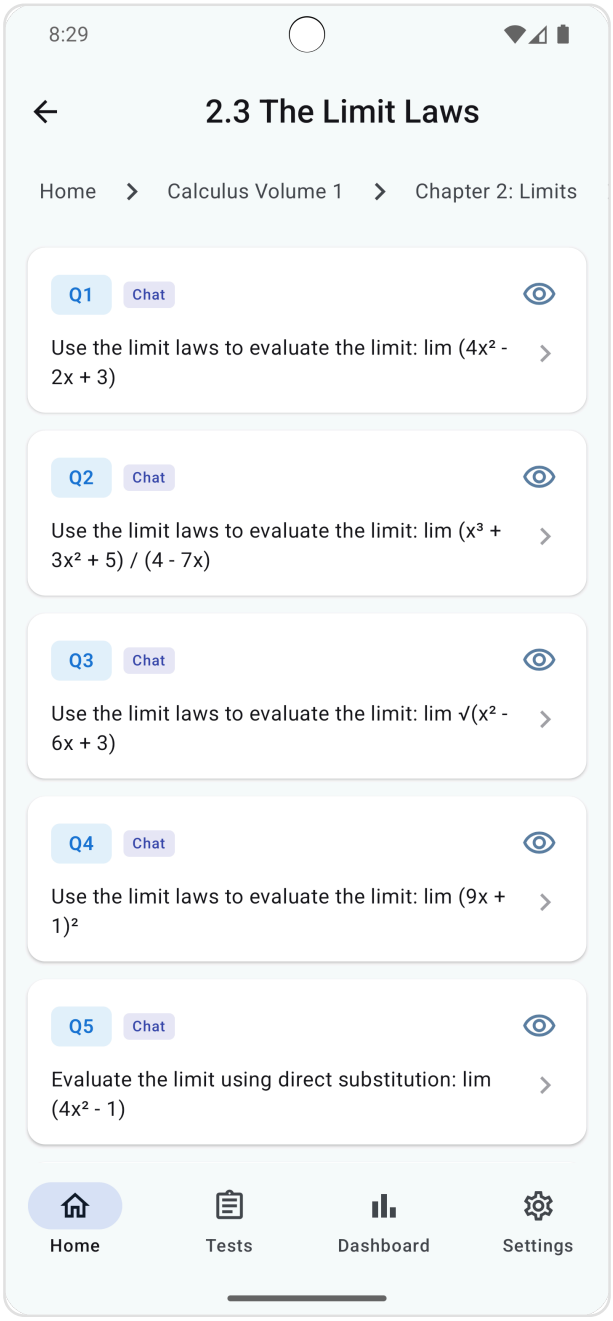

Figure 27: Figure 7.5 – Question navigation grid for jumping between questions

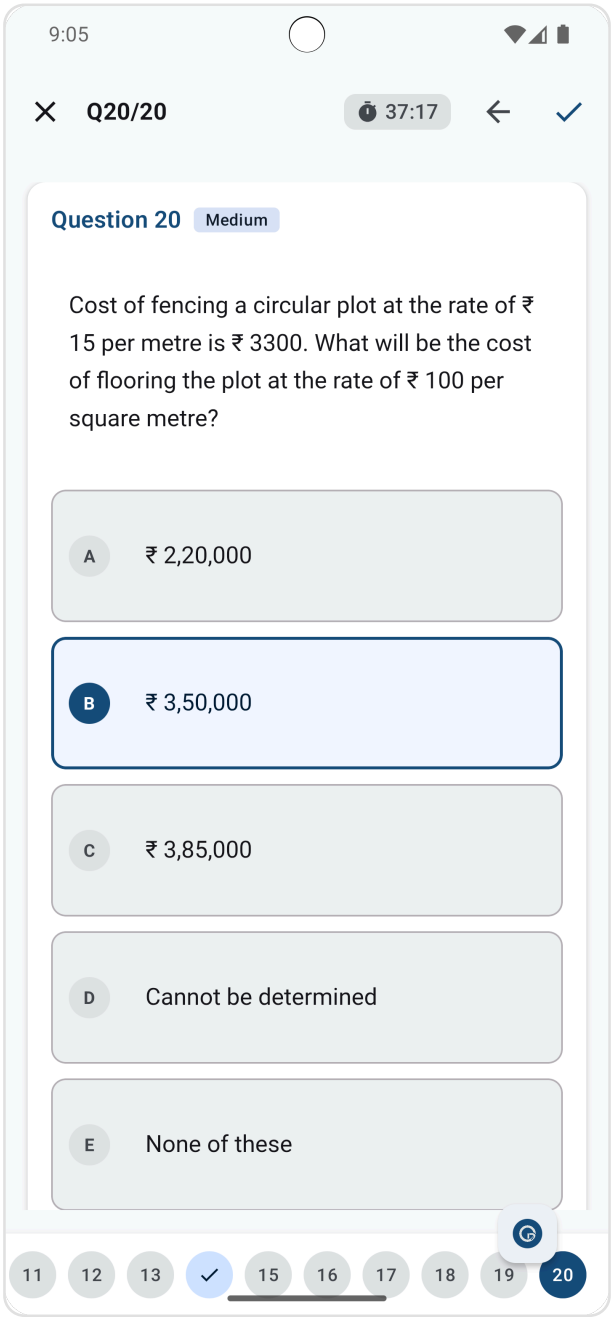

Figure 28: Figure 7.6 – Last question before submission

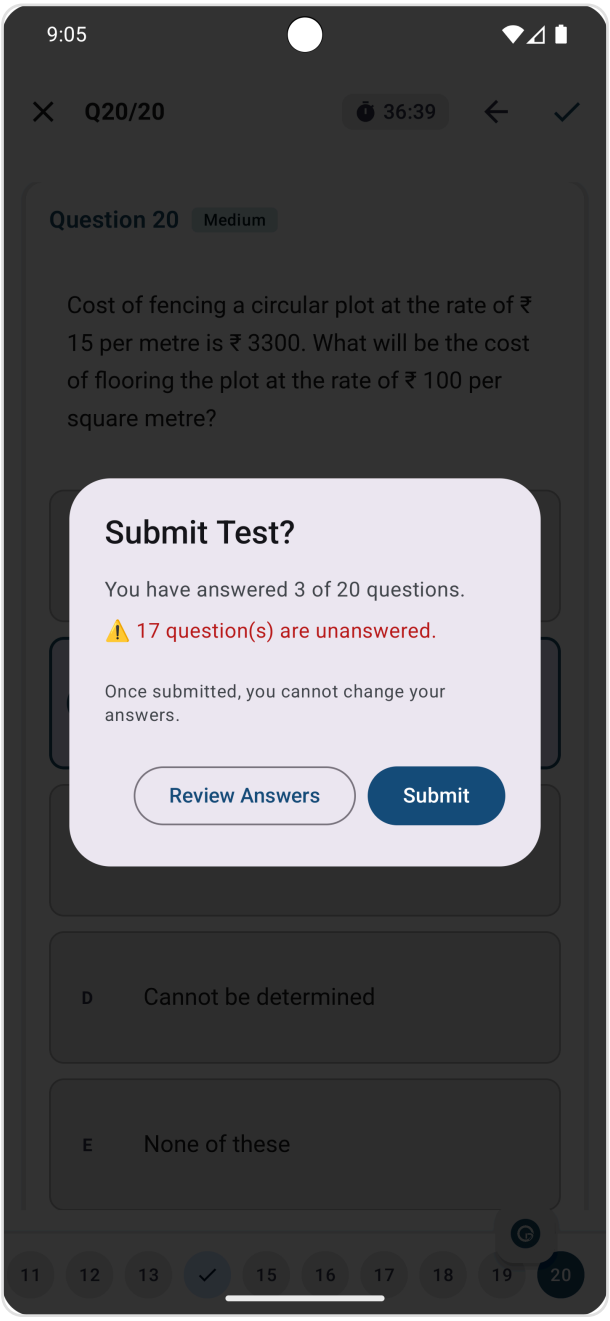

Figure 29: Figure 7.7 – Test submission confirmation dialog

7.4 Reviewing Results

After submitting, you can see your score, which questions you got right or wrong, and how much time you spent.

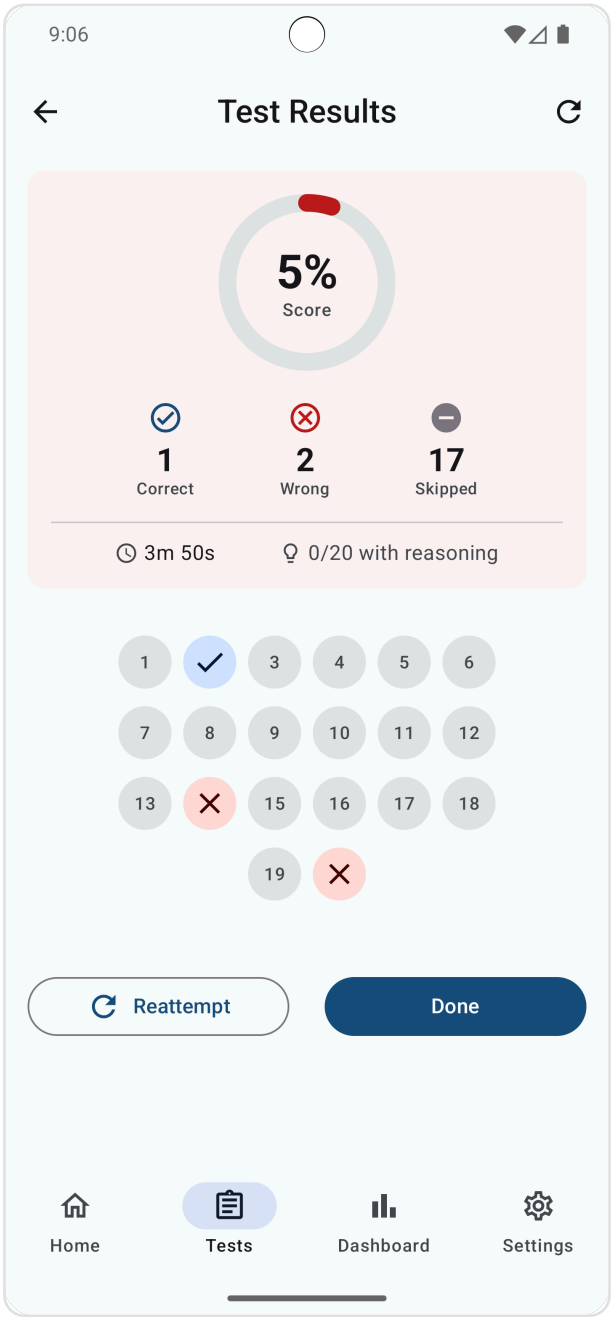

Figure 30: Figure 7.8 – Test results: score breakdown, correct/wrong/skipped counts, and question grid

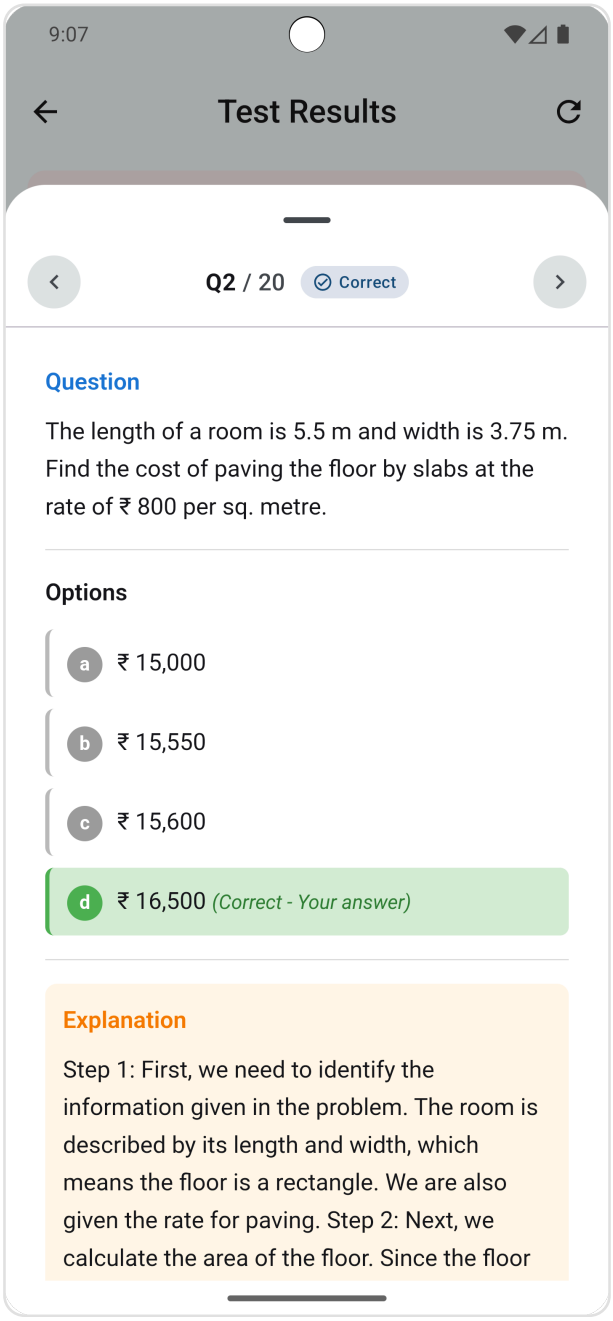

Figure 31: Figure 7.9 – Reviewing individual answers after a test

7.5 Using AI During Tests (GoCompanion)

During a test, you can open the GoCompanion AI assistant to ask questions about a specific problem. This is useful when you are stuck and need a hint or explanation.

1. During a test, tap the AI assistant button.
2. Type your question or use voice input.
3. The AI will help explain the problem without giving away the direct answer.

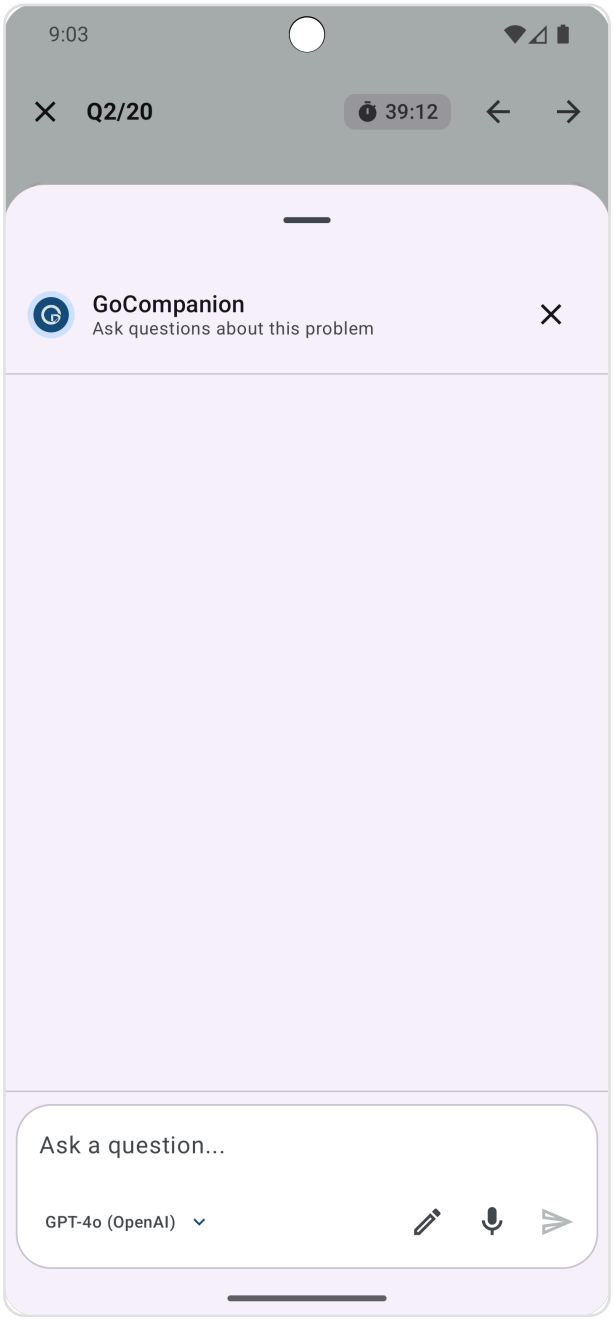

Figure 32: Figure 7.10 – GoCompanion AI assistant during a test

You can also use the Writing Pad inside GoCompanion to show your working:

Figure 33: Figure 7.11 – GoCompanion with the Writing Pad open for handwritten work

8. Dashboard (Analytics Hub)

8.1 What the Dashboard Shows

The Dashboard tab gives you an overview of your learning activity, including test performance trends and conversation history with the AI.

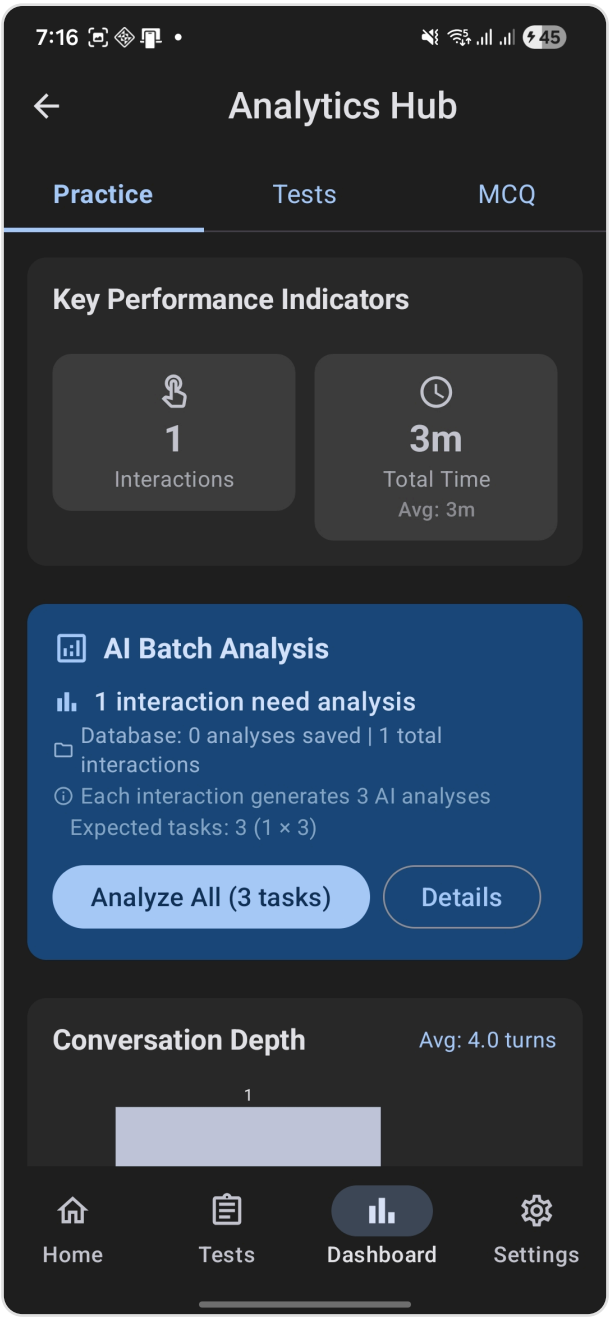

Figure 34: Figure 8.1 – Dashboard main screen

8.2 Conversation History

You can review your past AI conversations from the Dashboard. This is useful for revisiting explanations or finding previous solutions.

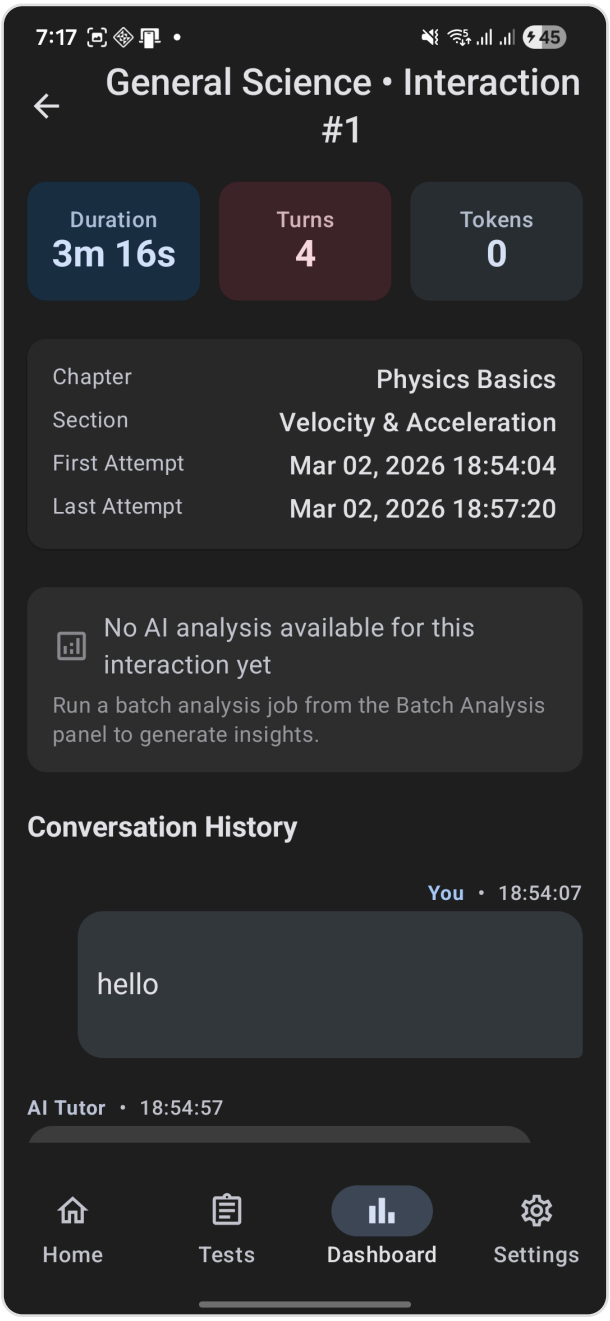

Figure 35: Figure 8.2 – Conversation history in the Dashboard

9. Responsible Use and Privacy

9.1 Academic Integrity

- Follow your school's rules on AI use. Some assignments may not allow AI help.
- If you use the AI to help with schoolwork, be honest about it. Ask your teacher how to cite AI-generated content.
- Do not submit AI-generated answers as your own original work unless your teacher permits it.

9.2 The AI Can Be Wrong

- AI responses are not always correct. Always double-check facts, math, dates, and citations against your textbook or teacher's notes.
- If an answer looks wrong, ask the AI to explain its reasoning or show its steps.

9.3 Privacy

- Do not share personal information with the AI (full name, address, phone number, school ID, passwords).
- Keep your API key secret. Do not share it with classmates or post it online.
- When using cloud providers, be aware that your messages may be processed by an external server.
- When using on-device models, your data stays entirely on your phone.

10. Common Troubleshooting

Problem	What to Do
AI response is slow	Switch to a faster model or an on-device model from the compose bar. Keep requests shorter and focused.
Attachments not appearing	Reopen the compose area and attach again. Make sure you are using an image-capable model.
"Invalid API Key" error	Go to Settings > LLM Providers and re-enter your API key. Check with your teacher that the key is still active.
Provider not responding	The provider may be temporarily down. Wait a few minutes and try again, or switch to a different provider or an on-device model.
Model not available	The model may have been removed by the provider. Go to LLM Providers and select a different model.
No internet connection	Use an on-device model for offline AI. Cloud providers need internet to work.
Images not understood by AI	Make sure the photo is clear and well-lit. Retake if blurry. Confirm you are using an image-capable model.
Permission denied (camera/storage)	Go to your device Settings > Apps > inertGo > Permissions and enable the required permissions.
On-device model download fails	Check your internet connection and available storage. Some models require an access token – ask your teacher.

11. Glossary

Term	Meaning
AI Helper	The tool that answers your questions using an LLM.
Analysis	A deeper explanation mode for complex questions.
API Key	A secret code that connects the app to a provider's service. Keep it private.
Compose Area	The bottom input area where you type messages and attach items.
GoCompanion	The AI assistant available during tests for hints and explanations.
LLM	Large Language Model – the technology that powers the AI helper.
Model	The type of AI helper you selected. Different models have different strengths.
On-Device Model	An AI model that runs directly on your phone, with no internet needed.
Provider	The company or service that hosts a model (for example, OpenAI, Google Gemini).
Writing Pad	The in-app drawing and writing tool for sketching notes and attaching them to chat.