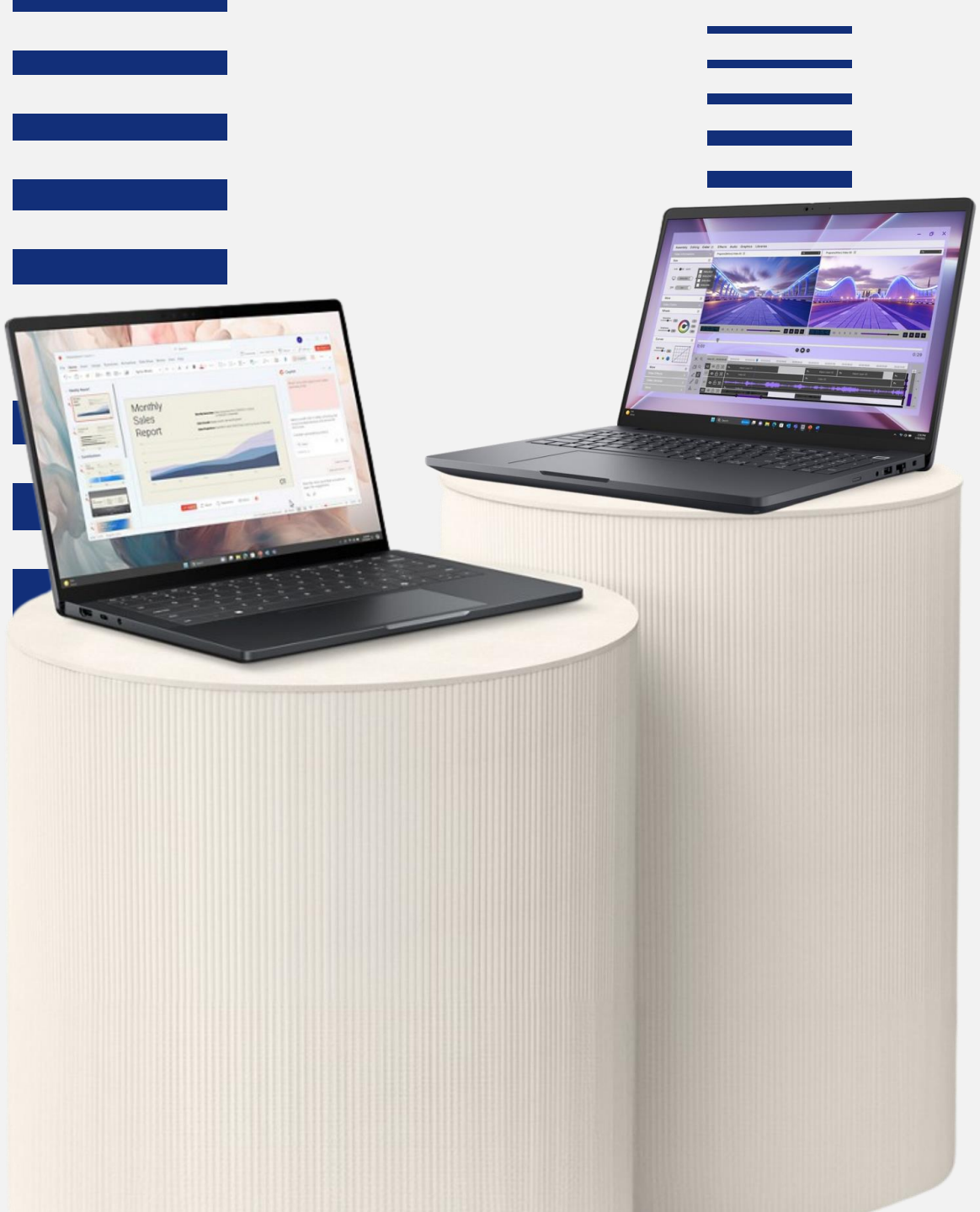
DELL Technologies

# AI Recipes from the Dell AI Kitchen

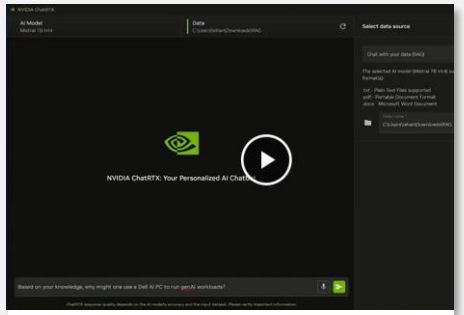
Master AI today with simple, actionable recipes that guide you through every step.



# Discover AI, One Recipe at a Time

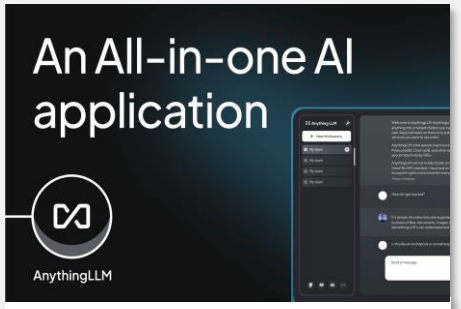
Select an AI recipe you'd like to explore


Run AI models locally with LMStudio

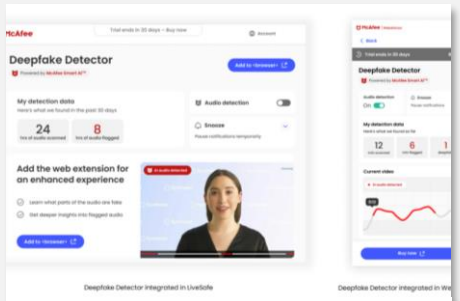


RAG App using NVIDIA ChatRTX


Content Gen App using NVIDIA AI Workbench




Running AnythingLLM on Dell Copilot+ PCs




McAfee Deepfake Detection

# AI RECIPES FROM THE DELL AI KITCHEN


Watch a quick video tutorial

## Run AI models locally with a Dell Pro Max PC

As companies embrace AI to make the day-to-day easier, many are drawn to cloud-based solutions due to the ease of setup. But the truth is that no matter how appealing cloud-based solutions sound, they can expose sensitive company data to risk.

At Dell Technologies, we advocate for running AI directly on your device to keep things more secure. AI PCs and AI Workstations not only make you more efficient, but it will also change *how* you get stuff done. The best part? You're not tied to an internet connection or reliant on cloud services. You can get your work done securely, independently, right on your device.

## What you need to get started

**Dell Pro Max PC**

With a Dell Pro Max PC with a discrete GPU, your machine can handle AI workloads faster and more efficiently

*Shop Dell Pro Max PCs* →

**LM Studio** (Or GPT4All, Ollama)

Run local Large Language Models (LLMs) through this application on your AI Workstation

# Get Started with LMStudio

## Steps:

### 1. Make your sure your PC can handle running LLMs locally

If you have 10GB+ of RAM, you should be able to run quantized versions of popular 3B - 7B models comfortably. With an NPU or a high performing discrete GPU your operations will run faster and will be less tasking on your device

### 2. Download an AI software platform

Directly download the application from lmstudio.ai or O llama or GPT4ALL

### 3. Load an LLM on LM Studio

Start with a popular open-source model like Llama 3 or Phi-3. You may have to experiment with which sized version of the model works well on your device.

### 4. Prompt your LLM with the relevant information about your task

Prepare your prompt, whether you are wanting to create formatted tables quickly with raw data or create content in minutes, prompt your LLM as you would with any other chatbot. Here's are a few example for you to easily copy and paste and customize to fit your needs

### 5. Press Enter and wait for the model to produce a response.

Some rules of thumb: If you have longer prompt it can take longer time to process, if you have a smaller model, it typically will generate a response faster, or if you have larger model, it is usually more capable (but not always true!)

You're now using LLMs locally in LM Studio, with full data privacy and no cost, (and you don't even need to be connected to the internet).

**EXAMPLE 1 (Create a formatted table in seconds)**
Save time and reduce errors in data entry by automating the conversion of raw and unorganized data into a structured table format.

INPUT EXAMPLE:

```
[Please create a table with the following columns: Name, Date, and Quantity. Populate the table with the data provided.]
```

**EXAMPLE 2 (Automate meeting summaries)**
Streamline the creation of meeting summaries and follow-up notes with minimal manual effort by inputting your meeting transcript into the LLM, ensuring clarity in your notes.

INPUT EXAMPLE:

```
[Please summarize the meeting notes. Include key points, decisions made, and any assigned responsibilities.]
```

Save the finalized summary. The LLM can then reference this meeting to incorporate additional content from follow-up transcripts

**EXAMPLE 3 (Create social campaigns in just minutes)**
With on-device LLMs, you can confidently create a comprehensive and secure social media campaign quickly that ensure proprietary and confidential launch information stays protected.

INPUT EXAMPLE:

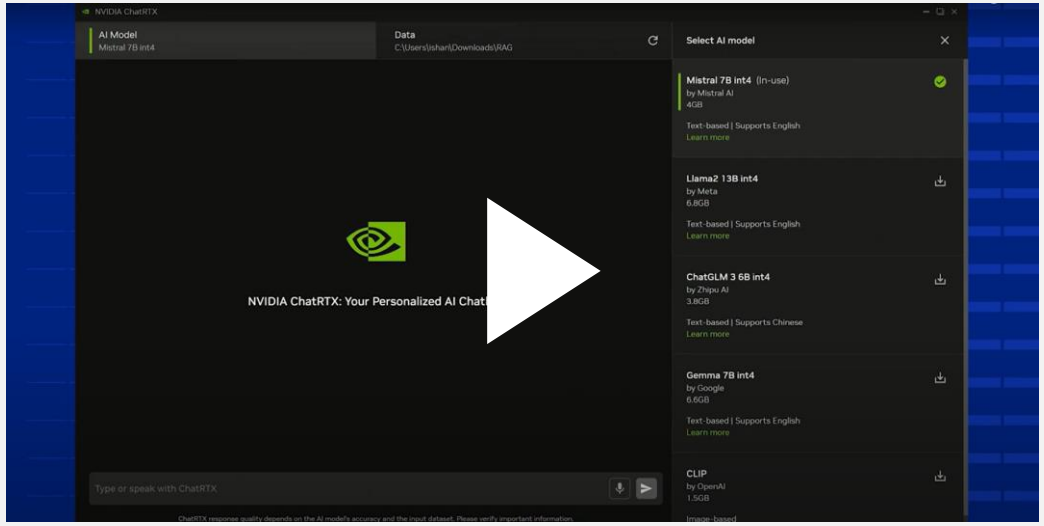
```
[Create a social campaign to boost engagement for our new summer product line, targeting young adults aged 18-30 who love outdoor activities. The theme is "Summer Fun," with key messages highlighting our products as ideal for summer adventures and outdoor enjoyment
```

## What more can you do?

- Turn on GPU acceleration in LM Studio for faster performance
- Experiment with different prompts
- Experiment with different models (e.g. coding models if you're a coder)

Explore use cases on Dell.com/AI

# AI RECIPES FROM THE DELL AI KITCHEN




Watch a quick video tutorial

## Create a Retrieval-Augmented Generation (RAG) app using NVIDIA ChatRTX and a Dell Pro Max PC

Transform ideas into AI-driven applications effortlessly with unmatched security and performance directly on your device. Dell Pro Max high-performance PCs, powered by NVIDIA RTX GPUs, provide the tools to harness this power, making it the foundation for running advanced AI workloads.

Now pair that with a Retrieval-Augmented Generation (RAG) app, which acts as your personal AI assistant. With RAG, generative AI combines with a knowledge base to deliver more accurate, context-aware responses. Whether you're enhancing customer support, analyzing internal documents or delivering actionable insights, RAG simplifies complex tasks while tailoring outputs to your data.

## What you need to get started

### Dell Pro Max PC

- With any NVIDIA RTX GPU
- 8+ GB of VRAM
- 16+ GB of RAM
- Windows 11

*Shop Dell Pro Max PCs* →

### NVIDIA ChatRTX

- Free download
- Approx 11GB download
  so stable internet connection needed

# Get Started with NVIDIA AI Workbench

## Steps:

### 1. Install NVIDIA ChatRTX

Download the application and extract the zip file to initiate installation. During installation, the system will verify your GPU compatibility. If compatible, choose an installation folder and follow the prompts to complete the installation. Launch NVIDIA ChatRTX via the desktop icon; it will open in your default browser and generate a command prompt for logs.

### 2. Load a Large Language Model (LLM)

Once NVIDIA ChatRTX is installed, locate the model dropdown menu on the right-hand side of the interface. Select a model suited for your task. For this recipe, we recommend using Mistral 7B in4 for its balance of performance and compression. Download and install the model by clicking on it. Verify it's ready by checking for the green checkmark next to the model name.

### 3. Prepare your Dataset

Create a folder containing the documents you want ChatRTX to use as its knowledge base (e.g., PDFs, text documents). Open the ChatRTX interface and locate the option to select a dataset folder. Point to your prepared folder and allow a few seconds for processing. ChatRTX will scan these documents and create a searchable library.

### 4. Test and Generate Responses

Use the chat interface to ask queries based on your uploaded documents. ChatRTX will analyze your dataset using retrieval-augmented generation, referencing your custom documents to create relevant, detailed responses.

Download NVIDIA ChatRTX here

**EXAMPLE 1 (Create accurate customer support responses)**
Quickly respond to customer queries with precise, context-driven answers generated from your knowledge base or document repository. NVIDIA ChatRTX ensures every response is accurate and tailored.

INPUT EXAMPLE:
```
[Answer this customer query using product manuals in the "Support Docs" folder: "What is the maximum storage capacity of the Dell Pro Max AI PC, and can I upgrade it?"]
```

**EXAMPLE 2 (Generate customized product recommendations)**
Deliver personalized product recommendations by analyzing user preferences and pulling relevant insights. NVIDIA ChatRTX ensures accurate, meaningful suggestions based on stored data.

INPUT EXAMPLE:
```
[Based on customer preferences stored in "Client Profiles," recommend a Dell workstation. Prioritize preferences like long battery life, portability, and powerful GPUs.]
```

**EXAMPLE 3 (Enhance e-commerce experiences)**
Amplify product discovery by creating relevant and engaging descriptions sourced from your product database. NVIDIA ChatRTX ensures descriptions are aligned with customer interests.

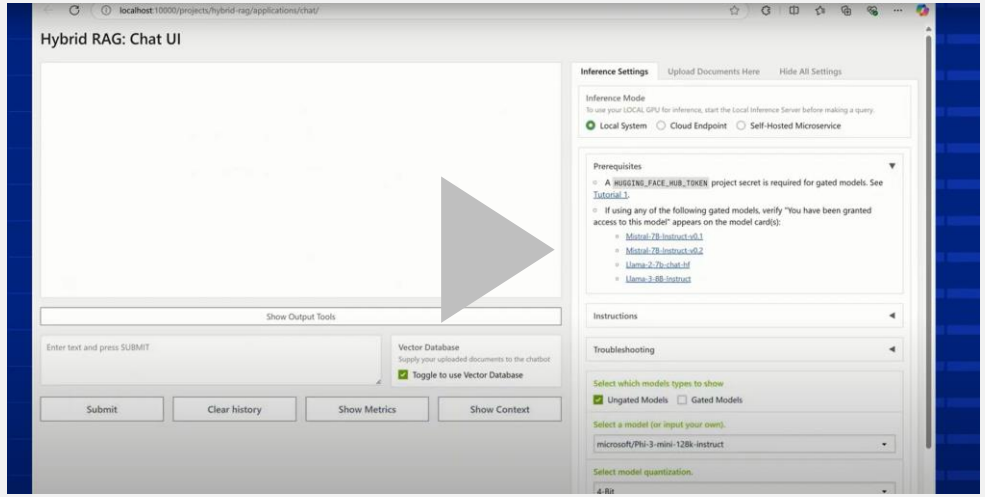
INPUT EXAMPLE:
```
[From "Product Descriptions," generate a customer-facing marketing description for an advanced Dell Pro Max AI PC, emphasizing its AI capabilities and NVIDIA GPUs.]
```

## Tips for building a RAG app using NVIDIA ChatRTX

- Turn on GPU acceleration in NVIDIA Chat RTX for faster processing
- Try different LLMs for diverse use cases (e.g. coding-specific models for development tasks)
- Experiment with prompt phrasing to get the most effective responses

DELL Technologies NVIDIA

# AI RECIPES FROM THE DELL AI KITCHEN




Watch a quick video tutorial

## Create a content generation app locally on a Dell Pro Max PC & NVIDIA AI Workbench

Picture this: inference tasks running directly on your device, your data staying secure, and you taking full control of your AI projects. No cloud dependencies, no privacy risks—just secure, unparalleled performance at your fingertips. Everything happens locally, with powerful GPUs and advanced workstations.

With NVIDIA AI Workbench on a Dell Pro Max high-performance PC, it's like having a high-powered AI sous chef right on your desk, ready to turn ideas into custom creations. Perfect for sandbox environments and rapid iterations and finetuning, these tools help you deploy AI-powered apps (e.g.: content generation) locally with confidence and precision. So, let's get cooking!

## What you need to get started

### Dell Pro Max PC

- CPU that supports virtualization
- 16+ GB of RAM
- 80+ GB of free disk space

*Shop Dell Pro Max PCs* →

### NVIDIA Software Suite

- NVIDIA AI Workbench
- NVIDIA NGC account
- Properly installed NVIDIA drivers
- Configured Docker with NVIDIA Container Toolkit (to interface with the GPU)

### Optional tools

- GitHub or GitLab account (for version control)
- Hugging Face API (to access gated LLMs)

# Get Started with NVIDIA AI Workbench

## Steps:

### 1. Set up your environment

Install NVIDIA AI Workbench on your local workstation and configure it according to the setup guide for your OS. Ensure all necessary NVIDIA drivers are installed. Install Docker & configure it to use NVIDIA GPUs by adding the NVIDIA Container Toolkit.

### 2. Clone a GitHub Project

Open NVIDIA AI Workbench and select **"Clone Project.“** Input the repository URL for the hybrid-rag project. Important: Fork the repository first to make a copy under your account. Customize the project name if needed to avoid duplicates, and then click **"Clone."**

### 3. Configure the Environment

Open the cloned project and access the **"Environment"** tab. Use the **"Secrets"** section to add API keys for GitHub, GitLab, or Hugging Face if using external gated models. For this recipe, secrets can remain empty when working with ungated, local models.

### 4. Deploy a Pre-Trained Model

Within the AI Workbench interface, select an ungated AI model from the preloaded options. Configure endpoints, batch sizes, and other settings necessary for your inference tasks. Download the selected model and click **"Start Server"** to begin working with it locally.

### 5. Test your AI application

Input sample data to test the deployed model. For instance, upload a document or provide specific prompts for content generation. Verify the output to ensure the app is functioning as desired.

Tools like NVIDIA AI Workbench, running locally on your Dell Pro Max PC, make it easy to customize workflows and rapidly experiment with different models and configurations.

**EXAMPLE 1 (Generate marketing copy in seconds)**
Turn raw product features into engaging marketing content effortlessly. The model will deliver polished, ready-to-use marketing copy tailored to your requirements.

INPUT EXAMPLE:

```
[Create a customer-facing product description for our new laptop. Highlight its lightweight design, long battery life, and high-resolution touchscreen. Use a positive and professional tone.]
```

**EXAMPLE 2 (Create Visual Content Ideas from Text)**
Simplify ideation for visual projects by generating detailed briefs. Receive precise, creative suggestions for visual communication projects that align with your brand message.

INPUT EXAMPLE:

```
[Suggest three visual storyboard ideas for a video campaign promoting Dell workstations for graphic designers. Highlight themes of innovation, speed, and user-friendly technology.]
```

**EXAMPLE 3 (Produce Research Summaries)**
Condense large quantities of data (whitepapers, research reports, or long-form articles) into manageable and actionable summaries tailored to specific needs such as executive briefs, blog posts, recap summaries. Generate structured and accessible insights from complex materials effortlessly.

INPUT EXAMPLE:

```
[Analyze and summarize the key findings from this academic research article. Present the information in bullet points, focusing on the implications for AI technology development.]
```

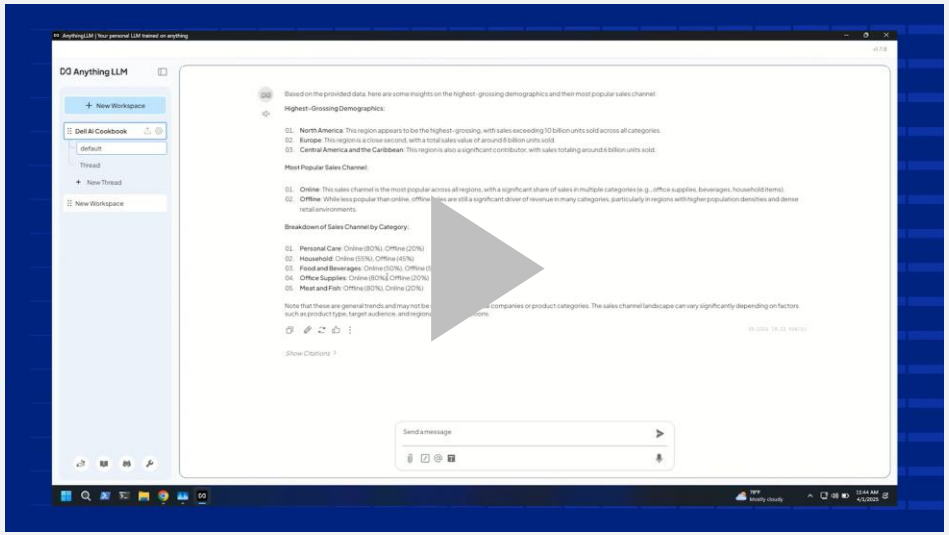
## Learn more about NVIDIA AI Workbench on Dell Pro Max

- Simplifying GenAI development: Dell Pro Max and NVIDIA AI Workbench eBook
- Download NVIDIA AI Workbench here

DELL Technologies NVIDIA

# AI RECIPES FROM THE DELL AI KITCHEN

DELL Technologies Qualcomm




Watch a quick video tutorial

## Run an all-in-one AI application locally with a Dell Copilot+ PCs

Imagine having the power of advanced AI right at your fingertips, no internet required. With Dell Copilot+ PCs powered by Snapdragon X Series processors, you can run tools like AnythingLLM securely and efficiently, all directly on your device. With these PCs equipped with Neural Processing Unit (NPU), tasks are processed in real time, giving you faster speed and more privacy

Whether you're chatting with documents, deploying AI agents, or running powerful models, you'll have everything you need for smarter, more productive work—all while keeping control of your data. It's AI, but on your terms.

## What you need to get started

**Dell Copilot+ PC**

Dell Copilot+ PCs powered by Snapdragon X Series processors are equipped with an NPU running at 45 TOPS that allow you to run AI workloads on-device

*Shop Dell Copilot+ PCs* ⟶

**AnythingLLM**

Run AI Models with an easy to use, all-in-one AI application that can do RAG, AI Agents, and much more with no code setup

*Download AnythingLLM* ⟶

# Get Started with AnythingLLM

## Steps:

### 1. Ensure your PC can handle running local AI models on your device

If you have a PC with dedicated NPU running at 45 TOPS with 16+ GB of RAM, 512+ GB of storage you should be able run AI models locally on your device. With a Snapdragon X Series processor, the dedicated NPU your operations will run with speed and efficiency.

### 2. Download and Install AnythingLLM

Directly download the application from AnythingLLM.com for your operating system

### 3. Choose an AI Model and Select an LLM

Pick an AI model or local LLM provider optimized for the NPU. Select multiple LLMs to use simultaneously across workspaces for flexibility and efficiency.

### 4. Create a Workspace

Upload files, documents, images, and PDFs that LLMs can use to answer your questions pertaining to those resources or general LLM knowledge

### 5. Prompt AnythingLLM

After uploading your resources and preparing your prompt, you can interact with the LLM just like a chatbot. Whether you're asking a simple question or extracting data for calculations, simply prompt the LLM. Here are some examples you can copy, customize, and use to fit your needs

Anything LLM is ready to use with a quick setup on your device. It supports custom models, works seamlessly with all document types, and ensures complete data privacy.

**EXAMPLE 1 (Chat with Documents)**
Easily analyze data or extract valuable insights from uploaded documents without the need for complex tools or formulas

INPUT EXAMPLE:
```
[Using the data in this document, identify which (e.g., product, region, or category) has the best performance based on (specific metric, e.g., growth rate, engagement, or revenue). Provide a summary of the findings.]
```

**EXAMPLE 2 (Summarize Documents with Agents)**
Quickly extract key points from lengthy or complex documents with citations pointing you directly to the original sections of the document where the information was derived. Perfect for summarizing reports, articles, research papers.

INPUT EXAMPLE:
```
[Can you summarize the key findings, highlight any actionable insights, and provide a brief overview for quick reference?
```

**EXAMPLE 3 (Scrape the Web with Agents)**
Leverage agents to efficiently extract and analyze web content, streamlining tasks such as gathering detailed information from websites

INPUT EXAMPLE:
```
[@agent Can you find the menu for [this] restaurant? Organize the dishes by category (e.g., appetizers, main courses, desserts) and provide a summary of the offerings.]
```

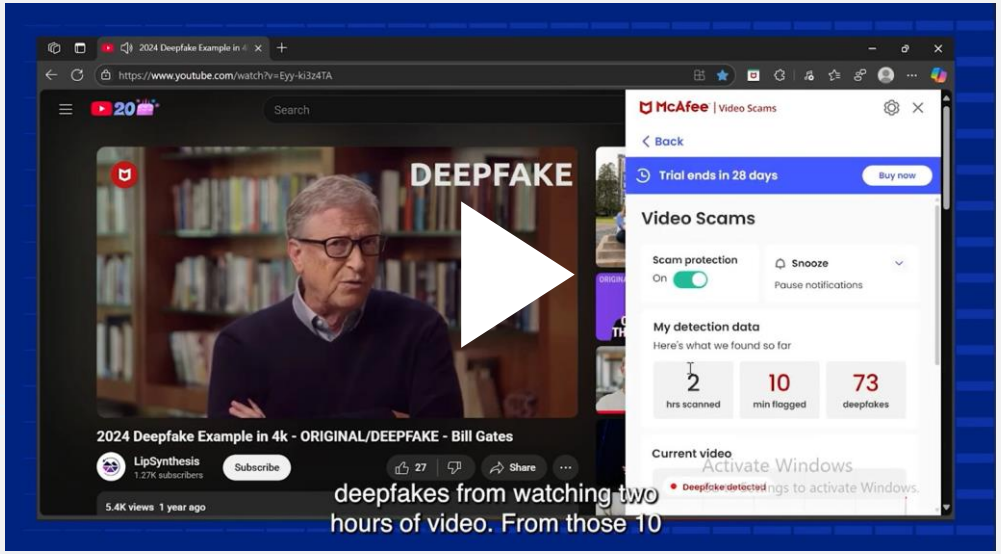
*Agents work in the background to gather data; they may take a bit longer than standard prompt responses to ensure comprehensive results*

## Learn more about Dell Copilot+ PCs and Snapdragon X Series

- Advanced NPUs deliver seamless performance for complex AI tasks.
- Superior energy efficiency maximizes power and minimizes consumption.
- Optimized for on-the-go productivity with power-saving features.
- Runs generative AI models with over 13 billion parameters on-device.

DELL Technologies Qualcomm

# AI RECIPES FROM THE DELL AI KITCHEN




Watch to see how it works

## Unmask Deepfakes with Advanced AI Solutions on your Dell AI PC

Discover how McAfee Deepfake Detection technology and Intel® Core™ Ultra processors on Dell AI PCs help you fight digital deception head-on. This advanced solution uses AI detection techniques and models to analyze audio and video in real-time, instantly flagging AI generated or manipulated content.

Deepfakes are AI-generated fakes that mimic real videos, audio, or images so convincingly they're nearly impossible to identify with the naked eye. These manipulations create serious challenges, undermining trust and causing real-world harm. With tools like McAfee Deepfake Detection, we can defend against these threats, rebuild confidence, and keep authenticity at the forefront of digital interaction

## What you need to get started

### Dell AI PCs

Dell AI PCs that feature an integrated NPU designed to handle AI workloads directly on the device.

*Shop Dell AI PCs* →

### McAfee Smart AI

Use McAfee® Deepfake Detector using advanced AI technology to spot deepfakes for you right in your browser, without any extra clicks.

*Start Using McAfee* →

# Get Started with McAfee Deepfake Detection

## Steps:

### 1. Ensure your PC can handle running local AI models on your device

If you have a Dell AI PC, you run AI models locally on your device. Featuring a dedicated neural processing unit (NPU) delivering anywhere from 11 to 48+ TOPS, these processors bring speed and efficiency to everyday AI tasks running seamlessly in the background.

### 2. Download and Install McAfee

Directly download the McAfee software. Deepfake Detector is included in the McAfee subscription. Set up your scam protection page

### 3. Open your web browser and test McAfee® Deepfake Detector

Find a deepfake fake video online and start playing. Once you have enabled the McAfee Deepfake Detector extension. You can see the that deepfake detector dashboard is checking real time if a deepfake has been detected. Within a few seconds of playing the video, it will flash a watermark to you flagging that AI audio has been detected.

### 4. See the NPU in action

McAfee Deepfake Detection runs locally on Dell AI PCs and sends the AI workload to the NPU, that sustained workload is now freeing up your CPU and GPU so there is little effect on your overall performance, saving you battery while running the background. Open task manager, use the side bar on the left to navigate to performance, and watch the NPU work while playing the video.

The NPU uses

**48% less power**

when running McAfee's AI-powered Deepfake Detector to efficiently identify misleading videos and audio vs the CPU alone.[1]

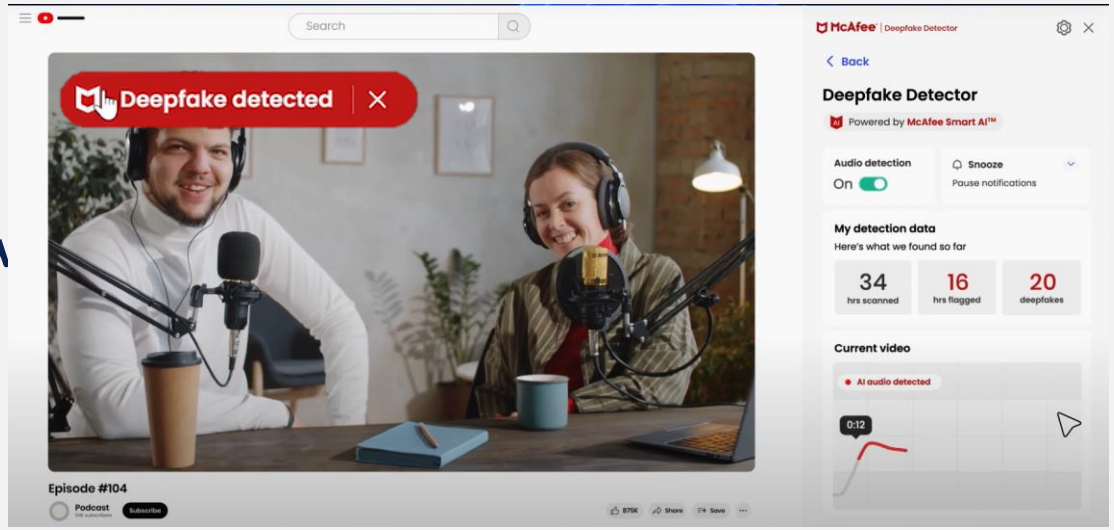
*1 Based on internal analysis and testing, November 2024. McAfee's Deepfake Detector running on devices with Intel® Core™ Ultra 200V series processors was compared to previous-generation Intel® Core™ Ultra processors. Results may vary.*

**EXAMPLE (Test Deepfake Detection)**

Play any deepfake video (ex: Deepfake Example of Bill Gates)

Behind the scenes, McAfee Deepfake Detection is monitoring what you are watching and will be alerted immediately when the deepfake video is detected showing you that portion of the video has been manipulated.

**McAfee Deepfake Detection Dashboard Example**




## Learn more about Dell AI PCs powered by Intel® Core™ Ultra

- Intel® Core™ Ultra processors deliver a dedicated engine to help possibilities on the PC
- Available with built-in Intel® Arc™ GPU.1 2 Created using the new Intel 4 process, Intel® Core™ Ultra processors deliver an optimal balance of performance and power efficiency
- Learn the top 4 reasons to refresh to Dell AI PCs powered by Intel® Core™ Ultra processors

DELL Technologies intel CORE ULTRA