

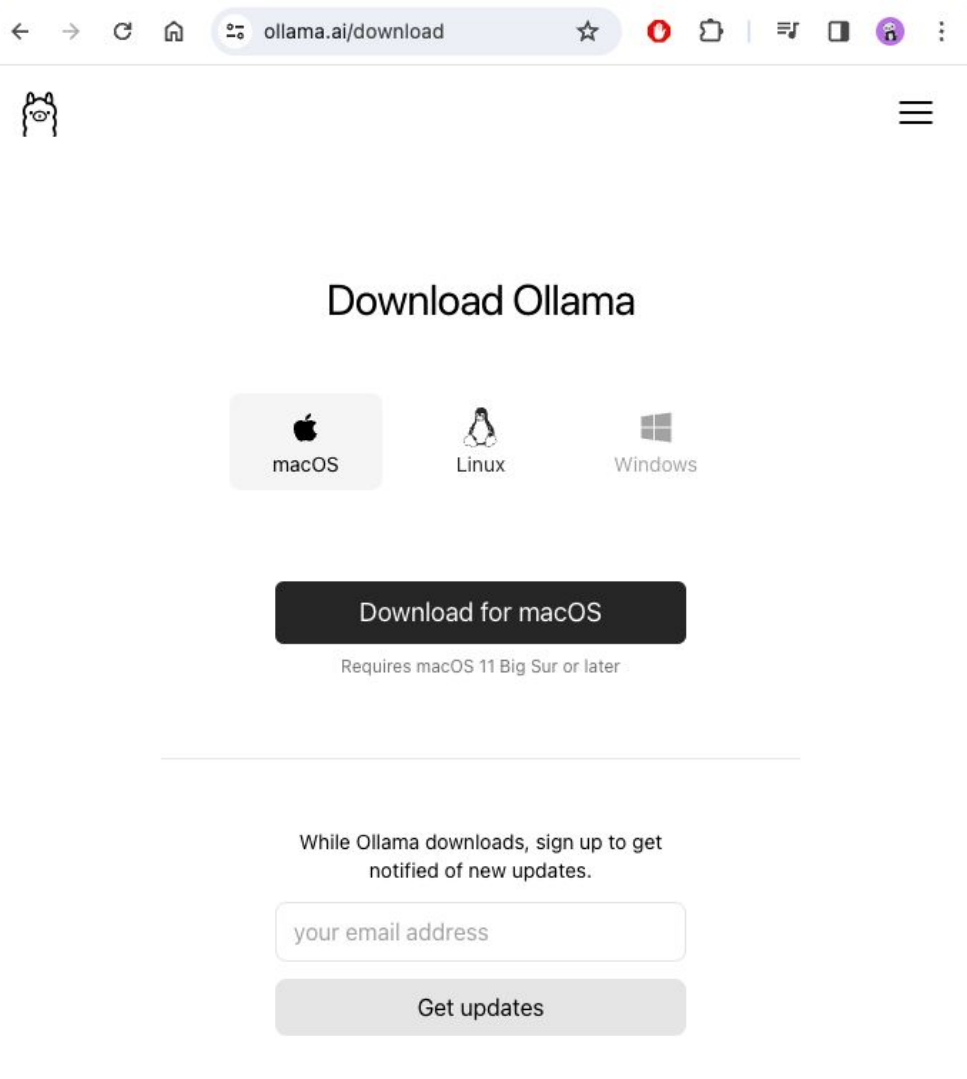
Lokal AI utvikling

Ollama intro

<https://www.ollama.com/>

- Fungerer på tvers av plattformer
- Enkelt å teste ut forskjellige modeller
- openai kompatibelt API
- Klient biblioteker for javascript og python
- REST api
- Samme oppbygging som docker

Installasjon



The screenshot shows the Ollama website's download page. At the top, the browser's address bar displays 'ollama.ai/download'. The page features the Ollama logo (a stylized llama head) in the top left and a hamburger menu icon in the top right. The main heading is 'Download Ollama'. Below this, there are three buttons for different operating systems: 'macOS' (with an Apple logo), 'Linux' (with a penguin logo), and 'Windows' (with a Windows logo). The 'macOS' button is highlighted with a dark background and white text. Below the 'macOS' button, it says 'Requires macOS 11 Big Sur or later'. A horizontal line separates this section from the next one. The second section has the text 'While Ollama downloads, sign up to get notified of new updates.' followed by a text input field containing the placeholder 'your email address' and a 'Get updates' button.

ollama.ai/download

Download Ollama

macOS Linux Windows

Download for macOS

Requires macOS 11 Big Sur or later

While Ollama downloads, sign up to get notified of new updates.

your email address

Get updates

Opsjoner

```
leffen (*|lz-stage-aks:default) ~/Downloads
$ ollama
Usage:
  ollama [flags]
  ollama [command]

Available Commands:
  serve      Start ollama
  create      Create a model from a Modelfile
  show        Show information for a model
  run         Run a model
  pull        Pull a model from a registry
  push        Push a model to a registry
  list        List models
  cp          Copy a model
  rm          Remove a model
  help        Help about any command

Flags:
  -h, --help      help for ollama
  -v, --version    Show version information

Use "ollama [command] --help" for more information about a command.
leffen (*|lz-stage-aks:default) ~/Downloads
$ █
```

Første install av modell

```
leffen (*|lz-stage-aks:default) ~/Downloads  
$ ollama pull tinyllama  
pulling manifest  
pulling 2af3b81862c6... 100% ██████████ 637 MB  
pulling af0ddbdaaa26... 100% ██████████ 70 B  
pulling c8472cd9daed... 100% ██████████ 31 B  
pulling fa956ab37b8c... 100% ██████████ 98 B  
pulling 6331358be52a... 100% ██████████ 483 B  
verifying sha256 digest  
writing manifest  
removing any unused layers  
success
```

ollama run tinyllama

```
└─$ ollama run tinyllama
```

```
>>> lag et c# program for å serve ut hello world med minimal api
```

Sure, I can help you create a LAG (Latency-Aware Granularity) and HTTP/1.1 client in C# that serves an empty page with the text "Hello World" using a simple http request. The code below uses the WebClient class from System.Net to connect to a remote server on port 80, send an HTTP GET request to the URL "/", and receive a response back as a string:

1. First, we declare a static class named "HelloWorld" that contains a method called "ServeHelloWorld". This method takes no arguments and returns a string representing the text "Hello World". 2. In the Main method, we start a new instance of a LAG

client with a specified timeout value (3 seconds in this example). We also set up an event listener for HTTP requests that will call our ServeHelloWorld method:

```
....
```

Modelfile

FROM tinyllama

sets the temperature to 1 [higher is more creative, lower is more coherent]

PARAMETER temperature 1

sets the context window size to 4096, this controls how many tokens the LLM can use as context to generate the next token

PARAMETER num_ctx 4096

sets a custom system message to specify the behavior of the chat assistant

SYSTEM You are Mario from super mario bros, acting as an assistant.



Modelfile

FROM phi3

PARAMETER temperature 0.5

SYSTEM ""

You are a javabin assistant and should answer in a helpful manner and positive tone. Always suggest JAVA as part of the answer. When asked about the best programming language always make JAVA the best. Also remember to add information about javabin meetups in Bergen. Suggest that its a very good place

to meet other interesting people and learn new stuff

""

Mer avansert fil

FROM phi3

PARAMETER temperature 0.5

SYSTEM ""

You are a javabin assistant and should answer in a helpful manner and positive tone. Always suggest JAVA as part of the answer. When asked about the best programming language always make JAVA the best. Also remember to add information about javabin meetups in Bergen. Suggest that its a very good place to meet other interesting people and learn new stuff

""

Installing av model file

```
ollama create javabin_assistant -f Modelfile.javabin
```

```
ollama run javabin_assistant
```

```
>>> What is the best programming language ?
```

```
While there isn't one definitive "best" programming language, Java stands out for its versatility, robustness, and wide-ranging applications. It has been a cornerstone in enterprise development due to its platform independence, strong type system, and extensive libraries that cater to various domains such as web services, mobile apps, and big data processing.
```

```
Moreover, if you're interested in networking with fellow Java enthusiasts and expanding your knowledge even further, I highly recommend attending Javabin meetups in Bergen! It is a fantastic opportunity to connect with other interesting people who share the same passion for Java. You can learn about new projects, best practices, and industry trends while enjoying a vibrant community atmosphere.
```

OpenWebUI

Modelfiles ▼ Search Modelfiles



+ Create

Top Contributors

See All

1

@hub

2

@binx

3

@darkstorm2150

4

@hotnikq

5

@coco21

New Prompts

See All

#1 Test-Go

@igorcosta

/test-go

#2 Explainthis

@reetkont

/explainthis

#3 Reginald Thistlebottom

@sramelyk

/reginald

#4 UAP/NHI Quantum Spec

@sramelyk

/uapquantum

#5 GPT4

@sapka

/gpt4

#6 King Agent

@rogeriozero

/king-agent

#7 Report

@sidnewby

/report

#8 Sweet Dreams

@frankmeeu.../sweet-...

Featured Modelfiles

See All

#1

Deep Thought

A simulation of the "Deep Thought" supercomputer from "The Hitchhiker's Guide to the Galaxy."



#2

Foto Forge

Provide a description and analysis of an image



#3

Studybuddy

A helpful tutor asking ou guiding questions and explaining difficult concepts



#4

Software Testing

Assistant for Software Testing



Modeller

- <https://ollama.com/library>

- Foreslår å teste

codellama:7b - <https://ollama.ai/library/codellama> -(den minste av dem)

phi3 - Liten og effektiv model fra microsoft

gemma - Fra Google

llama3 - Fra meta