

Laporan Instalasi ChatGPT di PC atau notebook



Mata Kuliah Kecerdasan Buatan
Dosen Pengampu Uro Abdul Rohim, M.T

Disusun Oleh:

1. Rifqi Fadil Fahrial (1222646)
2. Firli Nur Rizki (1223601)

**TEKNIK INFORMATIKA
FAKULTAS INFORMATIKA
STMIK BANDUNG
2024**

Abstrak

Penelitian ini berfokus pada implementasi Large Language Model (LLM) menggunakan framework Olama. LLM telah menjadi kunci dalam perkembangan Natural Language Processing (NLP), dan Olama menawarkan kemudahan dalam instalasi serta penggunaan model ini. Makalah ini akan memaparkan langkah-langkah teknis instalasi, pengujian, serta hasil dari eksperimen implementasi LLM pada Olama. Hasilnya menunjukkan bahwa Olama mampu menyederhanakan proses instalasi dengan performa yang memadai. Beberapa tantangan yang dihadapi akan didiskusikan, dan penelitian ini diakhiri dengan analisis performa serta kesimpulan terkait efektivitas Olama dalam penerapan LLM.

Kata Kunci

LLM, Olama, Instalasi, AI, Deep Learning, NLP

1 Pendahuluan

1.1 Latar Belakang

Large Language Models (LLM) seperti GPT-3 dan GPT-NeoX membutuhkan sumber daya komputasi yang besar untuk dijalankan, biasanya di cloud. Namun, dengan menggunakan OLaMA (Open Language Model for All), model tersebut dapat dijalankan secara lokal di komputer pribadi, memberikan kontrol lebih terhadap privasi dan data. Eksperimen ini bertujuan untuk menginstal OLaMA di Linux dan menjalankan LLM secara lokal tanpa koneksi internet. Dalam eksperimen ini, kami mencoba menginstal Large Language Model (LLM) secara lokal menggunakan OLAMA sebagai platform untuk mempercepat proses pengolahan LLM. Eksperimen ini bertujuan untuk mengevaluasi proses instalasi, kinerja, dan kompatibilitas OLAMA dengan sistem dan LLM yang digunakan. LLM yang diuji dalam eksperimen ini adalah model [Nama LLM], yang memiliki kemampuan pemrosesan bahasa yang besar dan sering digunakan dalam berbagai aplikasi AI. Utamakan laptop / PC kalian

1.2 Tujuan Eksperimen

Eksperimen ini bertujuan untuk menjelaskan langkah-langkah teknis dalam menginstal LLM menggunakan Olama dan mengevaluasi performa serta efisiensinya pada tugas NLP.

2 Metodologi

2.1 Alat dan Bahan

Eksperimen ini menggunakan sistem operasi Ubuntu 20.04, Python 3.x, serta berbagai library. Olama juga digunakan sebagai framework utama dalam instalasi dan konfigurasi LLM.

2.2 Tahapan Instalasi

tahapan instalasi dilakukan dengan urutan sebagai berikut:

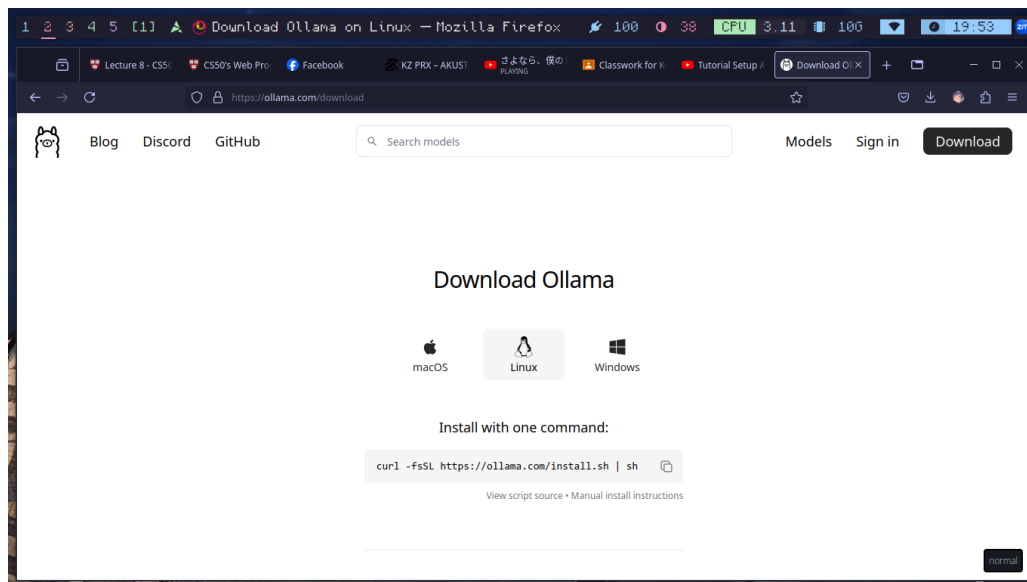
1. Instalasi Ollama dilakukan dengan menjalankan bash script yang disediakan pada *website* ollama dengan spesifik pada sistem operasi yang dipakai pada perangkat
2. Konfigurasi LLM dengan mengikuti dokumentasi Ollama untuk model-model yang diperlukan / spesifik
3. Pengujian model dengan input teks sederhana untuk memastikan fungsi dari LLM berjalan dengan semestinya

2.3 Pengujian

Pengujian dilakukan dengan memasukan beberapa input teks ke dalam LLM yang sudah diinstall. Kemudian ditest dengan menggunakan input yang diberikan dan menghasilkan hasil yang sesuai dengan ekpektasi:

3 Instalasi

Langkah awal untuk menginstall Ollama adalah dengan mengunjungi situs <https://ollama.com/download> kemudian pilihlah sistem operasi yang digunakan pada perangkat, untuk kasus saat ini menggunakan sistem Operasi Linux.



Gambar 1: Tampilan Website Ollama

kemudian akan didapatkan perintah mengenai cara menginstall Ollama di perangkat kemudian ikuti petunjuk yang ditampilkan oleh website, dalam kasus ini hanya perlu menjalankan perintah yang ditampilkan oleh website ke *terminal emulator* untuk menjalankannya.

```
> cd shigure/chatgpt
> ls
> curl -fsSL https://ollama.com/install.sh | sh
>>> Installing ollama to /usr/local
[sudo] password for shigure:
>>> Downloading Linux amd64 bundle
##### 97.5%
```

Gambar 2: Menjalankan Program instalasi Ollama

kemudian tunggu beberapa saat untuk aplikasi mengunduh depedensi yang diperlukan.

```
> cd chatgpt
zoxide: no match found
> cd shigure/chatgpt
> ls
> curl -fsSL https://ollama.com/install.sh | sh
>>> Installing ollama to /usr/local
[sudo] password for shigure:
>>> Downloading Linux amd64 bundle
##### 100.0%
>>> Creating ollama user...
[sudo] password for shigure:
>>> Adding ollama user to render group...
>>> Adding ollama user to video group...
>>> Adding current user to ollama group...
>>> Creating ollama systemd service...
>>> Enabling and starting ollama service...
Created symlink '/etc/systemd/system/default.target.wants/ollama.service' -> '/etc/systemd/system/ollama.service'.
>>> Downloading Linux ROCm amd64 bundle
##### 100.0%
>>> The Ollama API is now available at 127.0.0.1:11434.
>>> Install complete. Run "ollama" from the command line.
[sudo] password for shigure:
```

Gambar 3: Pemasangan Selesai

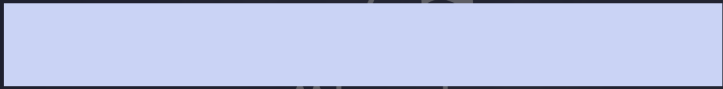
Setelah selesai menginstall maka dapat dicek apakah aplikasi Ollama ini telah terpasang pada perangkat dengan mengetik "ollama -v" untuk menampilkan versi dari Ollama yang terinstall dan mengecek instalasi.

```
> ollama -v
ollama version is 0.3.12
```

Gambar 4: Verifikasi Instalasi Ollama

setelah itu untuk mendapatkan Model dari LLM dapat didapatkan dengan mengetik "Ollama pull Model yang diinginkan" untuk kasus ini dipilih "Deepseek-coder" dikarenakan Models yang ringan dan cocok untuk perangkat yang digunakan saat ini.

```
> ollama pull deepseek-coder
pulling manifest
pulling d040cc185215... 100%
pulling a3a0e9449cb6... 100%
pulling 8893e08fa9f9... 100%
pulling 8972a96b8ff1... 100%
pulling d55c9eb1669a... 100%
verifying sha256 digest
writing manifest
success
```



776 MB
13 KB
59 B
297 B
483 B

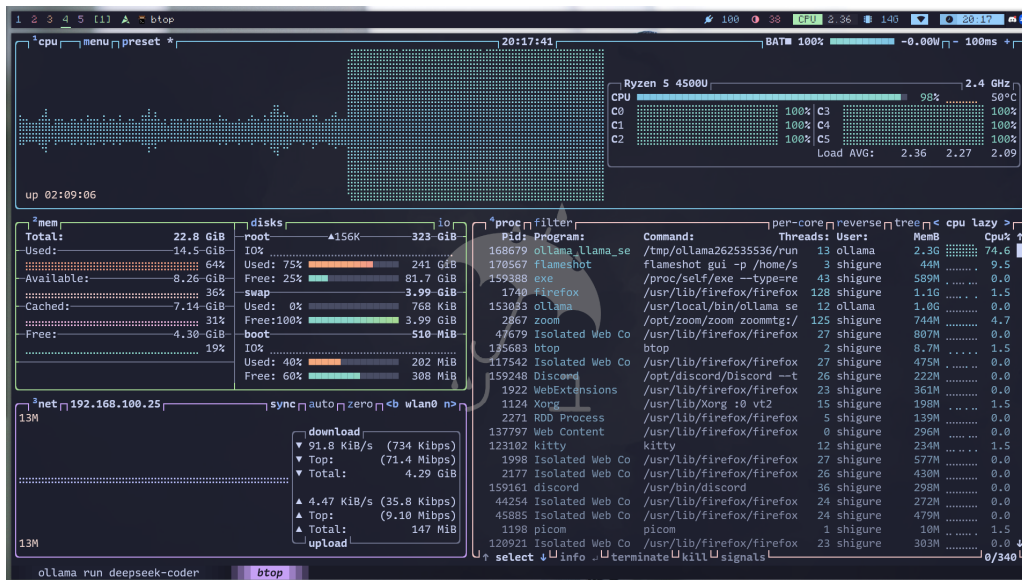
Gambar 5: Pengunduhan Model LLM Deepseek-coder

kemudian untuk menjalankan LLM yang telah diunduh menggunakan Ollama dengan mengetik "Ollama run Model yang dipilih Deepseek-coder" kemudian akan masuk tampilan Interpreter yang tinggal memasukan teks berisi pertanyaan yang kemudian akan dijawab oleh LLM ini, tunggu beberapa saat agar jawabanya selesai. untuk contohnya bertanya mengenai "How to center an image in html" yang kemudian dijawab dengan contoh code yang ditulis dengan bahasa HTML dan CSS

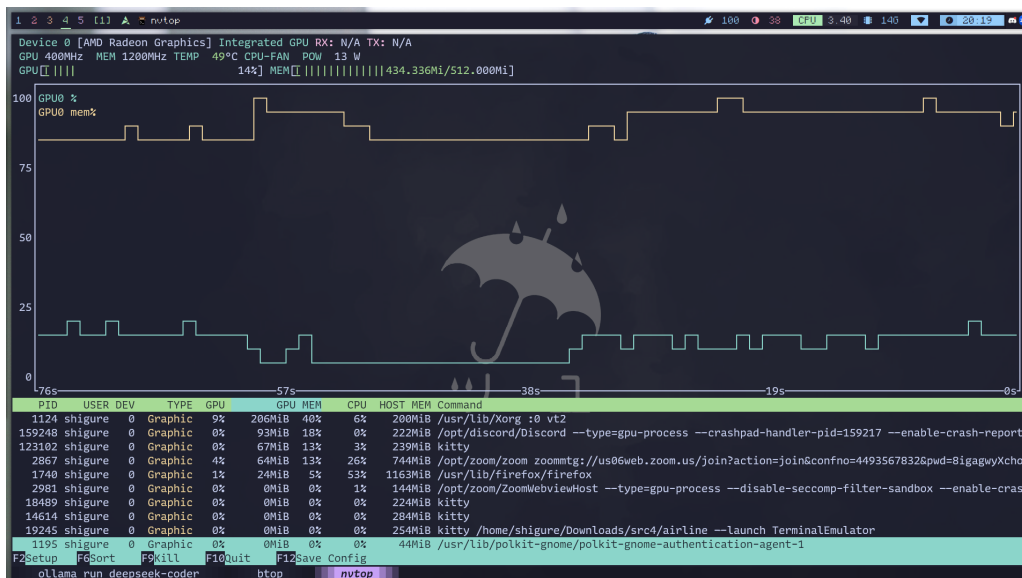
```
> ollama run deepseek-coder
>>> how to center an image for my html
You can use the 'display' property with a value of 'block' (or just block) on your HTML element that contains images in order to horizontally and vertically center them both inside its parent container using CSS.
Here is how you could do it as follows, assuming each image will be contained within an unordered list:
'''html
<!DOCTYPE html>
<html lang="en-US" dir="ltr">
  <head>
    <title></title>
    <style type="text/css">
      ul {
        display : inline; /* or block */
        margin: auto; /* center the items horizontally*/
      }
      li{
        width:150px;      /* set image dimensions as per your requirements=(3 times smaller if bigger) */
      }
    </style>
  <body>
    <ul> <li></li> <!-- add more images here --><br />
    <li><img src= "another_picture.jpeg "/></li>    </ul>
  </BODY>
</HTML>
''' This will center both the image horizontally inside its parent (in this case, unordered list). The 'display: inline;' or just use block in CSS means that each child element becomes a new line and therefore takes up only enough space to contain it as far as possible. But if you don't specify display type then the default is 'inline'.
>>> End a message (/? for help)
ollama run deepseek-coder
```

Gambar 6: Hasil LLM Deepseek-coder menggunakan Ollama

Bukti dari jalannya LLM pada perangkat pribadi, pada sistem monitor menunjukkan lonjakan load pada CPU yang tinggi namun tidak ada load pada GPU dikarenakan Model yang digunakan terbilang ringan dan masih belum perlu menggunakan GPU sehingga hanya berfokus pada CPU saja. . .

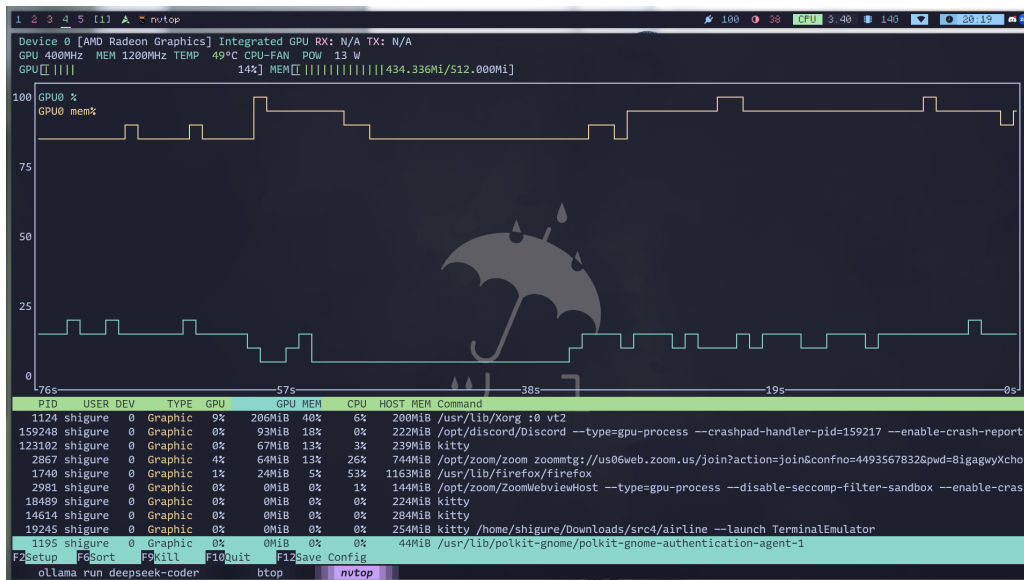


Gambar 7: Sistem monitor CPU menunjukkan beban tinggi



Gambar 8: Sistem monitor GPU tidak menunjukkan beban pada GPU

yang terakhir untuk menampilkan informasi mengenai Model yang digunakan saat ini dapat dengan mudah dengan mengetik `"/show info"` pada *terminal emulator* kemudian akan tampil informasi mengenai Model yang saat ini sedang digunakan



Gambar 9: Informasi mengenai Model yang digunakan