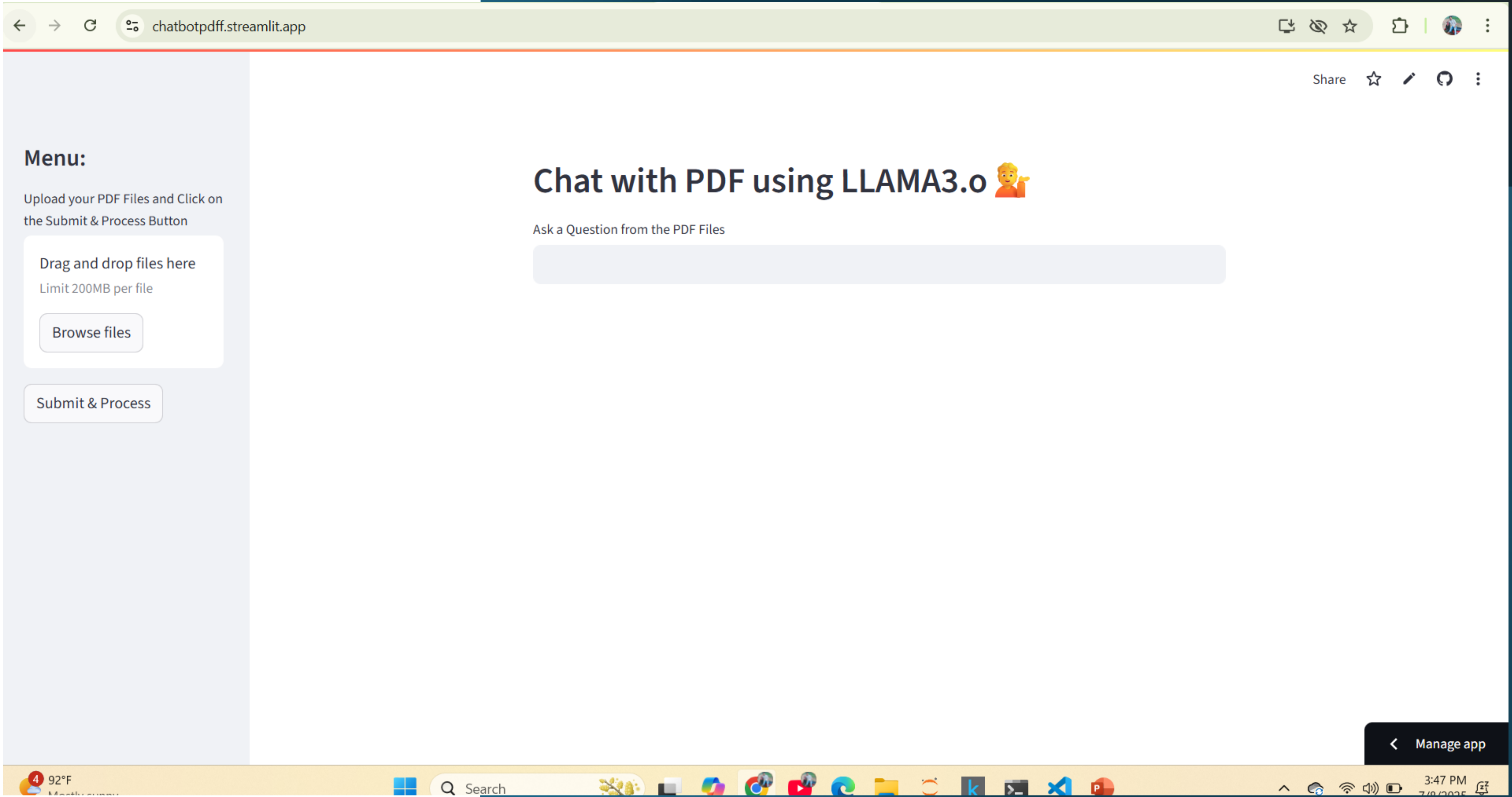


Menu:

Upload your PDF Files and Click on the Submit & Process Button

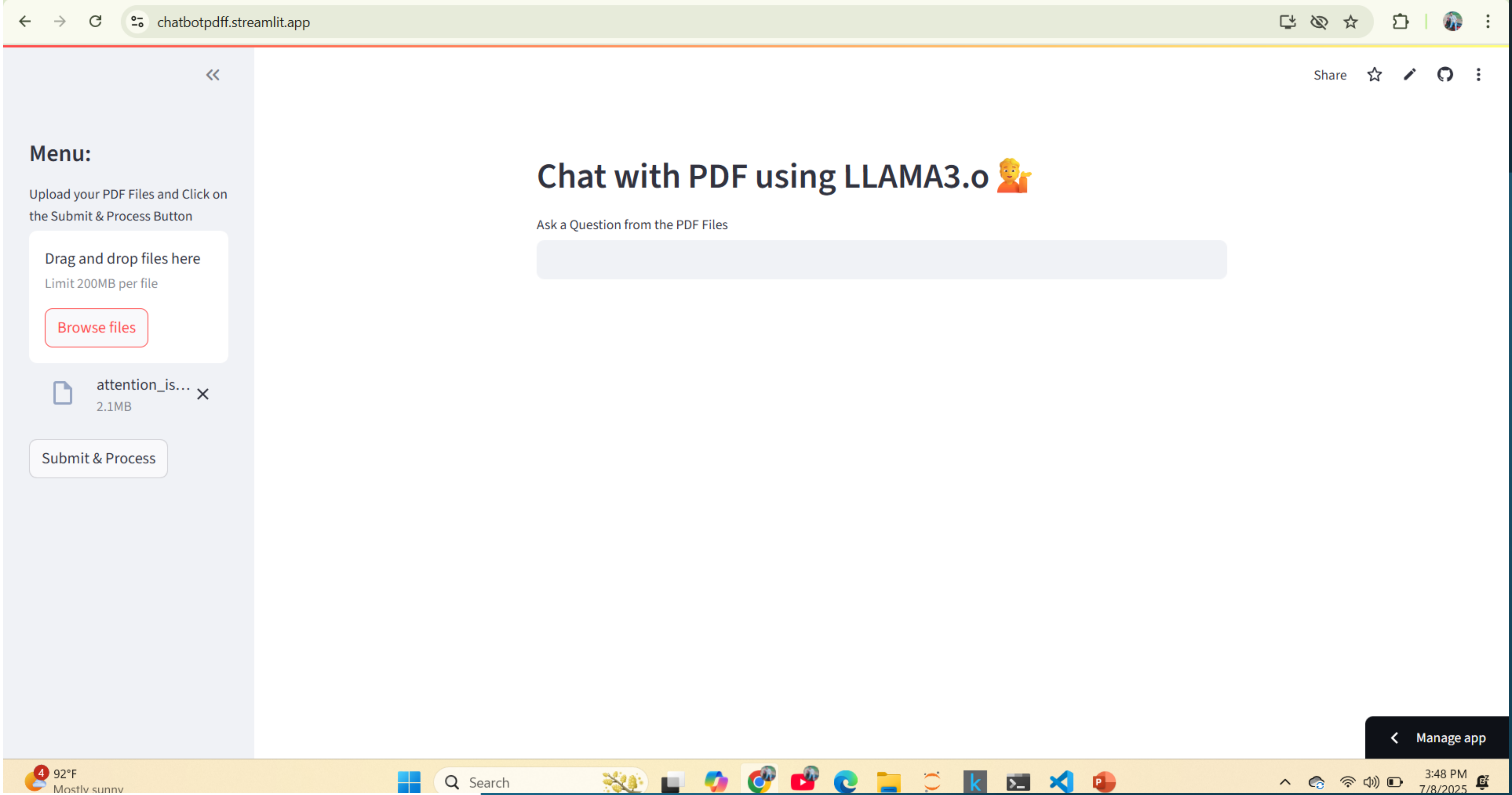
Drag and drop files here

Limit 200MB per file

Chat with PDF using LLAMA3.o 💁

Ask a Question from the PDF Files

Submit your file by clicking on Browse Files



chatbotpdff.streamlit.app

Chat with PDF using LLAMA3.o 💁

Ask a Question from the PDF Files

Menu:

Upload your PDF Files and Click on the Submit & Process Button

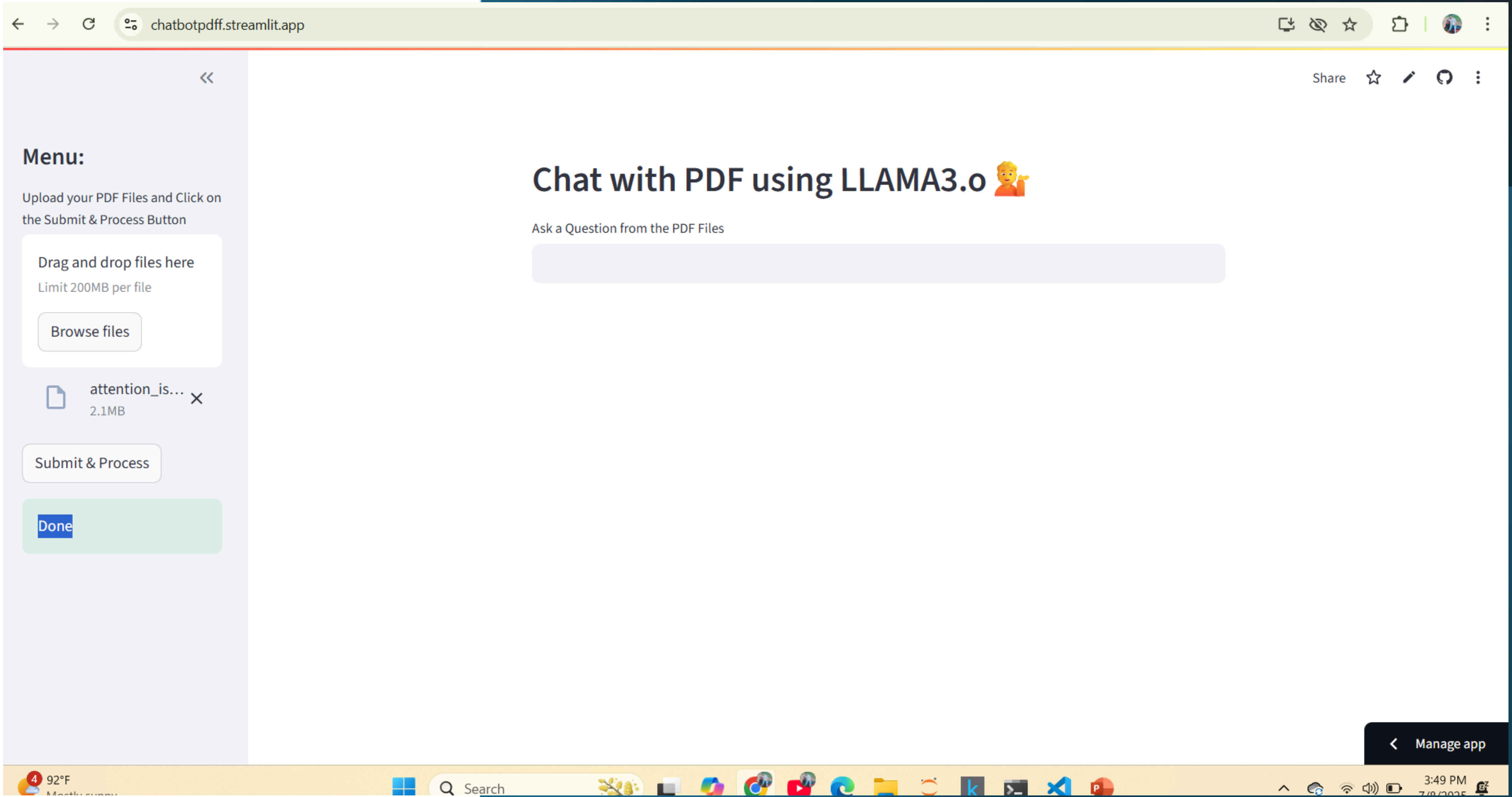
Drag and drop files here

Limit 200MB per file

attention_is…

2.1MB

Click on Submit & Process



chatbotpdff.streamlit.app

Menu:

Upload your PDF Files and Click on the Submit & Process Button

Drag and drop files here

Limit 200MB per file

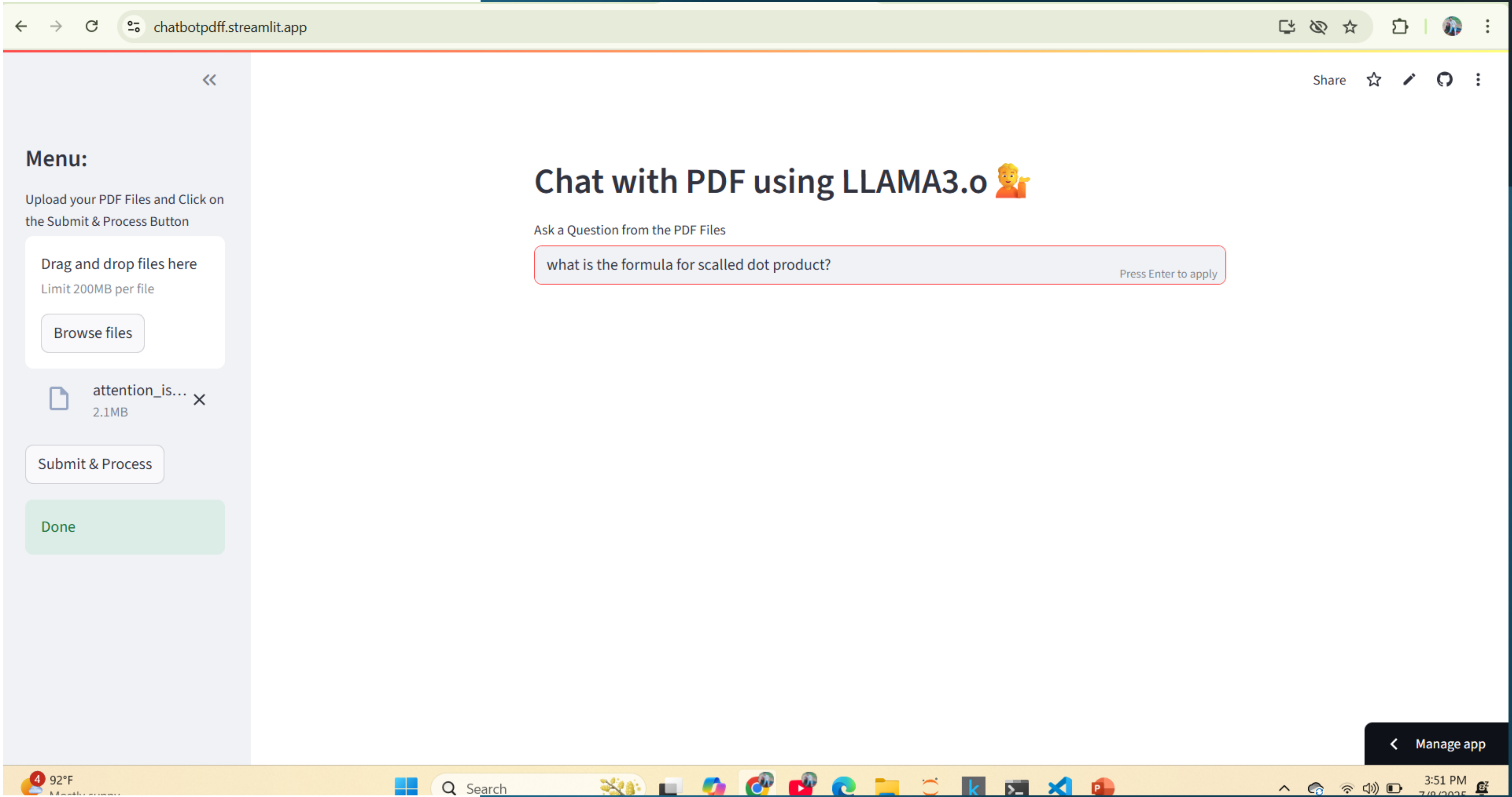
attention_is…
2.1MB

Done

Chat with PDF using LLAMA3.o

Ask a Question from the PDF Files

Ask Question to know from PDF



Menu:

Upload your PDF Files and Click on the Submit & Process Button

Drag and drop files here
Limit 200MB per file

attention_is…
2.1MB

Done

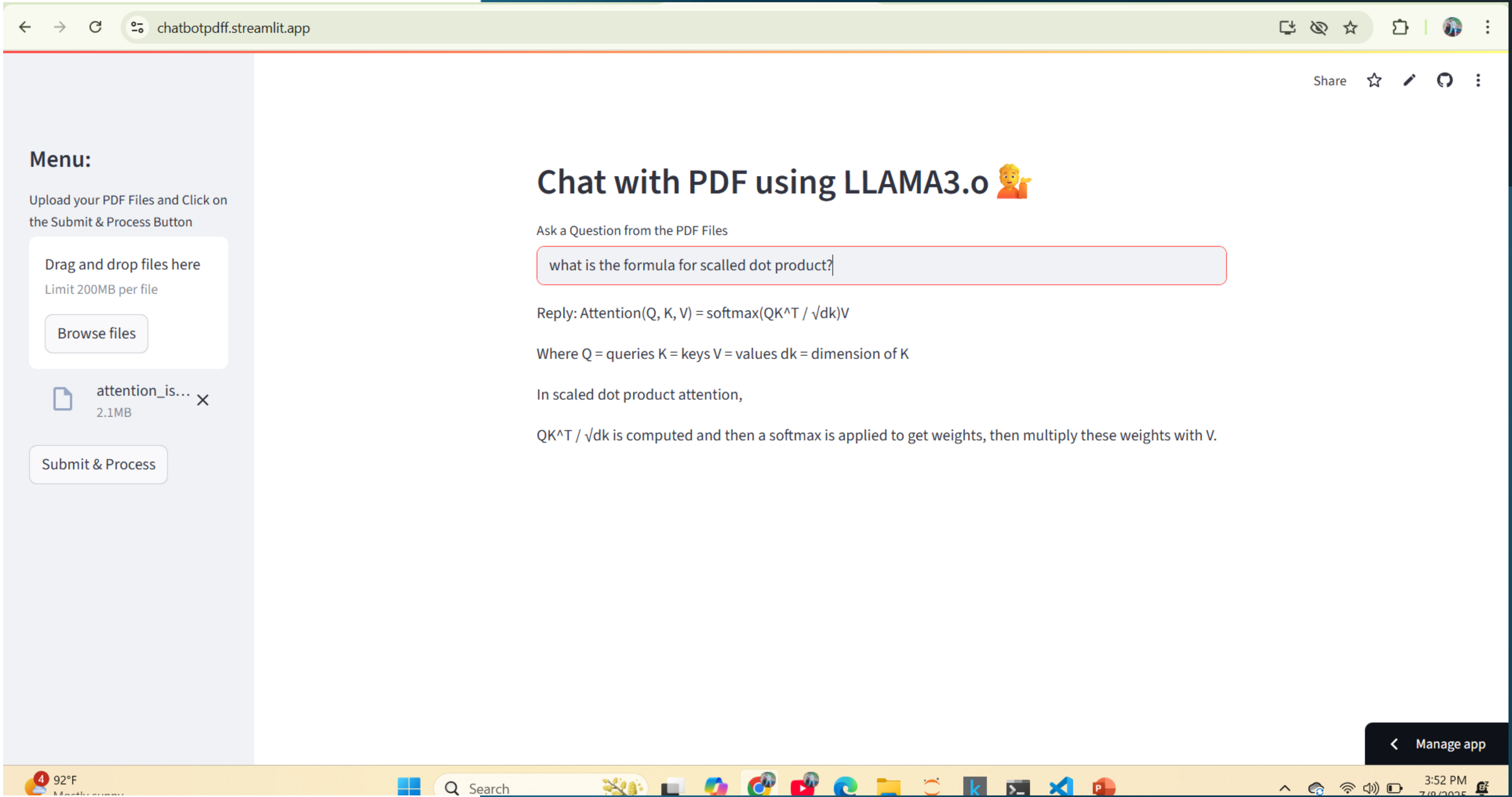
Chat with PDF using LLAMA3.o 💁

Ask a Question from the PDF Files

what is the formula for scalled dot product?

Press Enter to apply

Response Generated



Menu:

Upload your PDF Files and Click on the Submit & Process Button

Drag and drop files here

Limit 200MB per file

Chat with PDF using LLAMA3.o 💁

Ask a Question from the PDF Files

what is the formula for scalled dot product?

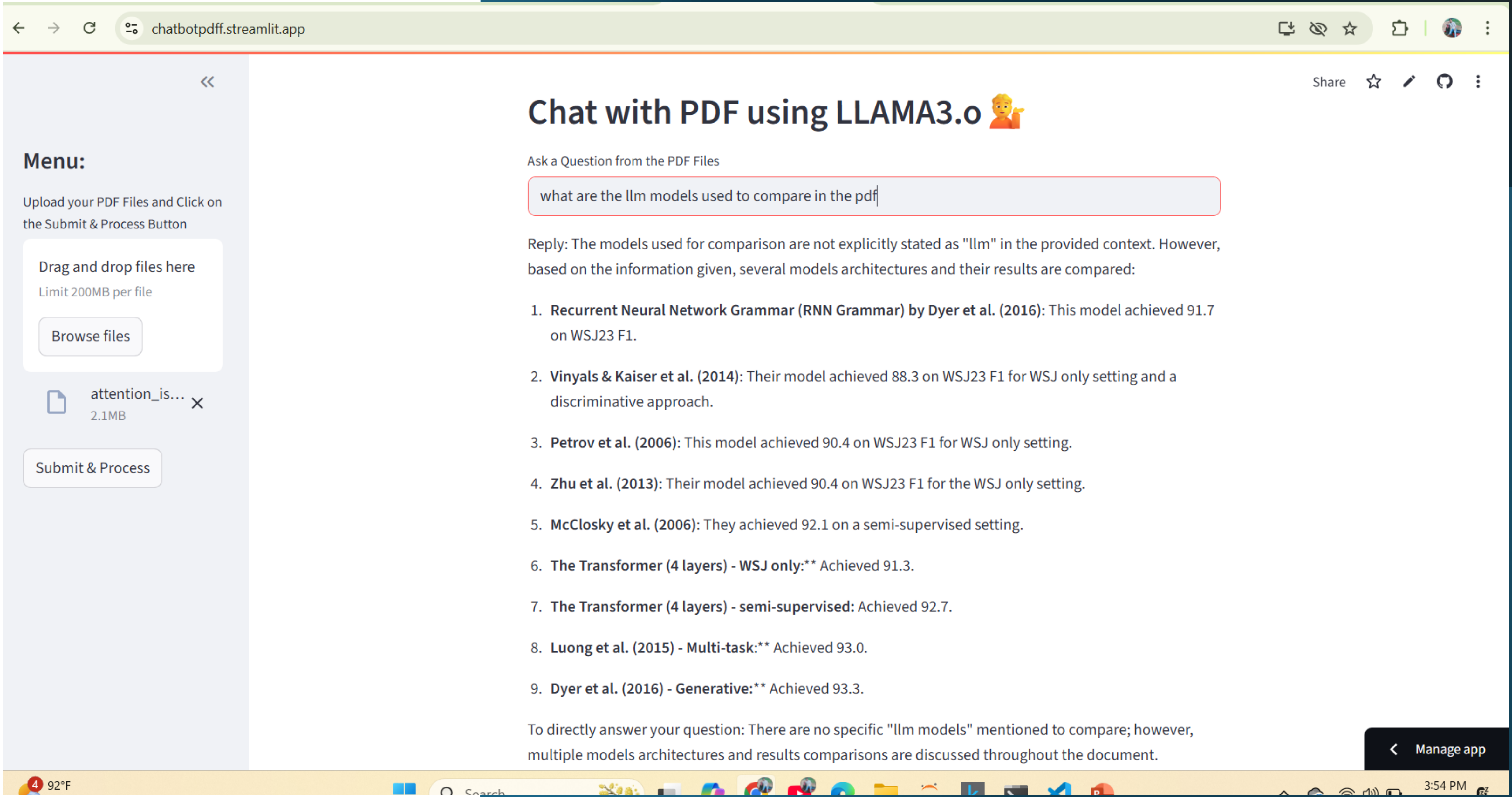
Reply: Attention(Q, K, V) = softmax(QK^T / √dk)V

Where Q = queries K = keys V = values dk = dimension of K

In scaled dot product attention,

QK^T / √dk is computed and then a softmax is applied to get weights, then multiply these weights with V.

Sample Question 2



Upload your PDF Files and Click on the Submit & Process Button

Chat with PDF using LLAMA3.o 💁

Ask a Question from the PDF Files

what are the llm models used to compare in the pdf

Reply: The models used for comparison are not explicitly stated as "llm" in the provided context. However, based on the information given, several models architectures and their results are compared:

1. **Recurrent Neural Network Grammar (RNN Grammar) by Dyer et al. (2016)**: This model achieved 91.7 on WSJ23 F1.
2. **Vinyals & Kaiser et al. (2014)**: Their model achieved 88.3 on WSJ23 F1 for WSJ only setting and a discriminative approach.
3. **Petrov et al. (2006)**: This model achieved 90.4 on WSJ23 F1 for WSJ only setting.
4. **Zhu et al. (2013)**: Their model achieved 90.4 on WSJ23 F1 for the WSJ only setting.
5. **McClosky et al. (2006)**: They achieved 92.1 on a semi-supervised setting.
6. **The Transformer (4 layers) - WSJ only:**** Achieved 91.3.
7. **The Transformer (4 layers) - semi-supervised:** Achieved 92.7.
8. **Luong et al. (2015) - Multi-task:**** Achieved 93.0.
9. **Dyer et al. (2016) - Generative:**** Achieved 93.3.

To directly answer your question: There are no specific "llm models" mentioned to compare; however, multiple models architectures and results comparisons are discussed throughout the document.