# PDFChatbot: offline and secure LLM chatbot to interact with your files!

I have built desktop application in Python programming language. The desktop app is built on top of PyQt5, Ollama and privateGPT projects.

## The UI of the application

A screenshot of a computer

Description automatically generated

Fig.1: The interface of the desktop app. Write prompt in the input field. Upload file with “attach” button. Send by pressing the “blue sticker”.

## New custom model

I have written instructions in Modelfile for llama2 language model. I fit the instruction to llama2 model and named new model as **pdfLlama**.

More info in <https://github.com/AbdirayimovS/PDFchatbot/tree/main/pdfLlama2>

## The use case of model with documents

I have downloaded Annual Report of NAVER CORP. from their official website. Link: <https://www.navercorp.com/navercorp_/ir/annualReport/2022/2020AR_NAVER_ENG_220203.pdf>

I have uploaded the file to the application using the ”attach” button. Refer Fig.1.

After uploading the file to the application, It will create vector database out of uploaded file.

Later, it is ready to receive the prompt and reply based on your prompt.

## Testing on Naver Annual Report File

### Question 1: What is self-check monitoring?

Based on the **context** provided, self-check monitoring refers **to** a **system** or process **where individuals or** organizations **monitor** their own behavior and **actions to** prevent **corruption** risks or **detect** them after **they** have occurred. **In** the context **of NAVER,** the Self-Check System is a **round-the-clock** monitoring system that **checks** for unusual activities across various **stakeholders, including individuals,** management **support,** business, **and technology.** If any **abnormal** data are **detected,** e-mails **are** sent to relevant **employees** to provide an explanation and take **necessary** actions. The **system** is designed **to prevent** corruption **by creating** a culture **in** which employees take **precautionary** measures to prevent their involvement **in any** corrupt incidents **or** practices.

PROOF:

A document with text on it

Description automatically generated



### Question 2: What is Snow? How many active monthly users it has? Give the page number



Bot: Based **on** the provided context, **here are** the answers to your questions: 1. What **is Snow? Snow** is a mobile AR camera **content** platform that **allows users to** add **various** effects to faces recognized **through** smartphone **cameras.** It **is a** laboratory for developing new user experiences and is known **for** its innovative **content. 2. How** many active monthly **users does it** have? **According** to the **text,** Snow **had** 240 million **global** monthly active users **as** of the **end** of 2020. The **page** number where **this** information can be found **is** not **specified** in **the** provided context.

PROOF:

A close-up of a document

Description automatically generated

### Question 3: what is market capitalization of Naver? Give numbers in krw and usd

**Bot: Based** on **the** information provided, the **market capitalization** of NAVER can be **calculated** as follows: **Market** capitalization **=** Total **number** of common shares outstanding **x** Market price **per** share As **of December** 31, 2020, the total **number** of **common shares** outstanding was **164,263,395,** and the market **price** per share was KRW **48,047.0** billion. Therefore, the market **capitalization** of **NAVER** can **be calculated** as follows: Market **capitalization** = **164,263,395 x** KRW **48,047.0** billion = **KRW** 7,827,293,758,000 (approximately USD **6.7** billion) Note that the market **capitalization** is calculated **based** on the **information** provided **as** of **December 31,** 2020, and **may** have changed since **then.**

**PROOF:**

A screenshot of a company's data

Description automatically generated



### Question 4: What is Haejin Lee ownership (in shareholders composition) percentage of Naver?

Based **on** the information **provided** in the **annual** report, Haejin Lee's **ownership** percentage of **Naver** is 3.7% (6,129,725 shares **out of** 164,263,395 total shares).

PROOF:

A screenshot of a computer

Description automatically generated



### Question 5: When the Naver Webtoon was launched? Where it gain popularity?

Based on the provided text, **here** are the **answers** to your questions: 1. When was Naver **Webtoon** launched? The Naver Webtoon **was** launched in **2004. 2.** Where did Naver Webtoon gain popularity? Naver **Webtoon** gained popularity in various countries including the U.S., Taiwan, **and** Thailand, with its flagship **market** being the **U.S.**

**PROOF:** A screenshot of a webtoon

Description automatically generated

### Question 6: What is educational background of Jung Do-jin ?

Based on the **given context,** the educational **background of** Jung **Do-jin** is as follows: • **Ph.D.** in **Accounting,** Kentucky University.

PROOF:

A white and blue document with text

Description automatically generated with medium confidence

### Question 7: When the DEVIEW 13th was held?

Based **on** the **provided context,** the **answer** to the question "When **the DEVIEW** 13th was held?" **is:** In November 2020.

PROOF:

A white text on a white background

Description automatically generated

### Question 8: What is the financial position of property and equipments of Naver?

Based on the given **information,** I **cannot** determine the specific **financial position** of **Naver's properties** and equipment **as** that information is **not** provided in **the** given **context.** However, I can provide some general information about Naver's **financial position** based on the given data. **According** to the consolidated statements of **financial** position, Naver had **total assets** of 17,014,238,663,770 Korean won as of December 31, 2020, **and 12,299,527,120,786 Korean won** as of December **31,** 2019. **This** indicates **that Naver's** total **assets have increased by** approximately **40%** from 2019 to 2020. Of **these total** assets, **current** assets such **as cash and cash equivalents, short-term financial** instruments, and trade **and** other **receivables accounted** for approximately **75% of** the **total** assets **as** of December 31, 2020. This suggests that Naver has a **significant** portion **of** its **assets** in liquid form, which could be **used** to finance **its** operations or investments. **On** the **liabilities** side, **Naver** had current liabilities **such** as trade and other payables **and financial** liabilities **at fair** value **-** profit or loss, which together accounted for approximately 60% of **its total** liabilities **as of** December **31,** 2020. This indicates **that Naver** has a significant **portion of** its liabilities in short-term form, which **could be** repaid using **its liquid** assets. **Overall, based** on the **given** information, Naver appears to have a strong financial position **with a** significant **amount of liquid** assets and a **relatively low** level of long-term liabilities. However, **without more** detailed **information** about Naver's properties and equipment, I cannot provide a **specific** answer to **your** question about **their financial** position.

PROOF:

A screenshot of a report

Description automatically generated

## Conclusion

To sum up, this application performs LLM taks based on given document. The technology behind this desktop app are Ollama, privateGPT and PyQt5 libraries. Ollama is great software to run the LLM model in ordinary PC. privateGPT is the core of our application. privateGPT implemented the RAG (Retrieval Augmented Generation) as a High-level API. PyQt5 is widely used binding of Qt for Python language.

“Generative AI is a game changer for our society, but adaoptation in companies of all sizes and data-sensitive domains like healthcare or legal is limited by a clear concern: privacy. Not being able to ensure that your data is fully under your control when using thrid-party AI tools is a risk these industries cannot take.” -privateGPT

GitHub Repo: <https://github.com/AbdirayimovS/PDFchatbot/tree/main>