

ハンズオン準備

①事前配布ファイルを取得する

Tech0day vol6 ご参加予定の皆様

平素は大変お世話になっております。

株式会社Tech0 イベント運営事務局でございます。

この度はTech0day Vol6へのご参加申し込みいただき、

誠にありがとうございます。

直前のご連絡となり恐縮ではございますが、

明日の集合時間、集合場所、参加ZoomURLにつきましてご連絡いたしました。

●集合時間および場所

- ・集合時間：19:20-19:30（イベントは19:30-開始です）
- ・集合場所：[〒141-0031東京都品川区西五反田1丁目17-3 plaza square五反田 4階](#)
- ・リモート参加の場合はこちらから：<https://zoom.us/j/95089999842?pwd=9rG2chwlbM2hNhYuo07TXL0MxRVETZ.1>

●ハンズオン資料

[こちら](#)より、ファイル二つとフォルダー一つをダウンロードください。

● 2ファイル

L20240802_Tech0day_vol6.ipynb

L【説明資料】ハンズオン環境立ち上げ.pdf

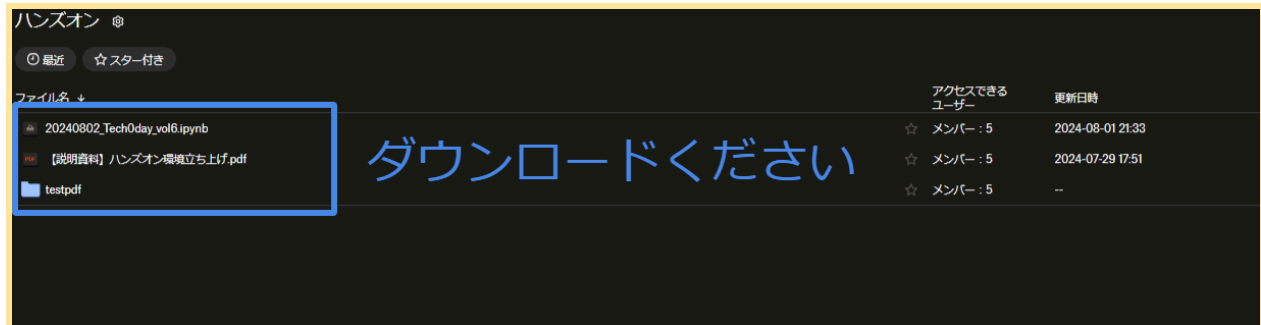
※当日は↑を開きながら、ハンズオン準備を進めるとスムーズです。

● 1フォルダ

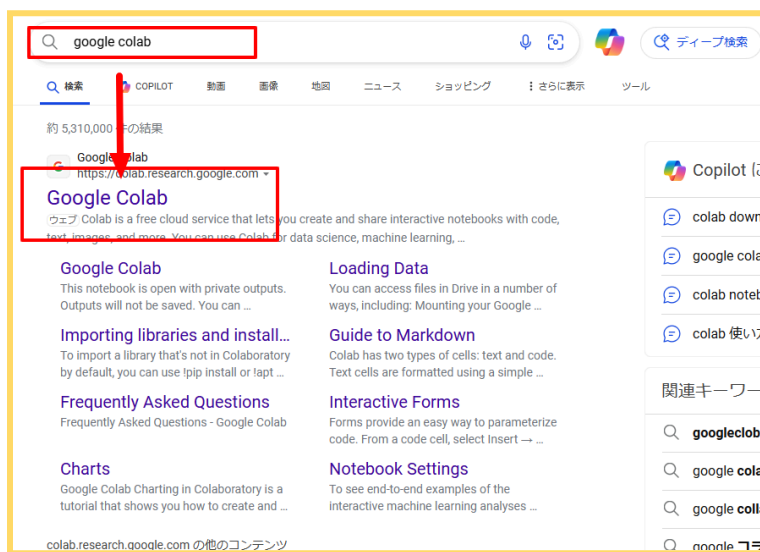
testpdf(①～④のPDFファイル)

※念のため添付ファイルもお送りいたします

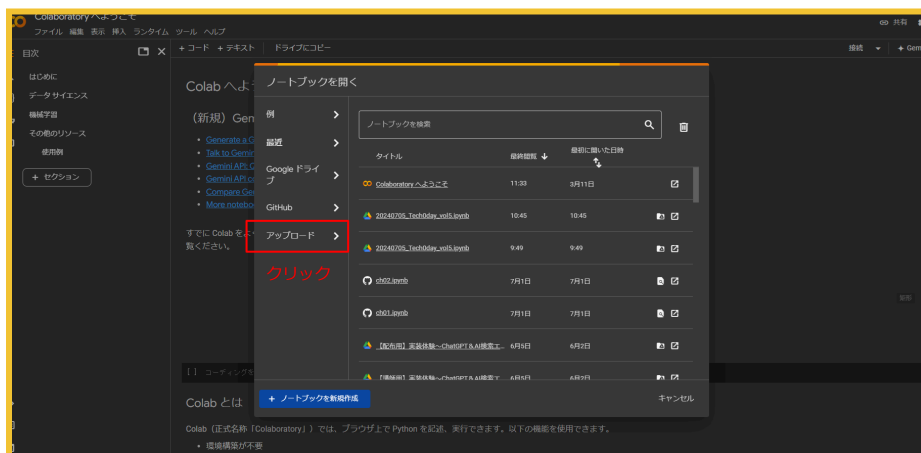
②Dropboxよりダウンロード

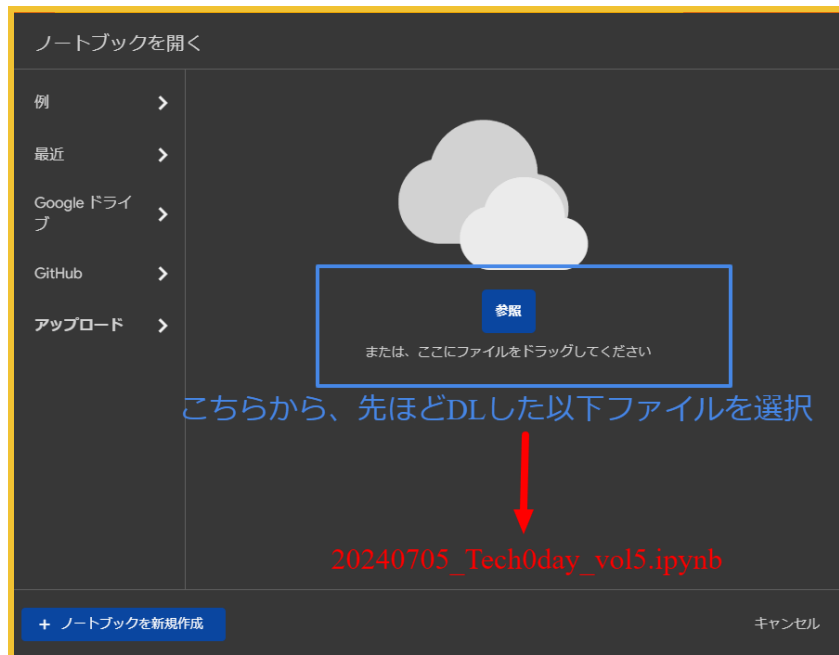


③Google Colabと検索



③ノートブックを開く→アップロードをクリック ※こちらが実行不可の場合、一番下へ。





この画面になれば準備完了です！！

```
20240705_Tech0day_vol5.ipynb ☆
ファイル 編集 表示 挿入 ランタイム ツール ヘルプ 最終保存: 11:54
+ コード + テキスト 接続 Gemini

[ ] pip install pdfminer
Requirement already satisfied: pdfminer in /usr/local/lib/python3.10/dist-packages (20181125)
Requirement already satisfied: pycryptodome in /usr/local/lib/python3.10/dist-packages (from pdfminer) (3.20.0)

# 接続テスト
from pdfminer.pdfinterp import PDFResourceManager
rsmgr = PDFResourceManager()

[ ] # 必要なpdfminer.sixモジュールのクラスをインポート
from pdfminer.pdfinterp import PDFResourceManager
from pdfminer.converter import TextConverter
from pdfminer.pdfinterp import PDFPageInterpreter
from pdfminer.pdfpage import PDFPage
from pdfminer.layout import LAParams
from io import StringIO

[ ] # 標準読み込み関数open()でモード指定をbinaryでFileオブジェクトを取得
fp = open("/content/01電気エネルギーを利用し大気CO2を固定する.pdf", 'rb')

[ ] # 出力先をPythonコンソールするためにIOストリームを取得
outfp = StringIO()

# 各種テキスト抽出に必要なPdfminer.sixのオブジェクトを取得する処理
rsmgr = PDFResourceManager() # PDFResourceManagerオブジェクトの取得
lparams = LAParams() # LAParamsオブジェクトの取得
device = TextConverter(rsmgr, outfp, lparams=lparams) # TextConverterオブジェクトの取得
iprtr = PDFPageInterpreter(rsmgr, device) # PDFPageInterpreterオブジェクトの取得

# PDFファイルから1ページずつ解析所(テキスト抽出)処理する
for page in PDFPage.get_pages(fp):
    iprtr.process_page(page)

text = outfp.getvalue() # Pythonコンソールへの出力内容を取得
```

③が実行不可だった場合→[yokishin1005/tech0handson_rag \(github.com\)](https://github.com/yokishin1005/tech0handson_rag)をクリックし、遷移先のURLをコピー([yokishin1005/tech0handson_rag](https://github.com/yokishin1005/tech0handson_rag))し、以下を実行

