Appendix３；

3.3.2 ：S-BERT（Sentence-BERT）を使って、MITRE ATT&CKの攻撃テクニックの説明文と総務省・デジタル庁のガイドラインPDFから抽出した文との意味的類似度をスコア化するGoogle Colab対応のPythonコードです。

**機能概要**

* s-BERTモデルで埋め込み（sentence embedding）を生成
* cosine類似度でスコア化（0〜1）
* PDFからテキストを抽出（PyMuPDF使用）
* MITRE説明文とガイドライン文書の組で最も高いスコアを記録
* 類似度が閾値を超えたら「実装済」と判定
* 出力はCSVでダウンロード可能

**Google Colab用Pythonコード（s-BERT + PDF対応 + 類似度スコア + CSV出力）**

#  1. 必要なライブラリをインストール  
!pip install -q sentence-transformers PyMuPDF pandas scikit-learn  
  
#  2. ライブラリのインポート  
import pandas as pd  
import fitz # PyMuPDF  
from sentence\_transformers import SentenceTransformer, util  
from sklearn.metrics.pairwise import cosine\_similarity  
import numpy as np  
  
#  3. s-BERTモデルのロード（軽量モデル）  
model = SentenceTransformer('all-MiniLM-L6-v2')  
  
#  4. PDFからテキストを抽出する関数  
def extract\_text\_from\_pdf(pdf\_path):  
doc = fitz.open(pdf\_path)  
texts = []  
for page in doc:  
text = page.get\_text().strip()  
if text:  
texts.append(text)  
doc.close()  
return texts  
  
#  5. 類似度スコア計算関数（最大スコアと該当文）  
def compute\_max\_similarity(mitre\_text, guideline\_sentences):  
mitre\_emb = model.encode([mitre\_text], convert\_to\_tensor=True)  
guideline\_embs = model.encode(guideline\_sentences, convert\_to\_tensor=True)  
scores = util.cos\_sim(mitre\_emb, guideline\_embs)[0]  
max\_idx = scores.argmax()  
return float(scores[max\_idx]), guideline\_sentences[max\_idx]  
  
#  6. MITRE説明CSV（列名：Technique\_ID, Description, Asset\_Type）をアップロード  
from google.colab import files  
uploaded = files.upload()  
mitre\_df = pd.read\_csv(next(iter(uploaded)))  
  
#  7. ガイドラインPDFをアップロード  
uploaded\_pdf = files.upload()  
pdf\_path = next(iter(uploaded\_pdf))  
guideline\_sentences = extract\_text\_from\_pdf(pdf\_path)  
  
#  8. スコアを計算しDataFrameに追加（閾値は0.7）  
similarities = []  
matched\_texts = []  
implementation = []  
  
for i, row in mitre\_df.iterrows():  
desc = row["Description"]  
score, matched = compute\_max\_similarity(desc, guideline\_sentences)  
similarities.append(round(score, 3))  
matched\_texts.append(matched)  
implementation.append("実装済" if score >= 0.7 else "要検討")  
  
mitre\_df["Similarity"] = similarities  
mitre\_df["Matched\_Text"] = matched\_texts  
mitre\_df["Implementation\_Status"] = implementation  
  
#  9. CSV出力  
output\_path = "/content/mitre\_similarity\_result.csv"  
mitre\_df.to\_csv(output\_path, index=False)  
  
#  10. ダウンロードリンクの表示  
from google.colab import files  
files.download(output\_path)