ASSIGNMENT – 2

Research Paper Comparator with KPI & Visual Insights

Code:

```
import streamlit as st
import PyPDF2
import openai
from io import BytesIO
from docx import Document
from wordcloud import WordCloud
import matplotlib.pyplot as plt
from collections import Counter
import re
# --- Streamlit Config ---
st.set_page_config(page_title="📑 Enhanced Research Paper Comparator", layout="centered")
st.title("📑 Research Paper Comparator with KPI & Visual Insights")
# --- OpenAI Key ---
api_key = st.text_input("Enter your OpenAI API Key", type="password")
# --- Upload PDFs ---
st.subheader("📤 Upload Two Research Papers (PDF)")
pdf1 = st.file_uploader("Upload First Research Paper", type=["pdf"], key="pdf1")
pdf2 = st.file_uploader("Upload Second Research Paper", type=["pdf"])
# --- Extract Text ---
def extract_text(uploaded_pdf):
    if not uploaded_pdf:
        return ""
    reader = PyPDF2.PdfReader(uploaded_pdf)
    text = ""
    for page in reader.pages:
        text += page.extract_text() or ""
```

```
    return text[:4000]
# --- GenAI Prompt for KPI ---
def get_kpi_response(text1, text2):
    prompt = f"""
Compare these two research papers and answer ONLY in the following KPI format:
1. Domain Similarity (Yes/No + Explanation)
2. Research Aspect Overlap (Yes/No + What aspects are common)
3. Innovation Uniqueness Score (1-10 with justification)
4. Content Similarity Index (0-100% estimation)
5. Theme Summary (1-2 lines for each paper)
6. Gap Analysis (What one covers that other doesn't)
7. Best Use Case for Each Paper (brief description)
--- Paper 1 ---
{text1}
--- Paper 2 ---
{text2}
"""
     openai.api_key = api_key
    response = openai.chat.completions.create(
        model="gpt-3.5-turbo",
        messages=[{"role": "user", "content": prompt}],
        max_tokens=1000
    )
    return response.choices[0].message.content.strip()
# --- GenAI Prompt for Highlights ---
def get_highlights(text, paper_number):
    prompt = f"Give 3 important bullet-point highlights of the following research paper:\n\n{text}\n\n"
    response = openai.chat.completions.create(
        model="gpt-3.5-turbo",
        messages=[{"role": "user", "content": prompt}],
        max_tokens=200
```

```
    )
    return response.choices[0].message.content.strip()
# --- Keyword Extraction & WordCloud ---
def extract_keywords(text, num_keywords=30):
    words = re.findall(r'\b[a-zA-Z]{5,}\b', text.lower())
    common = Counter(words).most_common(num_keywords)
    return dict(common)


def generate_wordcloud(keywords_dict, title):
    wordcloud = WordCloud(width=600, height=400,
background_color='white').generate_from_frequencies(keywords_dict)
    fig, ax = plt.subplots()
    ax.imshow(wordcloud, interpolation='bilinear')
    ax.axis('off')
    ax.set_title(title, fontsize=16)
    st.pyplot(fig)


# --- Compare Button ---
if st.button(" 🔍 Compare and Generate Insights"):
    if not all([pdf1, pdf2, api_key]):
        st.error("Please upload both PDFs and enter your OpenAI API key.")
    else:
        with st.spinner("Analyzing and generating insights..."):
            try:
                t1 = extract_text(pdf1)
                t2 = extract_text(pdf2)


                # KPI Response
                kpi_result = get_kpi_response(t1, t2)


                # Highlights
```

```
highlights1 = get_highlights(t1, 1)
highlights2 = get_highlights(t2, 2)

# Keyword clouds
keywords1 = extract_keywords(t1)
keywords2 = extract_keywords(t2)

st.success("✅ Insights Generated!")

# --- Display KPI Summary ---
st.subheader("📊 KPI Summary")
for line in kpi_result.split("\n"):
    if ":" in line:
        key, value = line.split(":", 1)
        st.markdown(f"*{key.strip()}*: {value.strip()}")
    else:
        st.markdown(line.strip())

# --- Visual Word Clouds ---
st.subheader("☁️ Keyword Cloud Comparison")
col1, col2 = st.columns(2)
with col1:
    generate_wordcloud(keywords1, "Paper 1 Keywords")
with col2:
    generate_wordcloud(keywords2, "Paper 2 Keywords")

# --- Key Highlights ---
st.subheader("🔎 Key Highlights from Each Paper")
st.markdown("*Paper 1 Highlights:*")
st.markdown(highlights1)
st.markdown("*Paper 2 Highlights:*")
```

```
            st.markdown(highlights2)

            # --- Export to DOCX ---
            doc = Document()
            doc.add_heading("Research Paper Comparison - KPI Report", 0)
            doc.add_paragraph(kpi_result)
            doc.add_heading("Paper 1 Highlights", level=1)
            doc.add_paragraph(highlights1)
            doc.add_heading("Paper 2 Highlights", level=1)
            doc.add_paragraph(highlights2)

            buf = BytesIO()
            doc.save(buf)
            buf.seek(0)
            st.download_button("📥 Download KPI Report (DOCX)", data=buf,
                               file_name="research_comparison_kpi.docx",
                               mime="application/vnd.openxmlformats-
officedocument.wordprocessingml.document")

        except Exception as e:
            st.error(f"Error: {e}")
```

New Features Added:

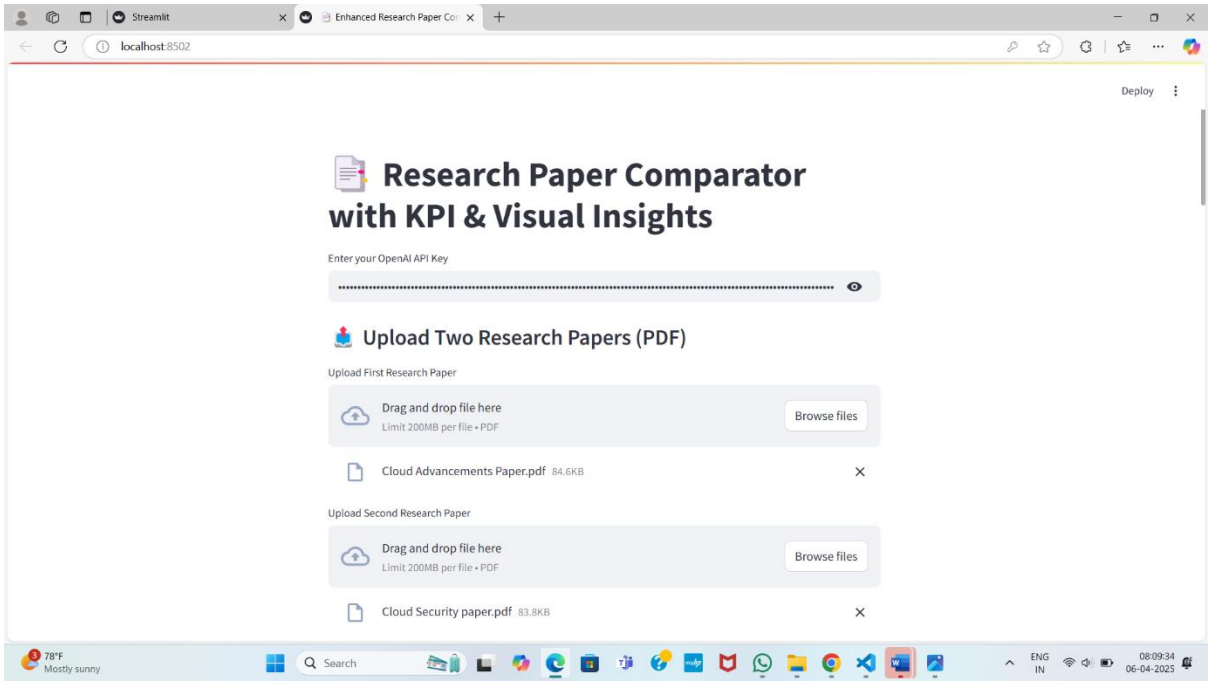
Keyword Cloud & Key Highlights Comparison

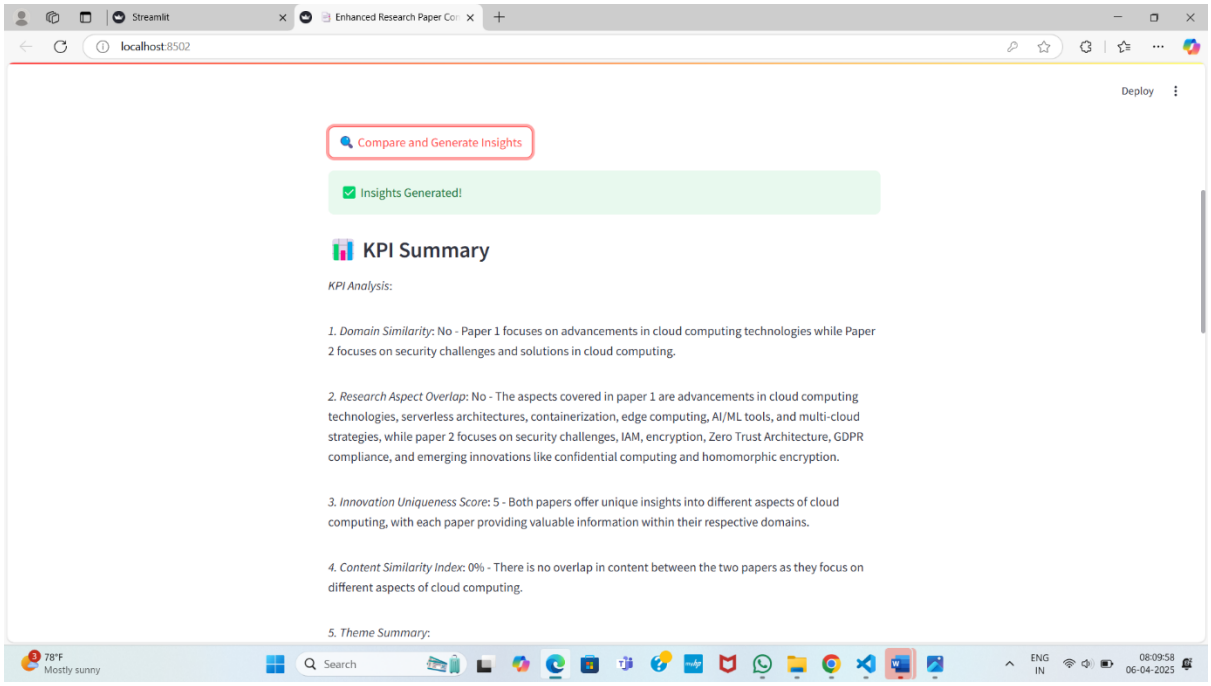
- Keyword Cloud: Extracts top keywords from both papers and displays them in a word cloud.

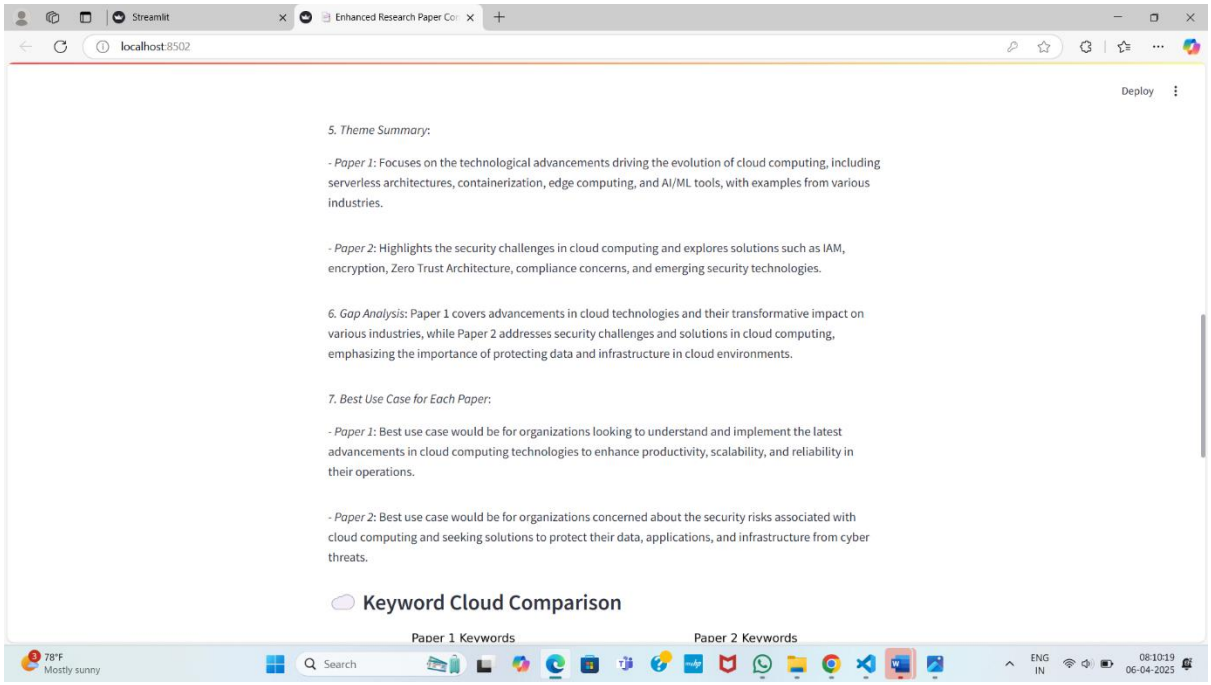
- Key Highlights: Extracts 3 key bullet-point highlights from each paper using OpenAI.

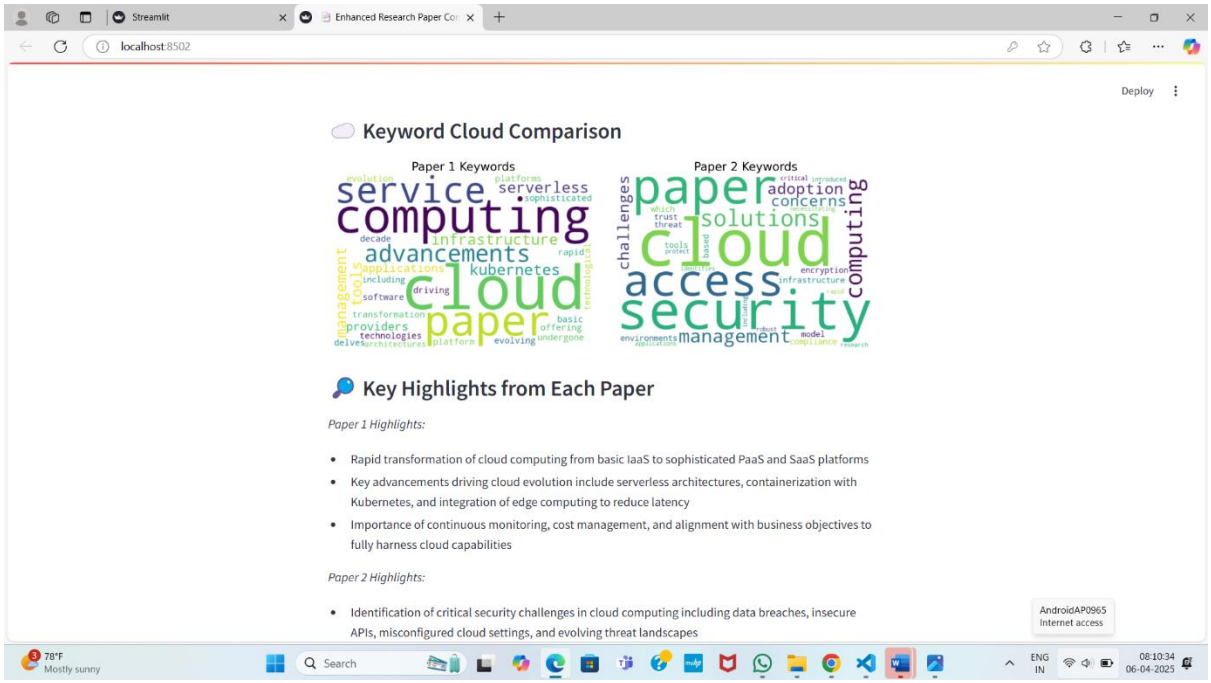
This makes the comparison visually appealing and adds depth to understanding beyond textual KPIs.

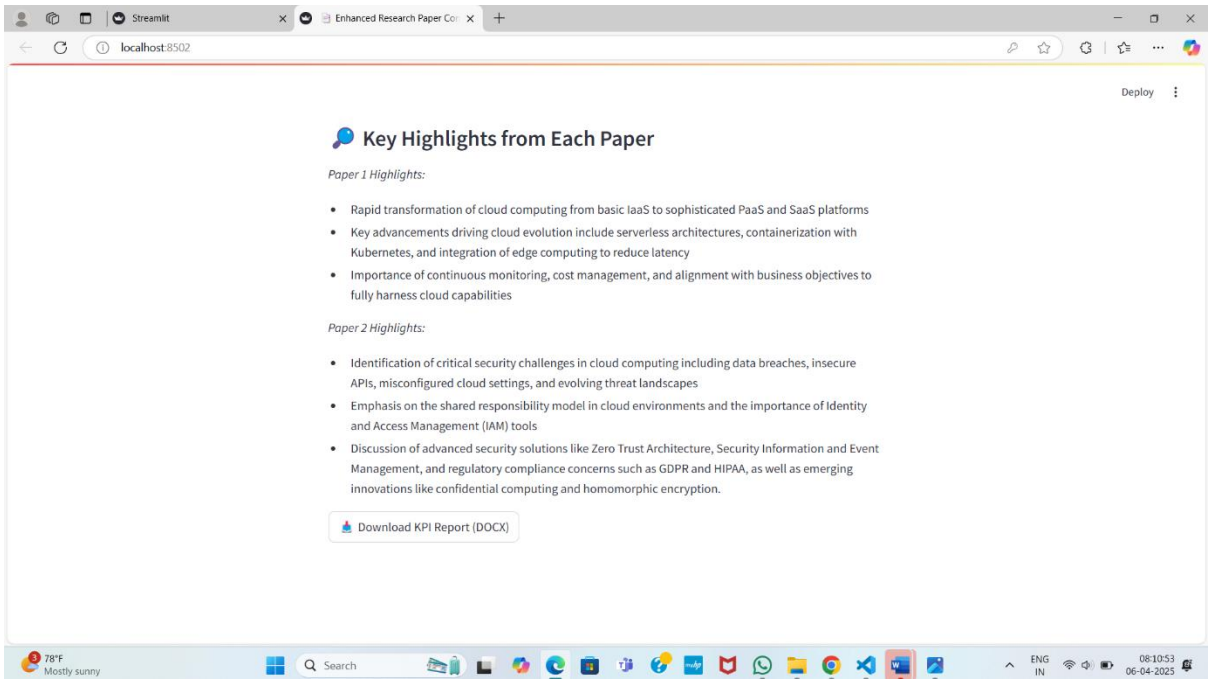
Output:

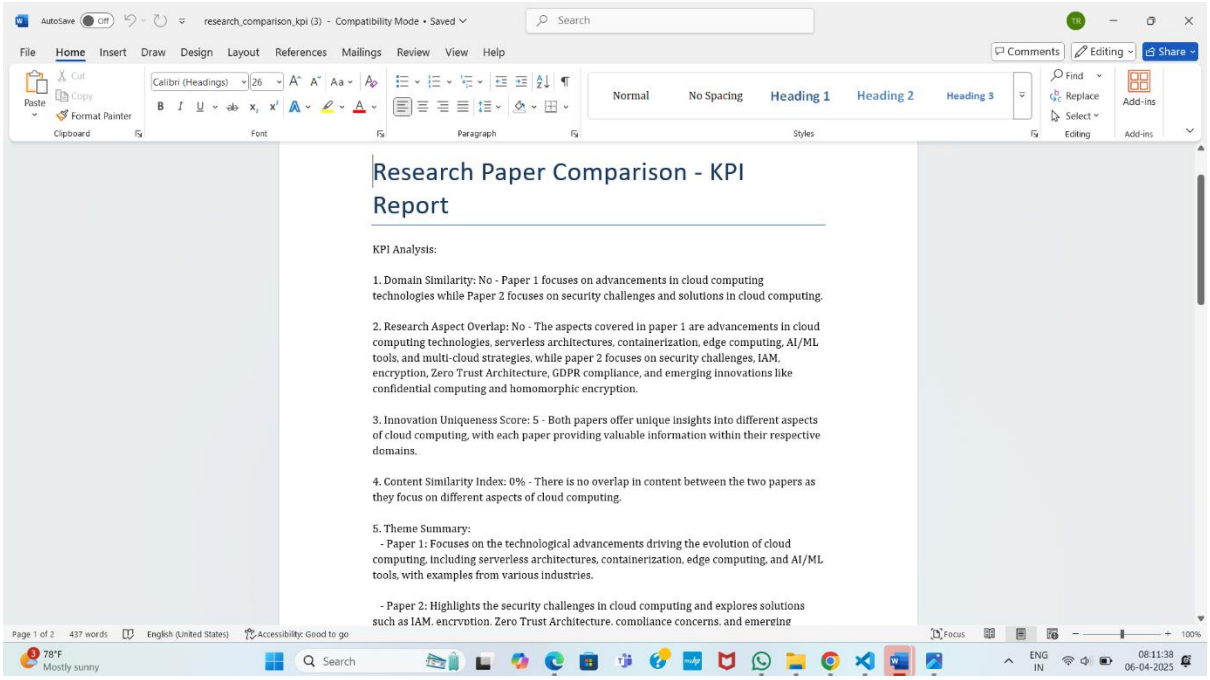



Streamlit
Enhanced Research Paper Co
localhost:8502
Deploy
Compare and Generate Insights
Insights Generated!
KPI Summary
KPI Analysis:
1. Domain Similarity: No - Paper 1 focuses on advancements in cloud computing technologies while Paper 2 focuses on security challenges and solutions in cloud computing.
2. Research Aspect Overlap: No - The aspects covered in paper 1 are advancements in cloud computing technologies, serverless architectures, containerization, edge computing, AI/ML tools, and multi-cloud strategies, while paper 2 focuses on security challenges, IAM, encryption, Zero Trust Architecture, GDPR compliance, and emerging innovations like confidential computing and homomorphic encryption.
3. Innovation Uniqueness Score: 5 - Both papers offer unique insights into different aspects of cloud computing, with each paper providing valuable information within their respective domains.
4. Content Similarity Index: 0% - There is no overlap in content between the two papers as they focus on different aspects of cloud computing.
5. Theme Summary:
78°F
Mostly sunny
Search
ENG
IN
08:09:58
06-04-2025


Streamlit
Enhanced Research Paper Co
localhost:8502
Deploy
5. Theme Summary:
- Paper 1: Focuses on the technological advancements driving the evolution of cloud computing, including serverless architectures, containerization, edge computing, and AI/ML tools, with examples from various industries.
- Paper 2: Highlights the security challenges in cloud computing and explores solutions such as IAM, encryption, Zero Trust Architecture, compliance concerns, and emerging security technologies.
6. Gap Analysis: Paper 1 covers advancements in cloud technologies and their transformative impact on various industries, while Paper 2 addresses security challenges and solutions in cloud computing, emphasizing the importance of protecting data and infrastructure in cloud environments.
7. Best Use Case for Each Paper:
- Paper 1: Best use case would be for organizations looking to understand and implement the latest advancements in cloud computing technologies to enhance productivity, scalability, and reliability in their operations.
- Paper 2: Best use case would be for organizations concerned about the security risks associated with cloud computing and seeking solutions to protect their data, applications, and infrastructure from cyber threats.
Keyword Cloud Comparison
Paper 1 Keywords
Paper 2 Keywords
78°F
Mostly sunny
Search
ENG
IN
08:10:19
06-04-2025


Streamlit
Enhanced Research Paper Co
localhost:8502
Deploy
Keyword Cloud Comparison
Paper 1 Keywords
Paper 2 Keywords
Key Highlights from Each Paper
Paper 1 Highlights:
Rapid transformation of cloud computing from basic IaaS to sophisticated PaaS and SaaS platforms
Key advancements driving cloud evolution include serverless architectures, containerization with Kubernetes, and integration of edge computing to reduce latency
Importance of continuous monitoring, cost management, and alignment with business objectives to fully harness cloud capabilities
Paper 2 Highlights:
Identification of critical security challenges in cloud computing including data breaches, insecure APIs, misconfigured cloud settings, and evolving threat landscapes
AndroidAP0965
Internet access
78°F
Mostly sunny
Search
ENG
IN
08:10:34
06-04-2025


Streamlit
Enhanced Research Paper Co
localhost:8502
Deploy
Key Highlights from Each Paper
Paper 1 Highlights:
Rapid transformation of cloud computing from basic IaaS to sophisticated PaaS and SaaS platforms
Key advancements driving cloud evolution include serverless architectures, containerization with Kubernetes, and integration of edge computing to reduce latency
Importance of continuous monitoring, cost management, and alignment with business objectives to fully harness cloud capabilities
Paper 2 Highlights:
Identification of critical security challenges in cloud computing including data breaches, insecure APIs, misconfigured cloud settings, and evolving threat landscapes
Emphasis on the shared responsibility model in cloud environments and the importance of Identity and Access Management (IAM) tools
Discussion of advanced security solutions like Zero Trust Architecture, Security Information and Event Management, and regulatory compliance concerns such as GDPR and HIPAA, as well as emerging innovations like confidential computing and homomorphic encryption.
Download KPI Report (DOCX)
78°F
Mostly sunny
Search
ENG
IN
08:10:53
06-04-2025



Research Paper Comparison - KPI Report

KPI Analysis:

1. Domain Similarity: No - Paper 1 focuses on advancements in cloud computing technologies while Paper 2 focuses on security challenges and solutions in cloud computing.

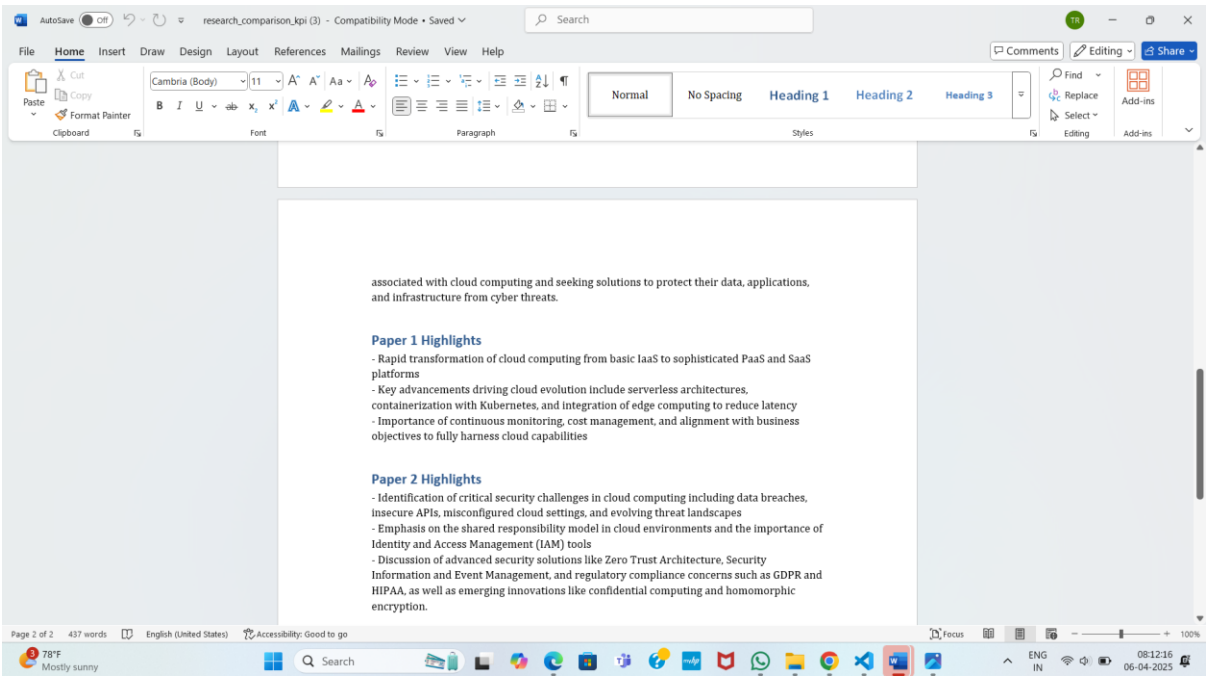
2. Research Aspect Overlap: No - The aspects covered in paper 1 are advancements in cloud computing technologies, serverless architectures, containerization, edge computing, AI/ML tools, and multi-cloud strategies, while paper 2 focuses on security challenges, IAM, encryption, Zero Trust Architecture, GDPR compliance, and emerging innovations like confidential computing and homomorphic encryption.

3. Innovation Uniqueness Score: 5 - Both papers offer unique insights into different aspects of cloud computing, with each paper providing valuable information within their respective domains.

4. Content Similarity Index: 0% - There is no overlap in content between the two papers as they focus on different aspects of cloud computing.

5. Theme Summary:
- Paper 1: Focuses on the technological advancements driving the evolution of cloud computing, including serverless architectures, containerization, edge computing, and AI/ML tools, with examples from various industries.

- Paper 2: Highlights the security challenges in cloud computing and explores solutions such as IAM, encryption, Zero Trust Architecture, compliance concerns, and emerging



associated with cloud computing and seeking solutions to protect their data, applications, and infrastructure from cyber threats.

Paper 1 Highlights

- Rapid transformation of cloud computing from basic IaaS to sophisticated PaaS and SaaS platforms
- Key advancements driving cloud evolution include serverless architectures, containerization with Kubernetes, and integration of edge computing to reduce latency
- Importance of continuous monitoring, cost management, and alignment with business objectives to fully harness cloud capabilities

Paper 2 Highlights

- Identification of critical security challenges in cloud computing including data breaches, insecure APIs, misconfigured cloud settings, and evolving threat landscapes
- Emphasis on the shared responsibility model in cloud environments and the importance of Identity and Access Management (IAM) tools
- Discussion of advanced security solutions like Zero Trust Architecture, Security Information and Event Management, and regulatory compliance concerns such as GDPR and HIPAA, as well as emerging innovations like confidential computing and homomorphic encryption.