SIBYL

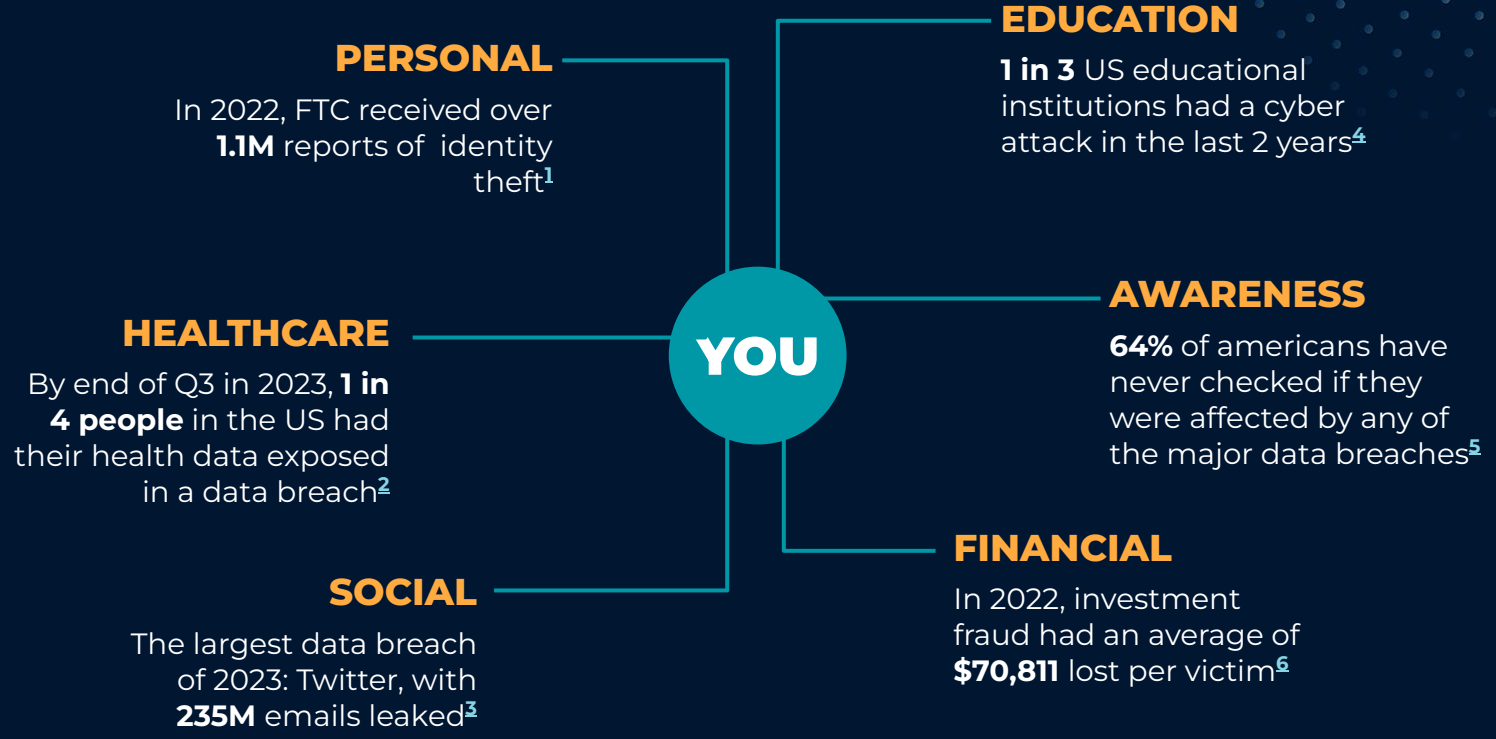
CYBERSECURITY COPILOT

Goldberg | Rahmatpanah | Williams

$9.5 TRILLION

$1B ESTIMATED HOURLY COST[1]

YOU & CYBERSECURITY



PERSONAL

In 2022, FTC received over **1.1M** reports of identity theft[1]

EDUCATION

1 in 3 US educational institutions had a cyber attack in the last 2 years[4]

HEALTHCARE

By end of Q3 in 2023, **1 in 4 people** in the US had their health data exposed in a data breach[2]

AWARENESS

64% of americans have never checked if they were affected by any of the major data breaches[5]

SOCIAL

The largest data breach of 2023: Twitter, with **235M** emails leaked[3]

FINANCIAL

In 2022, investment fraud had an average of **$70,811** lost per victim[6]

PROBLEM

Vulnerability exposes companies to exploitation and breaches

$4.4M average loss for a single non-compliance event[3]

Lengthy, dense security documents causes cyber fatigue and hinder updates

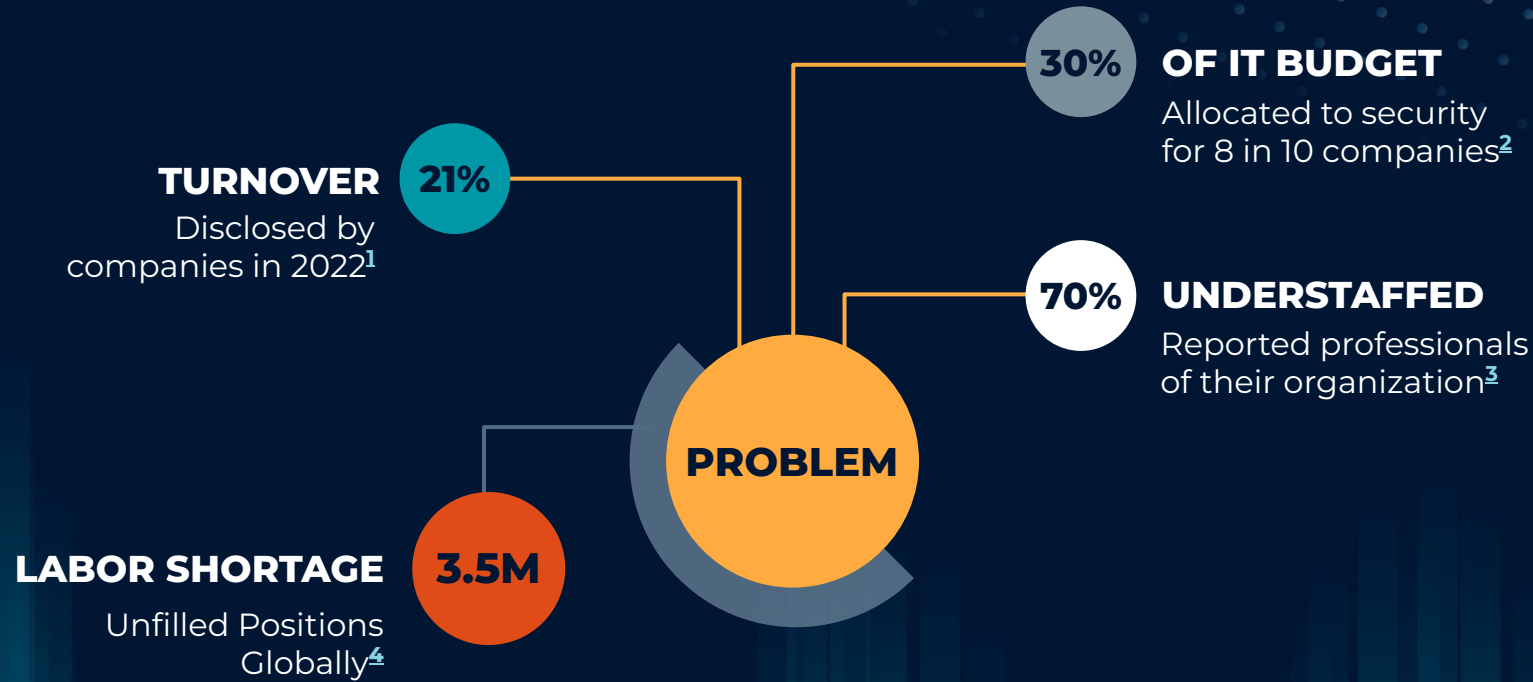
Cyber fatigue affects as much as **42 %** of companies[1]

National Institute of Standards and Technology (NIST)
International Organization for Standardization (ISO)

Cyber fatigue leaves systems vulnerable to potential attacks

A cyber incident occurs every **39** seconds[2]

INDUSTRY PROBLEM



SOLUTION

Our Cybersecurity Copilot:

SIBYL

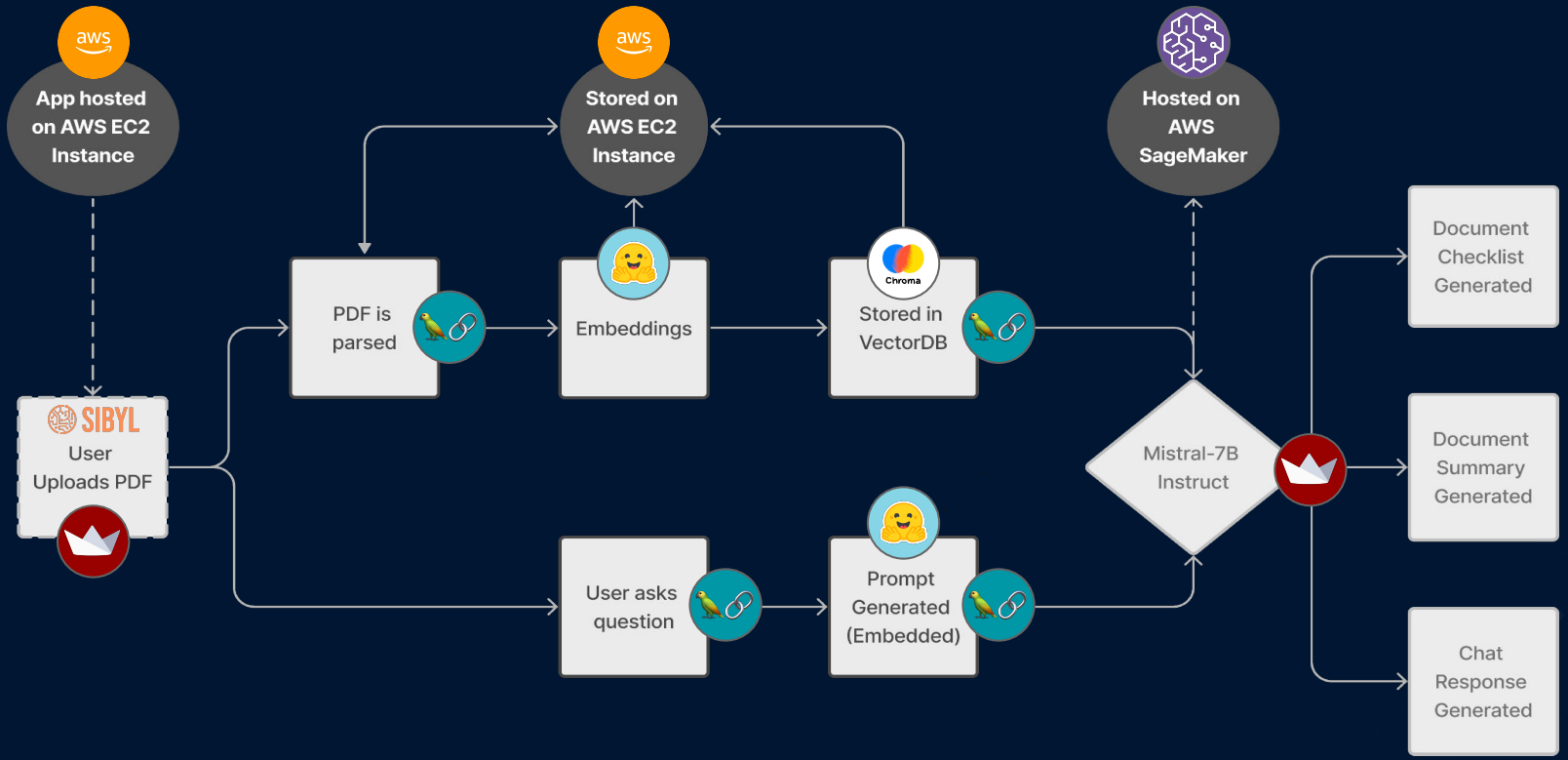
1. Alleviates Compliance Fatigue

2. Reduces Human Errors

3. Boosts Implementation Efficacy

View Demo

ARCHITECTURE



DATA & MODEL SELECTION

DATA

No public training data
Scraped 19,000 NIST PDFs
2.45GB significant tokens text file
Limited resources pivot:

- RAG implementation
- Vector database
- PDF upload
- Chunking strategies

MODELS

Llama 2
Falcon 7B Instruct
Falcon 180B Instruct
Pegasus-X
Mistral 7B Instruct
Zephyr 7B Alpha
Zephyr 7B Beta
OpenAI (ChatGPT)

Q&A EVALUATION

Q&A SAMPLE

Q: The company that Scott works for has experienced a data breach, and the personal information of thousands of customers has been exposed. Which of the following impact categories is not a concern as described in this scenario?

A. Financial
B. Reputation
C. Availability loss
D. Data loss

SECURITY+

90 minutes to complete 90 questions.

SUMMARY EVALUATION

RUBRIC

Grade summaries systematically based on five criteria

	HIGH	MED	LOW
RELEVANCE	✓	✓	✓
CLARITY	✓	✓	
COHERENCE	✓	✓	✓
LENGTH	✓		
ACCURACY	✓	✓	✓

SCORING

Scores based on reader discretion of summary generated for NIST.IR.8270

ADVANTAGE

Tailored for our specific use case and expected summaries

NIST SUMMARY

Space is a newly emerging commercial critical infrastructure sector that is no longer the domain of only national government authorities. Space is an inherently risky environment in which to operate, so cybersecurity risks involving commercial space – including those affecting commercial satellite vehicles – need to be understood and managed alongside other types of risks to ensure safe and successful operations.

This report provides a general introduction to cybersecurity risk management for the commercial satellite industry as they seek to start managing cybersecurity risks in space.

This document is by no means comprehensive in terms of addressing all of the cybersecurity risks to commercial satellite infrastructure, nor does it explore risks to satellite vehicles, which may be introduced through the implementation of cybersecurity controls. The intent is to present basic concepts, generate discussions, and provide sample references for additional information on pertinent cybersecurity risk management models.

Summary written by authors of the adopted NIST Standard.

SIBYL SUMMARY

S: ...to address core cybersecurity areas such as identifying assets, protecting assets, detecting cybersecurity incidents, responding to those incidents, and recovering from them.

...mitigate potential threats and vulnerabilities, such as intentional jamming and spoofing, interception and theft, denial-of-service attacks, malicious code injection, and hijacking and unauthorized commands.

...cybersecurity that are relevant to commercial satellite operations, including security requirements for cryptographic modules, security and privacy controls for federal information systems and organizations, leadership in space commerce, critical infrastructure cybersecurity, risk assessments, national information assurance policy for space systems used to support national security missions, licenses and permits for commercial space transportation, and licensing of private remote sens space systems.

RUBRIC

Grade summaries systematically based on five criteria

MODEL EVALUATION

MODEL	Q&A	SUMMARY	VERSION
Mistral	45/50	High	7B-Instruct
Falcon	44/50	High	180B-Instruct
Llama	45/50	High	Llama 2-70b
ChatGPT BASELINE	50/50	High	3.5 + Ask YourPDF

OTHERS VS. SIBYL

TASK	OTHER CHATBOTS	SIBYL
Question & Answering	Relies on outdated and broad-topic data that can lead to inaccurate answers.	Leverages uploaded document and retrieval-augmentation generation (RAG) for precise answers.
Summary & Checklist	Limited word count results in incomplete use of the full document, missing information.	Harnesses map-reduce to handle the entire document to craft a complete summary and checklist.

What does the document say about how long do commercial satellites stay in orbit?

The document does not mention how long commercial satellites stay in orbit.
Pulled from Pages: [8, 10, 14, 34, 37, 39, 40, 41]

PIVOTAL DISCOVERIES

Refine dynamic retriever methods and prompts to boost response accuracy

A retriever obtains relevant info from a vector database

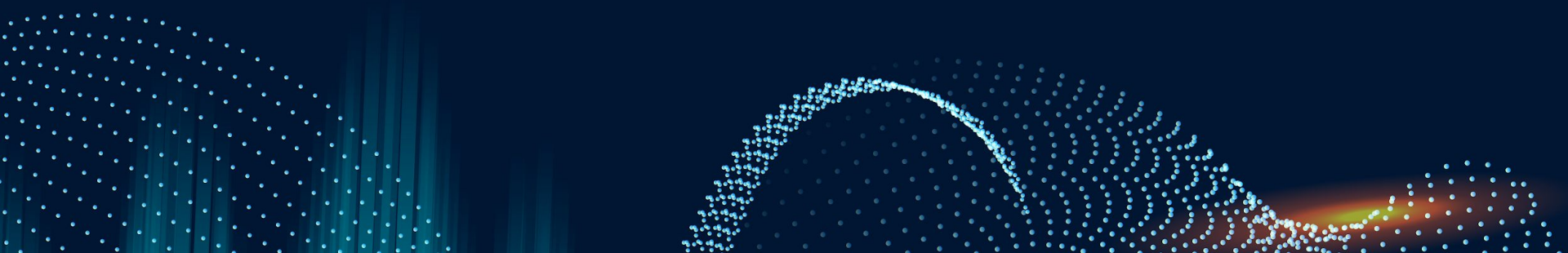
Minor variations in ETL of data greatly impacted response accuracy

ETL stands for extract, transform, and load

Scalable resources for evolving architecture of model and available data

FUTURE WORK

1. Simultaneous Document Engagement
2. Source Page, Figures & Tables Images
3. Further LLM Refinement & Training
4. Standards Gap Mapping Feature



SIBYL
CYBERSECURITY COPILOT

“The amount of time it takes for an organization to understand complex cybersecurity regulations is underappreciated. I believe a tool like this brings something new to the table - and can make cybersecurity compliance more accessible to every perspective in an organization from legal, to quality assurance, to R&D”

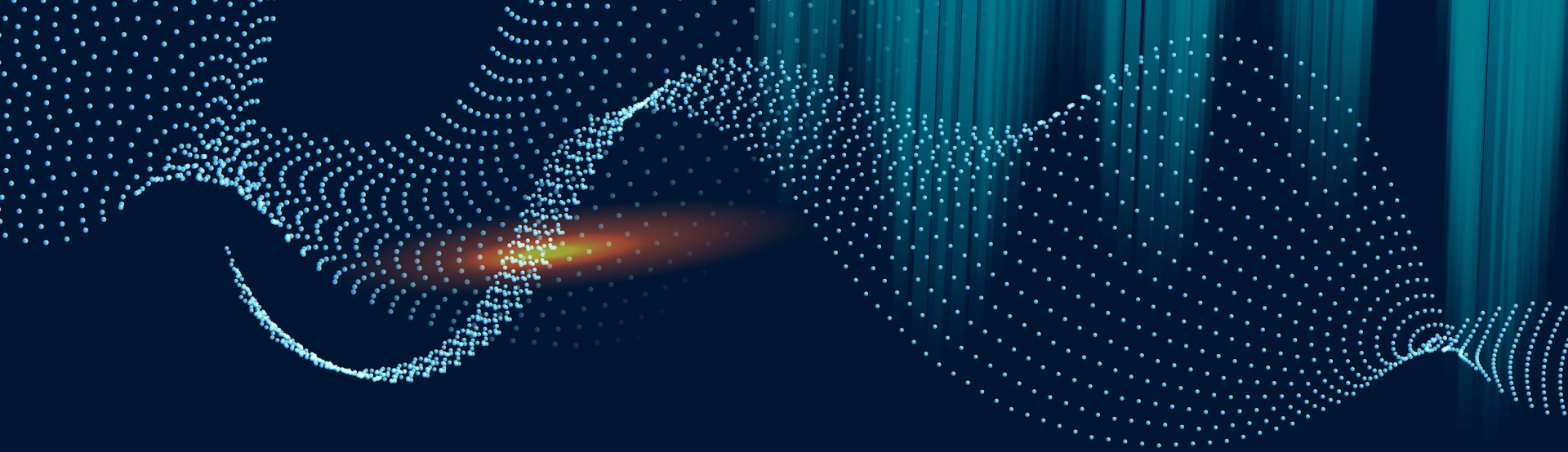
Jason Young, Cybersecurity Expert (Medical Devices)

MISSION

Make compliance both efficient and effective, so cybersecurity professionals can focus on what matters most - protecting their organizations from cybersecurity risks and threats.

THANK YOU

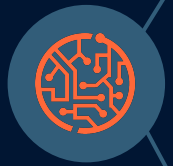
Over the past 14 weeks, a substantial amount of effort has been dedicated to the development of our AI Copilot, Sibyl. We wish to express our sincere gratitude to our families, friends, colleagues, classmates, cybersecurity professionals and instructors for their invaluable support during this project.



APPENDIX

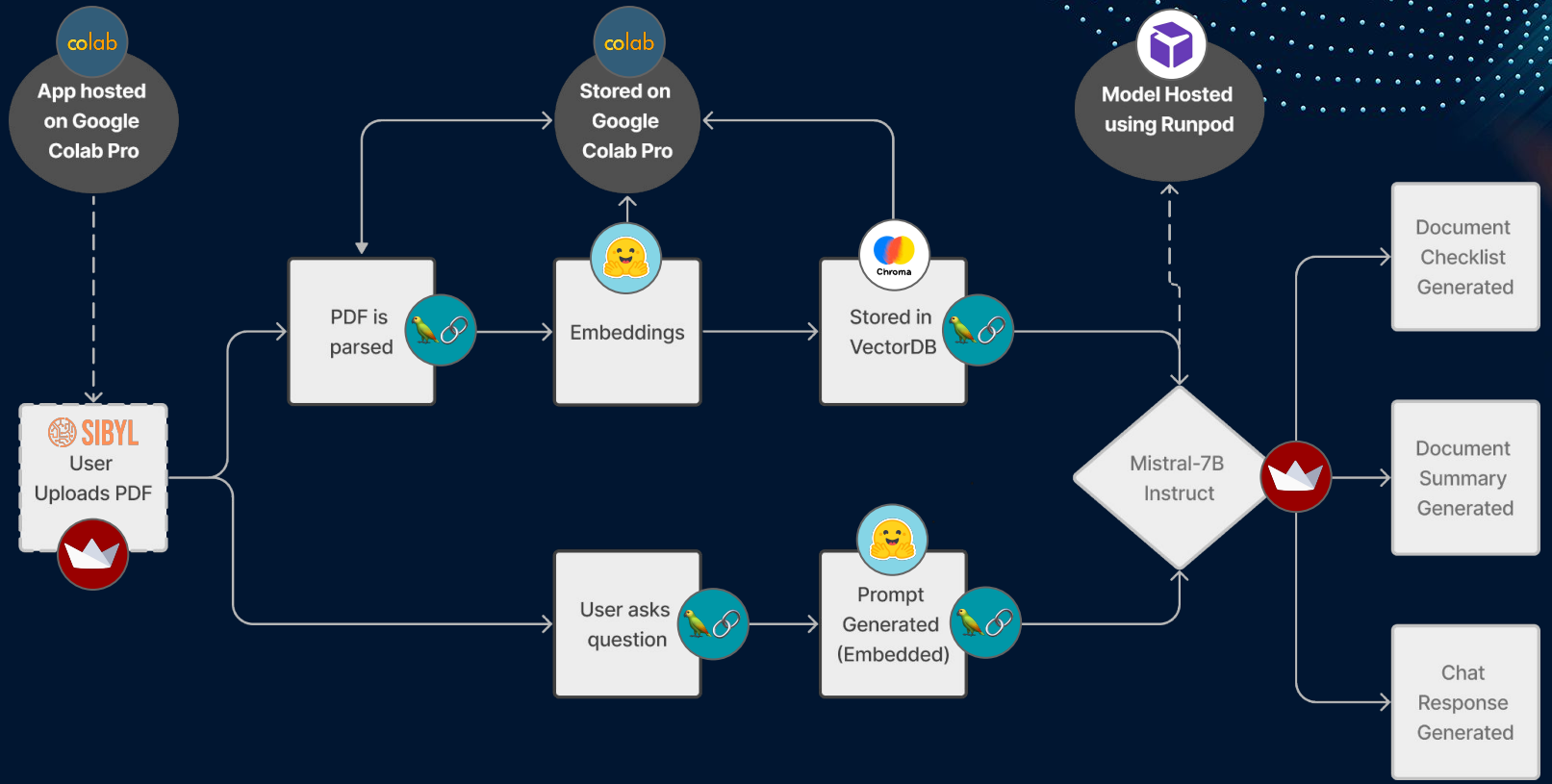
Additional Information

SIBYL SUMMARY



The provided excerpts discuss the importance of cybersecurity for commercial satellite operations and the need to implement robust measures to protect sensitive data and systems. The Cybersecurity Framework (CSF) is a risk management approach that can be tailored to various industries and consists of five primary functions: Identify, Protect, Detect, Respond, and Recover. The CSF can be applied to a notional low-Earth orbit (LEO) "small satellite vehicle" to address core cybersecurity areas such as identifying assets, protecting assets, detecting cybersecurity incidents, responding to those incidents, and recovering from them. The risk assessment process involves prioritizing and validating cybersecurity outcomes, considering costs and potential risks, consulting authorities, applying risk assessment principles, and determining outcomes that will achieve the desired risk posture in a cost-effective way. The excerpts also provide subcategories and response strategies for each function to mitigate potential threats and vulnerabilities, such as intentional jamming and spoofing, interception and theft, denial-of-service attacks, malicious code injection, and hijacking and unauthorized commands. The main theme is the importance of implementing effective cybersecurity measures to protect commercial satellite operations from potential threats. Additionally, the excerpts cover various aspects of cybersecurity that are relevant to commercial satellite operations, including security requirements for cryptographic modules, security and privacy controls for federal information systems and organizations, leadership in space commerce, critical infrastructure cybersecurity, risk assessments, national information assurance policy for space systems used to support national security missions, licenses and permits for commercial space transportation, and licensing of private remote sensing space systems. The excerpts also touch on the topic of hacking satellites and the potential risks associated with such activities.

AGILE ARCHITECTURE & BACKUP



MARKET RESEARCH

Launched in Fall 2023, interactive platform that links standards and guidelines into one view

DRATA

One of many enterprise MSP that handle all or part of an organization's cybersecurity functions

The global security market value is forecast to reach $424.97 billion in 2030[2]

TARGET PROFESSIONALS

- **1.1M** Cybersecurity Professionals[1]
- **83%** Male[2]
- **60%** Aged 40+[2]
- **56%** Hold a Bachelor's Degree[2]
- **600K** Unfilled Positions in US[1]
- **3.5M** Unfilled Positions Globally[3]
- **1-2** Years Average Tenure[2]
- **$102,600** Median Wage[3]

76% Manually Scan Websites for Regulatory Changes[4]

DEMO QUESTIONS

QUESTION	ANSWER	PAGES
Based on the document are satellites crewed or uncrewed?	The satellites discussed in the document are uncrewed commercial space vehicles that will not dock with human-occupied spacecraft.	5, 8, 10, 11, 14, 34, 37, 40
What does the document say about how long do commercial satellites stay in orbit?	The document does not mention how long commercial satellites stay in orbit.	8, 10, 14, 34, 37, 39, 40, 41
What does the document say about outsourcing?	The document says that ground operations can be outsourced in whole or in part. Even at launch, the payload operator may not be collocated with the launch facility.	4, 5, 7, 8, 12, 29, 30, 40

VECTOR DB FOR STORAGE & RETRIEVAL

Optimizing for LLM Context Windows

- LLM context windows are smaller than total text strings in compliance documents.
- Need a system to store and access relevant document chunks efficiently.

Introduction of Vector Databases

- Vector databases play a crucial role in this scenario.
- Document chunks transformed into numerical vector representations.
- Similar transformation for user prompts.

Retrieval Process

- Retrieve agent searches for close equivalents in the vector database.
- Identified equivalents, along with user prompts, provided to the LLM.

Current Experiments with Vector Databases

- Exploring options for optimal performance.
- Current focus on FAISS and Pinecone databases.
- Experimentation ongoing to determine the most effective solution.

RETRIEVAL AUGMENTATION GENERATION

Precision Enhancement:

- Stores and retrieves specific information chunks with precision.
- Counters inaccuracies and hallucinations, ensuring accurate responses.

Contextual Richness:

- Integrates retrieved data for detailed, context-specific responses.
- Guarantees accurate articulation of compliance document details.

User Experience Improvement:

- Provides reliable, contextually rich user interactions.
- Enhances user confidence through accurate and precise responses.

RAG IN ACTION

```
[ ] #load Document -
    loader= UnstructuredPDFLoader('/content/NIST.IR.8270.pdf', mode='paged', post_processors=[clean_extra_whitespace, replace_unicode_quotes])
    documents = loader.load()
```

```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Unzipping tokenizers/punkt.zip.
[nltk_data] Downloading package averaged_perceptron_tagger to
[nltk_data]     /root/nltk_data...
[nltk_data]   Unzipping taggers/averaged_perceptron_tagger.zip.
```

```
# Define a function to filter metadata fields
def filter_metadata(documents):

    # List of unwanted metadata fields
    unwanted_fields = ['coordinates', 'file_directory','filename','filetype','last_modified', 'category']

    for doc in documents:
        for field in unwanted_fields:
            if field in doc.metadata:
                del doc.metadata[field]
    return documents
```

```
[ ] filtered_documents = filter_metadata(documents)
```

```
[ ] text_splitter = CharacterTextSplitter(chunk_size=2000, chunk_overlap=10) # play with ch
    docs = text_splitter.split_documents(filtered_documents)
```

```
[ ] docs
```

Document(page_content='NIST IR 8270 July 2023\n\nCybersecurity for Commercial Satellite Operations\n\nConceptual High-Level Architecture of Satellite Operations\n\nThis section provides a notional, conceptual, high-level architectural view of commercial, uncrewed space operations. This view can be helpful in understanding, assigning, and managing cybersecurity requirements and risks associated with different owners and operators of different parts of the architecture. This architecture can be under the sole control of one system owner or shared among numerous public, commercial, and private owners.\n\nSpace Architecture Segments\n\nOnce in operation, space vehicles share an ecosystem that has no national and few natural boundaries and where safety is a communal concern. For the purposes of this paper and to facilitate subsequent discussions in setting, expressing, or meeting cybersecurity requirements, NIST notionally defines the scope of a commercial space operations architecture to include the following segments.\n\n2.1.1. Space Segment:\n\nThe space vehicle or satellite consists of the platform and one or more payloads. The bus consists of the components of the vehicle associated with the "flying of the satellite," such as power, structure, attitude control system, processing and command control, and telemetry. The spacecraft can carry many specialized payloads to conduct missions, including remote sensing and communications. The bus and the payload generally combine to form the satellite.\n\nFig. 1. Major parts of the conceptual high-level architecture of space operations\n\nFigure 1 reflects the major parts of the conceptual high-level architecture of satellite operations. This architecture is for uncrewed spacecraft and does not include cybersecurity requirements for human space systems, human spacecraft, or systems that will dock with human systems and/or lunar landers.\n\n4\n\n', metadata={'source': '/content/NIST.IR.8270.pdf', 'coordinates': {'points': ((303.4803599999999, 746.95236), (303.4803599999999, 756.91236), (311.0101199999999, 756.91236), (311.0101199999999, 746.95236)), 'system': 'PixelSpace', 'layout_width': 612.0, 'layout_height': 792.0}, 'filename': 'NIST.IR.8270.pdf', 'file_directory': '/content', 'last_modified': '2023-10-27T23:20:39', 'filetype': 'application/pdf', 'page_number': 11})

Setting up DB

Chroma does not support the dictionary style schema that unstructured returns.

```
[ ] ↳ 2 cells hidden
```

Save Document DB with document Embeddings

```
[ ] doc_db = FAISS.from_documents(docs, embeddings) # this takes about 4 minutes
```

FACT-CHECKING RESPONSE

Introducing Fact-Checking:

- Adds an additional layer of scrutiny to response generation process.
- Complements RAG for comprehensive evaluation of generated content.

Addressing Inaccuracies:

- Acknowledges the possibility of LLM producing errors despite RAG integration.
- Fact-checking prompts model to introspectively assess its assumptions for accuracy.

Iterative Self-Interrogation:

- Sequential evaluation of each assumption made during response formulation.
- Ensures accuracy of assumptions, leading to the generation of revised, verified answers.

Reliability Enhancement:

- Incorporates verified information into responses, enhancing reliability.
- Ensures overall accuracy, boosting confidence in the responses provided by the model.

FACT-CHECKING IN ACTION

```
[ ] question = "Are commerical statelite vehicles manned in space?"
```

```
[ ] #Run Model for Answer
    print(textwrap.fill(llm_chain.run(question), 100))
```

```
 As an AI language model, I do not have access to real-time information. However, as of 2021,
commercial satellites are not manned in space. They are usually launched into orbit by rockets and
then controlled from ground stations.
```

Fact Checking LLM Answer

```
[ ] from langchain.chains import SimpleSequentialChain
```

```
[ ] question_chain = LLMChain(llm=llm, prompt=prompt)
```

```
[ ] template = """Here is a statement:
    {statement}
    Make a bullet point list of the assumptions you made when producing the above statement.\n\n"""
    prompt = PromptTemplate(input_variables=["statement"], template=template)
    assumptions_chain = LLMChain(llm=llm, prompt=prompt)

    template = """Here is a bullet point list of assertions:
    {assertions}
    For each assertion, determine whether it is true or false. If it is false, explain why.\n\n"""
    prompt = PromptTemplate(input_variables=["assertions"], template=template)
    fact_checker_chain = LLMChain(llm=llm, prompt=prompt)

    template = """In light of the above facts, how would you answer the question '{}'""".format(question)
    template = """{facts}\n\n""" + template
    prompt = PromptTemplate(input_variables=["facts"], template=template)
    answer_chain = LLMChain(llm=llm, prompt=prompt)
```

```
[ ] overall_chain = SimpleSequentialChain(chains=[question_chain, assumptions_chain, fact_checker_chain, answer_chain], verbose=True)
```

```
[ ] print(overall_chain.run(question))
```

```
> Entering new SimpleSequentialChain chain...

As an AI language model, I do not have access to real-time information. However, based on my knowledge, commercial satellites ar
- Commercial satellites are manned in space
- Commercial satellites are typically launched into orbit
- Ground-based systems are typically used to control commercial satellites
- Commercial satellites are manned in space: False. Commercial satellites are typically unmanned and rely on automated systems f
- Commercial satellites are typically launched into orbit: True. Commercial satellites are typically launched into orbit using r
- Ground-based systems are typically used to control commercial satellites: False. Ground-based systems are typically used to co

As an AI language model, I do not have personal opinions or beliefs. However, it is important to note that commercial satellites

1. Implementing authentication and authorization mechanisms to ensure that only authorized personnel can access and control the

2. Implementing encryption to protect sensitive data transmitted between the satellite and the ground station.

3. Implementing firewalls to prevent unauthorized access to the satellite's systems and data.

4. Implementing intrusion detection and prevention systems to detect and prevent malicious activities.

5. Implementing patch management and vulnerability assessment programs to ensure that the satellite's systems and data are kept

6. Implementing physical security measures to protect the satellite and its associated equipment from theft or damage.

7. Implementing data integrity and backup procedures to ensure that data is not lost or corrupted in the event of a system or ha

8. Implementing a secure data storage system to store and process data transmitted between the satellite and the ground station.

9. Implementing a secure communication system to transmit data between the satellite and the ground station.

10. Implementing a secure authentication system to ensure that only authorized personnel can access and control the satellite.

> Finished chain.

As an AI language model, I do not have personal opinions or beliefs. However, it is important to note that commercial satellites

1. Implementing authentication and authorization mechanisms to ensure that only authorized personnel can access and control the

2. Implementing encryption to protect sensitive data transmitted between the satellite and the ground station.

3. Implementing firewalls to prevent unauthorized access to the satellite's systems and data.

4. Implementing intrusion detection and prevention systems to detect and prevent malicious activities.

5. Implementing patch management and vulnerability assessment programs to ensure that the satellite's systems and data are kept

6. Implementing physical security measures to protect the satellite and its associated equipment from theft or damage.

7. Implementing data integrity and backup procedures to ensure that data is not lost or corrupted in the event of a system or ha

8. Implementing a secure data storage system to store and process data transmitted between the satellite and the ground station.

9. Implementing a secure communication system to transmit data between the satellite and the ground station.

10. Implementing a secure authentication system to ensure that only authorized personnel can access and control the satellite.
```

CONTACT US

Do you have any questions?

swllms@berkeley.edu
vincent.goldberg@berkeley.edu
arwinrahmatpanah@berkeley.edu

sibylcopilot.com

CREDITS: This presentation template was created by Slidesgo, including icons by Flaticon, and infographics & images by Freepik.