P4.3-883

Enhancing Information Access to CTBTO Staff through Generative AI: A Smart Chat Application for Simplified Access to Complex Directives and Protocols

Marko Bosancic
Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO)

CTBTO
PREPARATORY COMMISSION

INTRODUCTION AND MAIN RESULTS

Presenting a Generative AI based application that enhances organization staff access to complex directives and protocols. Providing natural language conversation queries with accurate, reference backed responses and iterative refinement, the system reduces the time and effort of manual search and interpretation of documentation, thereby improving operational efficiency and knowledge accessibility

DISCLAIMER: This presentation has been generated for illustration purposes only.

SnT 2025
CTBT: SCIENCE AND TECHNOLOGY CONFERENCE
8 SEPTEMBER ONLINE DAY
9 TO 12 SEPTEMBER AT HOFBURG PALACE, VIENNA & ONLINE

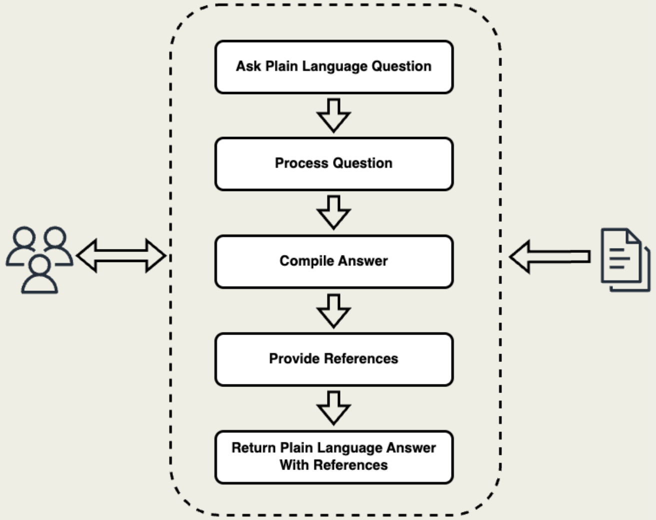
Enhancing Information Access to CTBTO Staff through Generative AI: A Smart Chat Application for Simplified Access to Complex Directives and Protocols

Marko Bosancic
Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO)

P4.3-883

Introduction

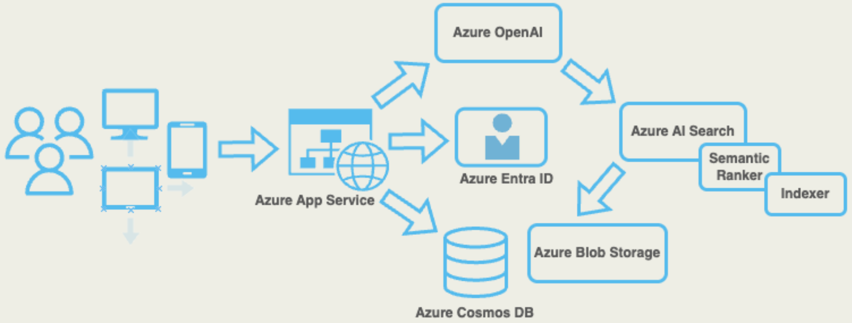
Successful market adaptation of large language models (LLMs) and the rapidly expanding capabilities of generative AI has demonstrated how such technologies can be applied to various organizational challenges with growing demand for assistive software that supports staff in navigating information heavy processes and minimizing cognitive burden. CTBTO approach was to utilize applications based on Generative AI aimed at enhancing access to complex directives and protocols by enabling natural language conversation-like queries, delivering accurate and reference-backed responses and allowing iterative refinement of results.


Ask Plain Language Question
Process Question
Compile Answer
Provide References
Return Plain Language Answer With References

Smart Search Workflow

DISCLAIMER: This presentation has been generated for illustration purposes only.

System Architecture and Implementation



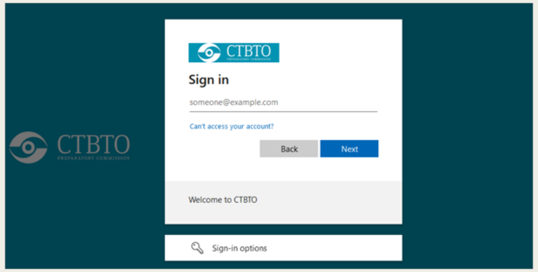
Retrieval Augmented Generation (RAG) architecture

Generative AI chat application was implemented using Microsoft Azure cloud services to provide rapid development of scalable, secure and high-performance solutions for CTBTO staff. The system integrates multiple Azure resources to handle document storage, content search and AI processing orchestrated according to a market proven Retrieval Augmented Generation (RAG) design pattern, together with authentication and conversation history management.

- **Azure AI Search**: Indexing the directives, protocols and other related documentation to provide best ranked search results to the LLM
- **Azure OpenAI**: GPT-4 LLM is deployed for natural language reasoning capabilities to generate responses to user prompts based on provided results
- **Azure Blob Storage**: Central repository for targe-domain directive and protocol documents, accessible by search indexers, to constrain generative AI responses
- **Azure Cosmos DB**: Stores conversation histories, enabling iterative refinement of user queries and continuity across sessions
- **Azure App Service**: Hosts the user-facing custom developed application, providing a responsive UI for all devices and secure interface for staff queries
- **Azure Entra ID**: Ensures secure, role-based organizational 2F Authentication and access to the application for selected or all organization staff, depending on the business domain sensitivity the application covers

Authentication and Authorization

CTBTO utilizes Azure Entra ID services for authentication and authorization across different systems. Federated identity for seamless Authentication and Role-Based Access Control (RBAC) for managing authorization through Azure App Services container provided endpoints /.auth/login/aad and /.auth/me for principal data.



Azure Entra ID 2FA

CTBTO
PREPARATORY COMMISSION

SnT 2025
8 SEPTEMBER ONLINE DAY
9 TO 12 SEPTEMBER AT HOFBURG PALACE, VIENNA & ONLINE
CTBT: SCIENCE AND TECHNOLOGY CONFERENCE

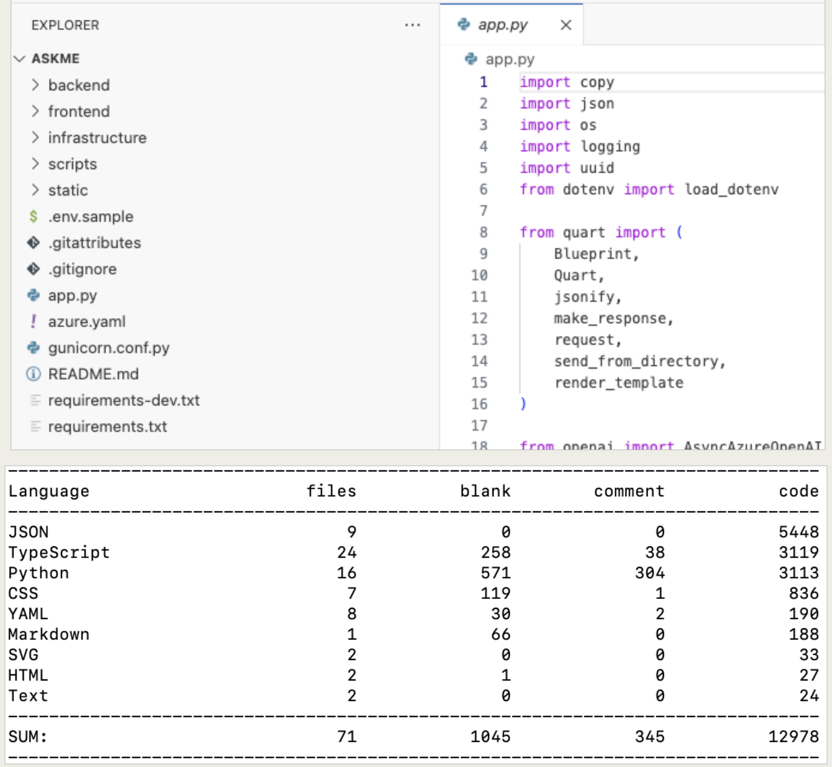
Enhancing Information Access to CTBTO Staff through Generative AI: A Smart Chat Application for Simplified Access to Complex Directives and Protocols

Marko Bosancic
Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO)

P4.3-883

Software development and deployment

To enable intuitive and effortless utilization of AI capabilities on targeted domain documentation, custom software was developed using Node.js, TypeScript, React, and FluentUI components on the Frontend and lightweight Quart Python framework with multiple Azure libraries on the Backend.



```
EXPLORER                          ...
∨ ASKME
  > backend
  > frontend
  > infrastructure
  > scripts
  > static
  $ .env.sample
  .gitattributes
  .gitignore
  app.py
  ! azure.yaml
  gunicorn.conf.py
  README.md
  requirements-dev.txt
  requirements.txt
```

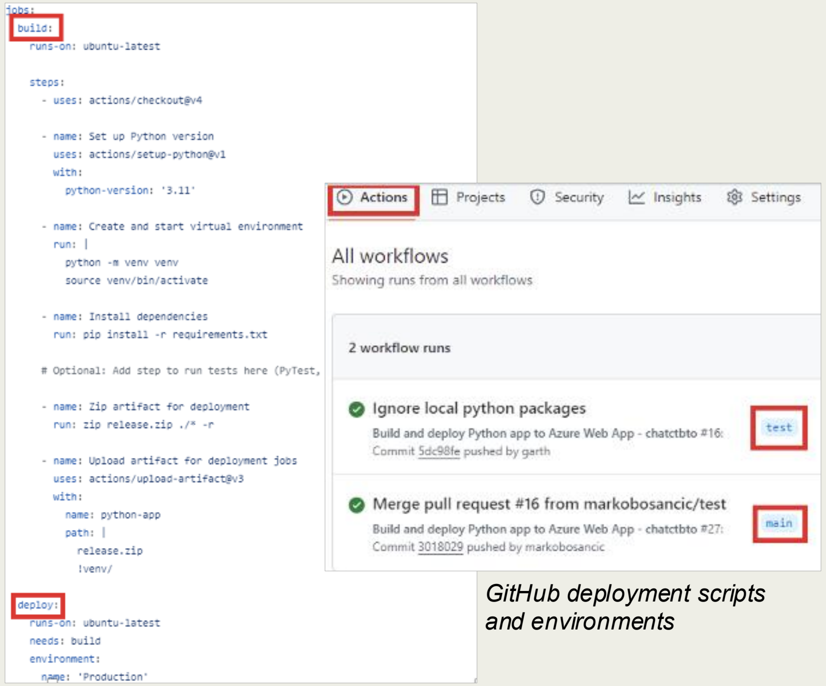
```python
# app.py
import copy
import json
import os
import logging
import uuid
from dotenv import load_dotenv

from quart import (
    Blueprint,
    Quart,
    jsonify,
    make_response,
    request,
    send_from_directory,
    render_template
)

from openai import AsyncAzureOpenAI
```

```
-------------------------------------------------------------------------------
Language                     files          blank        comment           code
-------------------------------------------------------------------------------
JSON                             9              0              0           5448
TypeScript                      24            258             38           3119
Python                          16            571            304           3113
CSS                              7            119              1            836
YAML                             8             30              2            190
Markdown                         1             66              0            188
SVG                              2              0              0             33
HTML                             2              1              0             27
Text                             2              0              0             24
-------------------------------------------------------------------------------
SUM:                            71           1045            345          12978
-------------------------------------------------------------------------------
```

Lightweight codebase

DISCLAIMER: This presentation has been generated for illustration purposes only.

GitHub repository is used for software versioning with GithHub Actions for seamless Azure App Service deployments compose out of main PremiumV3 (P1v3 - 2 vCPU and 8GB of memory) and multiple "Slot" instances for different development environments and stages.



```yaml
jobs:
  build:
    runs-on: ubuntu-latest

    steps:
      - uses: actions/checkout@v4

      - name: Set up Python version
        uses: actions/setup-python@v1
        with:
          python-version: '3.11'

      - name: Create and start virtual environment
        run: |
          python -m venv venv
          source venv/bin/activate

      - name: Install dependencies
        run: pip install -r requirements.txt

      # Optional: Add step to run tests here (PyTest,

      - name: Zip artifact for deployment
        run: zip release.zip ./* -r

      - name: Upload artifact for deployment jobs
        uses: actions/upload-artifact@v3
        with:
          name: python-app
          path: |
            release.zip
            !venv/

  deploy:
    runs-on: ubuntu-latest
    needs: build
    environment:
      name: 'Production'
```

All workflows
Showing runs from all workflows

2 workflow runs

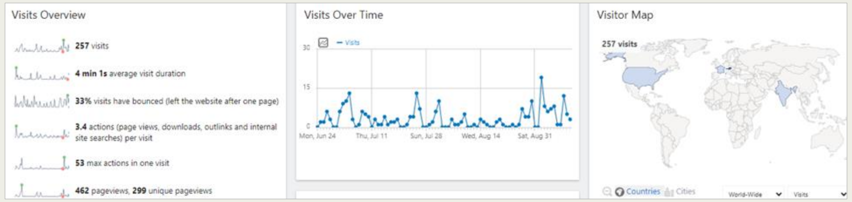
✓ Ignore local python packages — test
Build and deploy Python app to Azure Web App - chatctbto #16:
Commit 5dc98fe pushed by garth

✓ Merge pull request #16 from markobosancic/test — main
Build and deploy Python app to Azure Web App - chatctbto #27:
Commit 3018029 pushed by markobosancic

GitHub deployment scripts and environments

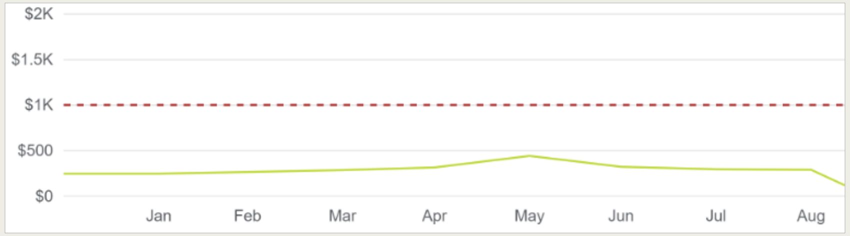
Analytics and costs

Implemented Matomo analytics showed early traction with staff members shortly after release with development and post-release Azure costs remaining negligible.



Visits Overview
- 257 visits
- 4 min 1s average visit duration
- 33% visits have bounced (left the website after one page)
- 3.4 actions (page views, downloads, outlinks and internal site searches) per visit
- 53 max actions in one visit
- 462 pageviews, 299 unique pageviews

2-Months Analytics after the initial release



Azure costs for the development and post-release period

Results

After an initial investment of time and effort to develop a generic and highly customizable smart chat solution, multiple instances of the application were rapidly deployed to serve various business domains within the organization. Access to each instance is tailored based on the sensitivity of the information, ranging from restricted access for select staff to full access for all personnel. Instances and POC-s for different organizational contexts cover analysis of the International Labor Organization documents for the Legal department, operational Manuals for the On-Site Inspection teams, clarification documents for results presented on the Secure Web Portal and simplifying Administrative Directives for CTBTO Staff.

CTBTO PREPARATORY COMMISSION

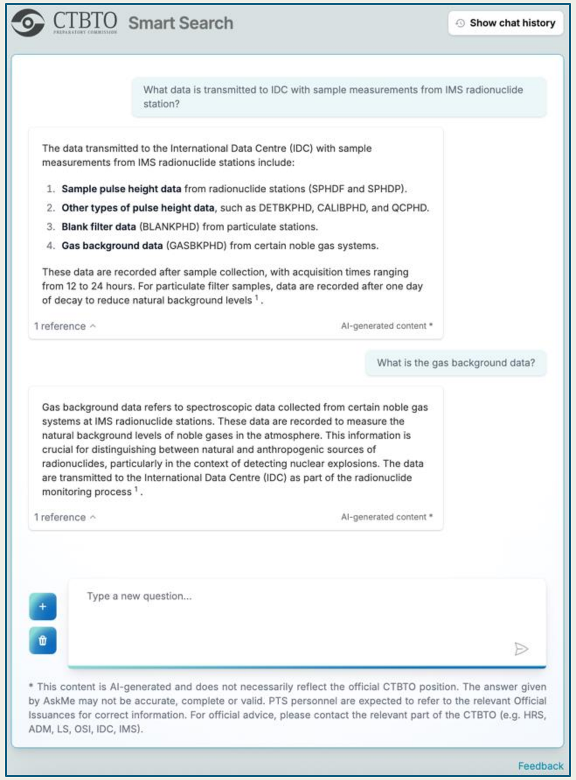
SnT 2025
CTBT: SCIENCE AND TECHNOLOGY CONFERENCE
8 SEPTEMBER ONLINE DAY
9 TO 12 SEPTEMBER AT HOFBURG PALACE, VIENNA & ONLINE

Enhancing Information Access to CTBTO Staff through Generative AI: A Smart Chat Application for Simplified Access to Complex Directives and Protocols

Marko Bosancic
Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO)

P4.3-883

The following graphics illustrate the design, functionality and practical application of the solutions utilized in the CTBTO organization.



What data is transmitted to IDC with sample measurements from IMS radionuclide station?

The data transmitted to the International Data Centre (IDC) with sample measurements from IMS radionuclide stations include:

1. **Sample pulse height data** from radionuclide stations (SPHDF and SPHDP).
2. **Other types of pulse height data**, such as DETBKPHD, CALIBPHD, and QCPHD.
3. **Blank filter data** (BLANKPHD) from particulate stations.
4. **Gas background data** (GASBKPHD) from certain noble gas systems.

These data are recorded after sample collection, with acquisition times ranging from 12 to 24 hours. For particulate filter samples, data are recorded after one day of decay to reduce natural background levels[1].

1 reference ^ AI-generated content *

What is the gas background data?

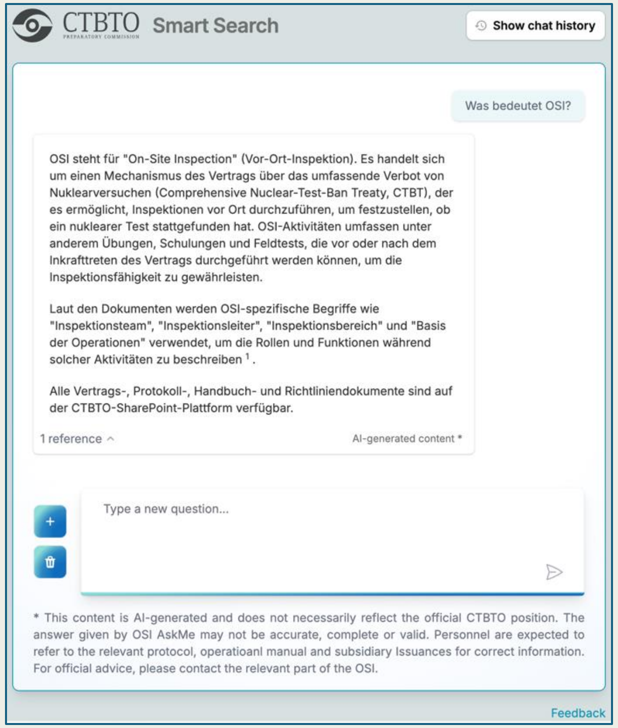
Gas background data refers to spectroscopic data collected from certain noble gas systems at IMS radionuclide stations. These data are recorded to measure the natural background levels of noble gases in the atmosphere. This information is crucial for distinguishing between natural and anthropogenic sources of radionuclides, particularly in the context of detecting nuclear explosions. The data are transmitted to the International Data Centre (IDC) as part of the radionuclide monitoring process[1].

1 reference ^ AI-generated content *

* This content is AI-generated and does not necessarily reflect the official CTBTO position. The answer given by AskMe may not be accurate, complete or valid. PTS personnel are expected to refer to the relevant Official Issuances for correct information. For official advice, please contact the relevant part of the CTBTO (e.g. HRS, ADM, LS, OSI, IDC, IMS).

Simplifying complex topics

Utilizing the tool in different languages



Was bedeutet OSI?

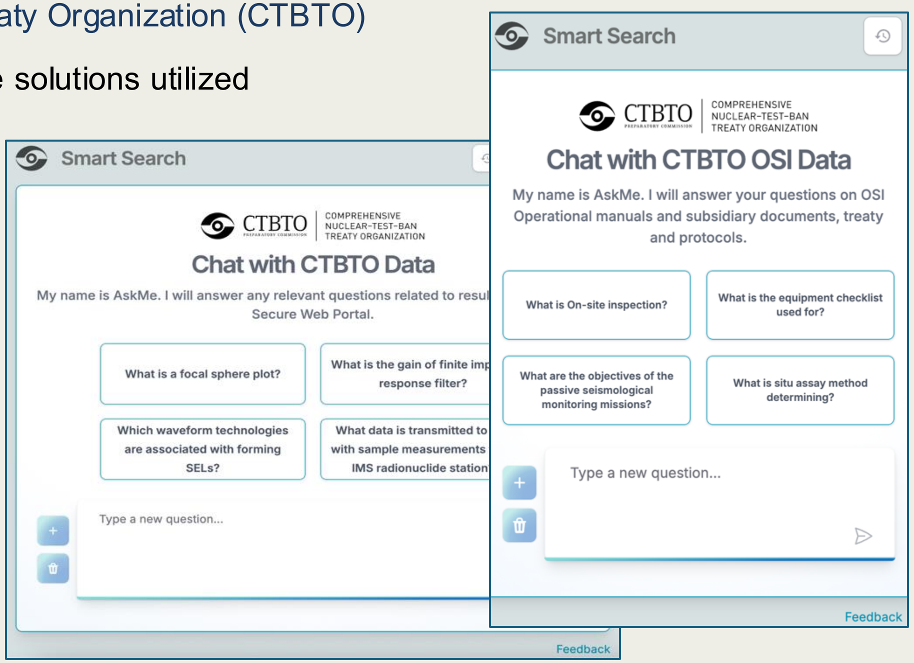
OSI steht für "On-Site Inspection" (Vor-Ort-Inspektion). Es handelt sich um einen Mechanismus des Vertrags über das umfassende Verbot von Nuklearversuchen (Comprehensive Nuclear-Test-Ban Treaty, CTBT), der es ermöglicht, Inspektionen vor Ort durchzuführen, um festzustellen, ob ein nuklearer Test stattgefunden hat. OSI-Aktivitäten umfassen unter anderem Übungen, Schulungen und Feldtests, die vor oder nach dem Inkrafttreten des Vertrags durchgeführt werden können, um die Inspektionsfähigkeit zu gewährleisten.

Laut den Dokumenten werden OSI-spezifische Begriffe wie "Inspektionsteam", "Inspektionsleiter", "Inspektionsbereich" und "Basis der Operationen" verwendet, um die Rollen und Funktionen während solcher Aktivitäten zu beschreiben[1].

Alle Vertrags-, Protokoll-, Handbuch- und Richtliniendokumente sind auf der CTBTO-SharePoint-Plattform verfügbar.

1 reference ^ AI-generated content *

* This content is AI-generated and does not necessarily reflect the official CTBTO position. The answer given by OSI AskMe may not be accurate, complete or valid. Personnel are expected to refer to the relevant protocol, operatioanl manual and subsidiary Issuances for correct information. For official advice, please contact the relevant part of the OSI.


Smart Search
CTBTO
PREPARATORY COMMISSION
COMPREHENSIVE
NUCLEAR-TEST-BAN
TREATY ORGANIZATION
What is a focal sphere plot?
What is the gain of finite imp response filter?
Which waveform technologies are associated with forming SELs?
What data is transmitted to with sample measurements IMS radionuclide station
Type a new question...
Feedback
Smart Search
CTBTO
PREPARATORY COMMISSION
COMPREHENSIVE
NUCLEAR-TEST-BAN
TREATY ORGANIZATION
What is On-site inspection?
What is the equipment checklist used for?
What are the objectives of the passive seismological monitoring missions?
What is situ assay method determining?
Type a new question...
Feedback

Chat with CTBTO Data

My name is AskMe. I will answer any relevant questions related to resul Secure Web Portal.

Chat with CTBTO OSI Data

My name is AskMe. I will answer your questions on OSI Operational manuals and subsidiary documents, treaty and protocols.

Web client Interface on different devices

Conclusion

In conclusion, the AI based systems significantly reduces the time and effort associated with manual search and interpretation of complex or domain specific documentation. Removing repetitive workloads, fostering consistency in interpretation, enables staff to focus more effectively on their core responsibilities, benefiting not only those who seek answers but also staff responsible for interpreting and providing relevant information. In doing so, such systems strengthens both operational efficiency and knowledge accessibility across the organization, operated at negligible costs with highly controllable and predictable expenses.

Presented solutions are sustainable for long-term organizational adoption and more importantly, the potential for further expansion and contribution to organizational efficiency is endless.

DISCLAIMER: This presentation has been generated for illustration purposes only.

CTBTO
PREPARATORY COMMISSION