



Fake Conversation Generator

This Python script creates fake conversations and synthesizes audio using OpenAI's text-to-speech model. The vCon files are then uploaded to an Amazon S3 bucket.



by Thomas Howe



The Problem: Redacting and Detecting PII in Conversations

At the upcoming vCon hackathon, participants will be tackling the challenge of redacting and detecting personally identifiable information (PII) in conversations. This is crucial for ensuring data privacy and security. We're looking for conversations with abundant personal information to test our solutions. Join us in this exciting endeavor!

vCons : A PDF for Conversations

Parties

Identities of the participants



Dialogs

What was communicated



Analysis

Tracks AI, ML interactions



Attachments

Supporting and associated data



Tamper proof, signed, encryptable JSON

**vCons carry
conversations across
networks, enabling
responsible
management of the
most personal of data.**

The Plan?

- ☐ Pick some fake names, E.164s and emails
- ☐ Pick a random business (bike store), pick a random problem (shipping), and pick a random emotion (ennui)
- ☐ Ask an LLM to create a fake transcript with (some, lots, none) of PII
- ☐ Use a TTS engine to create an MPG file of the conversation
- ☐ Create the vCon, load it into S3

ake Conversation enerator

This app generates fake conversations between customer and an agent. The conversation is generated based on a prompt and includes the names of the agent and customer, the business, problem, and the emotion of the customer. The conversation is then synthesized into an audio file. A vCon is created then it is uploaded into S3.

Number of Conversations to Generate

4

Generate Conversation(s)

Generating conversation 1 of 4

Synthesizing conversation audio

It worked.



Features

Fake Conversations

Generates fake conversations based on user-defined prompts.

Text-to-Speech

Synthesizes audio for each line of the conversation using OpenAI's model.

vCon Files

Creates vCon files containing conversation metadata, transcripts, and audio URLs, but without any of the actual PII

You can too!

- https://github.com/vcon-dev/vcon_faker contains the complete code (streamlit, python, OpenAI, S3)
- <https://github.com/vcon-dev/fake-vcons> contains the deceptions and misstatements. But, properly formatted.