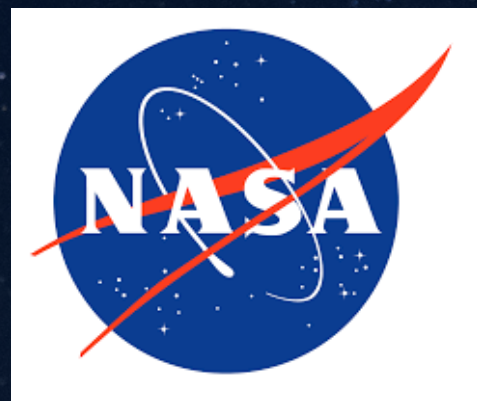




GAMAN
Present...

BUILDING RAG APPLICATIONS



A MARKETPLACE FOR OPEN SCIENCE PROJECTS

In collaboration
with **NASA**

“When I decided to join *TruEra Challenge: Build LLM Applications with Google Cloud Vertex AI Hackathon*, first thing that came up to my mind was that I aim to expand Large Language Model (LLM) with Google Vertex AI & TruLens for my project which is **NASA: A MARKETPLACE FOR OPEN SCIENCE PROJECTS!**

I did this NASA project when I participated in *NASA Space Apps Challenges Hackathon 2023* last October 2023.

Thus, I want to use NASA Marketplace as my case for this *TruEra Challenge Hackathon*.”

FICKY ALKARIM

*CTO & COO GAMAN AI
COFOUNDER*



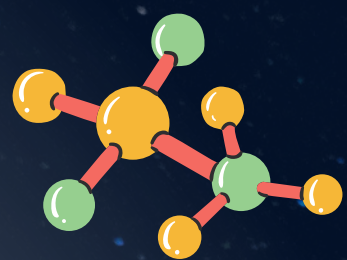
**OKAY, LET ME START THE
PRESENTATION NOW ...**



PAIN POINT

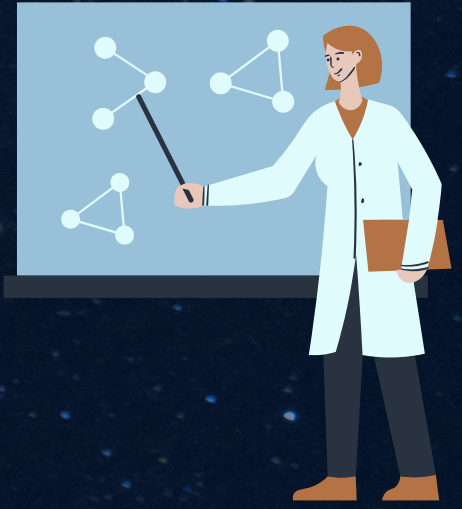
Recently, there is no marketplace facilitating the connection between creators and contributors in open science projects.

This absence hinders the opportunity for collaboration and the development of professional relationships for those interested in working on compelling open science projects & research.



Project Creators are the one who initiate the science projects (such as: enterprises, universities, government agencies, etc)

Project Contributors are the one who contribute to projects and they have particular capabilities which match with the skill requirements in science projects (such as: researchers, scientists, students, professionals, etc)



Also, another Pain Point...

For contributors, who want to search for jobs in open science projects in NASA & Space Agency Partner Web, will be challenging to find detail of project and data, since it's too sophisticated and too much materials in one page.

The screenshot shows the homepage of the Federal Crowdsourcing and Citizen Science Catalog. The URL is <https://www.citizenscience.gov/catalog/#>. The page features a search bar under "Search Projects" and a "Filter Projects" section with a "View by Status" dropdown. Three project cards are visible:

- Chesapeake Water Watch**: Status: Active - recruiting volunteers; Geographic Scope: Tidal regions of the Chesapeake Bay.
- Cooperative Tagging Center**: Status: Active - recruiting volunteers; Geographic Scope: Atlantic Ocean, Gulf of Mexico, and adjacent...
- Are we alone in the...**: Status: In Planning or Pending - recruiting volunteers; Geographic Scope: Worldwide.

The screenshot shows the GeneLab Open Science website. The URL is <https://genelab.nasa.gov>. The page is organized into several columns of project cards:

- TRANSCRIPTOMICS**: Includes projects like OSD-554 (Biofilm Formation), OSD-613 (Single nuclei transcriptomics of brain tissue), OSD-612 (Single nuclei transcriptomics of brain tissue), OSD-580 (Transcriptional profiling of heart tissue), OSD-576 (Transcriptional profiling of tibialis anterior muscle), and OSD-577 (3D Human Umbilical Vein Endothelial Cells).
- GENOMICS (including metagenomics, epigenomics)**: Includes projects like OSD-594 (Automation of plant nucleic acid extraction), OSD-386 (Persistence of Escherichia coli in mizuna), OSD-385 (Persistence of Escherichia coli in red Romaine), OSD-417 (Specific Host Metabolite and Gut Microbiome Alterations), OSD-582 (Resequencing of microbial isolates), and OSD-487 (The Maletth Program).
- PROTEOMICS / METABOLOMICS**: Includes projects like OSD-417 (Specific Host Metabolite and Gut Microbiome Alterations), OSD-581 (Molecular study of Italian ryegrass on Martian regolith), GLDS-514 (Artificial gravity effects on neurotoxicity), GLDS-505 (Proteomic approaches to identify metabolic changes), and GLDS-360 (Effects of low dose space radiation on splenic metabolism).
- PROCESSED**: Includes projects like OSD-374 (Bi-directional and shared epigenomic signatures), OSD-201 (Impact of Antiorthostatic Suspension on Mouse response to Tetanus), OSD-120 (Genetic dissection of Arabidopsis spaceflight transcriptome), OSD-423 (Transcriptional profiling of thymus tissue from rats), OSD-323 (Evaluating the effect of spaceflight on host-pathogen interaction), and OSD-462 (Transcriptional profiling, protein expression, and phosphoproteomics).

At the bottom, there are sections for "WATCH" (GeneLab Chats Interviews Dr Chris Mason) and "PUBLICATIONS" (Skeletal muscle gene expression dysregulation in long-term spaceflights and aging is clock-dependent).

**Thus, we aim to upgrade the
concept of marketplace of open
science projects in NASA...**

**We want everyone to be able to
participate in open science
project all around the world in
just one platform...**



WE OFFER THE SOLUTION...

PROJECT RECOMMENDATION SYSTEM

PROJECT SEARCH ENGINE - AI CHATTING BOT

**COMMUNICATION TOOLS BETWEEN CREATORS &
CONTRIBUTORS**

TRULENS FOR EVALUATING LLM

PROJECT RECOMMENDATION SYSTEM (FEATURE NO.1)

NASA Open Science Project Creator + Contributor Zone

A Project by GAMAN for NASA Space Apps Challenge 2023

We want to create a marketplace for open science projects so that open science projects will be more accessible for everyone who eager to contribute to science projects related to their interests. Also, we want to build a marketplace for project creators to find creators who match with all skills & expertise required in the project. We will be building this marketplace in web application using Machine Learning, Large Language Model (LLM), Generative AI and Natural Language Processing (NLP).

We are hoping that all interactions between project creators and contributors in any NASA Open Science Project will become more accessible, interactive, and fun. In addition, this web app will make NASA Open Data become easier to find and more organized.



NASA Open Science Project Creator + Contributor Zone

Elaborate the details below to look for many Open Science Project opportunities!

Please wait and be patient!

Select a project location where you are interested in:

Florida

Select a agency sponsor or institution that you are interested in:

U.S. Environmental Protection Agency (EPA)

Select the top 10 fields of science of your expertise:

Nature and outd... x Ecology and env... x

Project	Description	Location	keywords	email
Florida Keys Water Watch	The University of Florida IFAS Monroe County Extension developed Florida Keys Water Watch, a community-based volunteer water quality monitoring program to promote awareness of the importance of water quality, reduce nonpoint source pollution, and involve students and citizens to monitor coastal habitats from man-made canals to bays to beaches. Florida Keys Water Watch teaches volunteers to test a site of their choice for dissolved oxygen, pH, salinity and temperature after attending a 4-hour workshop.	Florida	water quality, dissolved oxygen, pH, salinity, Florida Keys, South Florida, coastal	shellykrueger@ufl.edu

First feature, we have developed Machine Learning to arrange the list of open science projects in NASA to be easier for potential contributors to track available open science projects. We use several parameters such as: project ID, Name of Project, Project URL, Project Description, keywords, fields of science, agency sponsor, govt contact name, email, geographic scope, participant tasks, start date, and so on...

This is the Link of App for feature 01: Just Click <https://nasaproject2023-dzgwiak2nirgiw4ibixxjc.streamlit.app>

NOTES: We use NASA Open Science Project Database <https://science.nasa.gov/researchers/open-science/>

<https://www.citizenscience.gov/#>

DEMO FEATURE NO.1 (PLEASE CLICK BELOW)

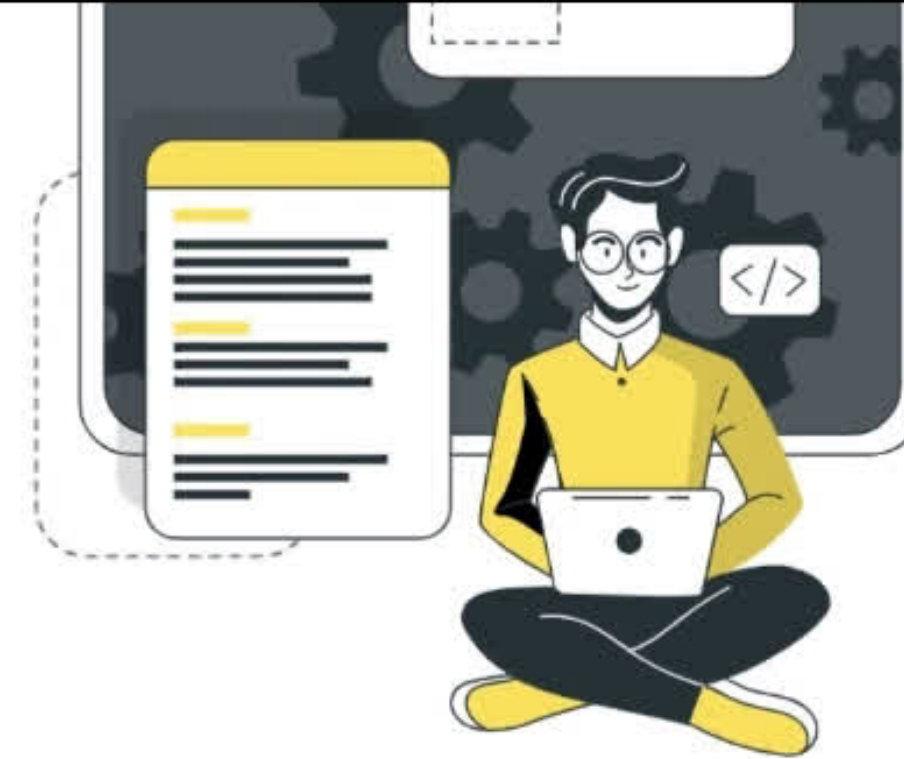
Contributor Zone

A Project by GAMAN for NASA Space Apps Challenge 2023

We want to create a marketplace for open science projects so that open science projects will be more accessible for everyone who eager to contribute to science projects related to their interests. Also, we want to build a marketplace for project creators to find creators who match with all skills & expertise required in the project. We will be building this marketplace in web application using Machine Learning, Large Language Model (LLM), Generative AI and Natural Language Processing (NLP).

We are hoping that all interactions between project creators and contributors in any NASA Open Science Project will become more accessible, interactive, and fun. In addition, this web app will make NASA Open Data become easier to find and more organized.

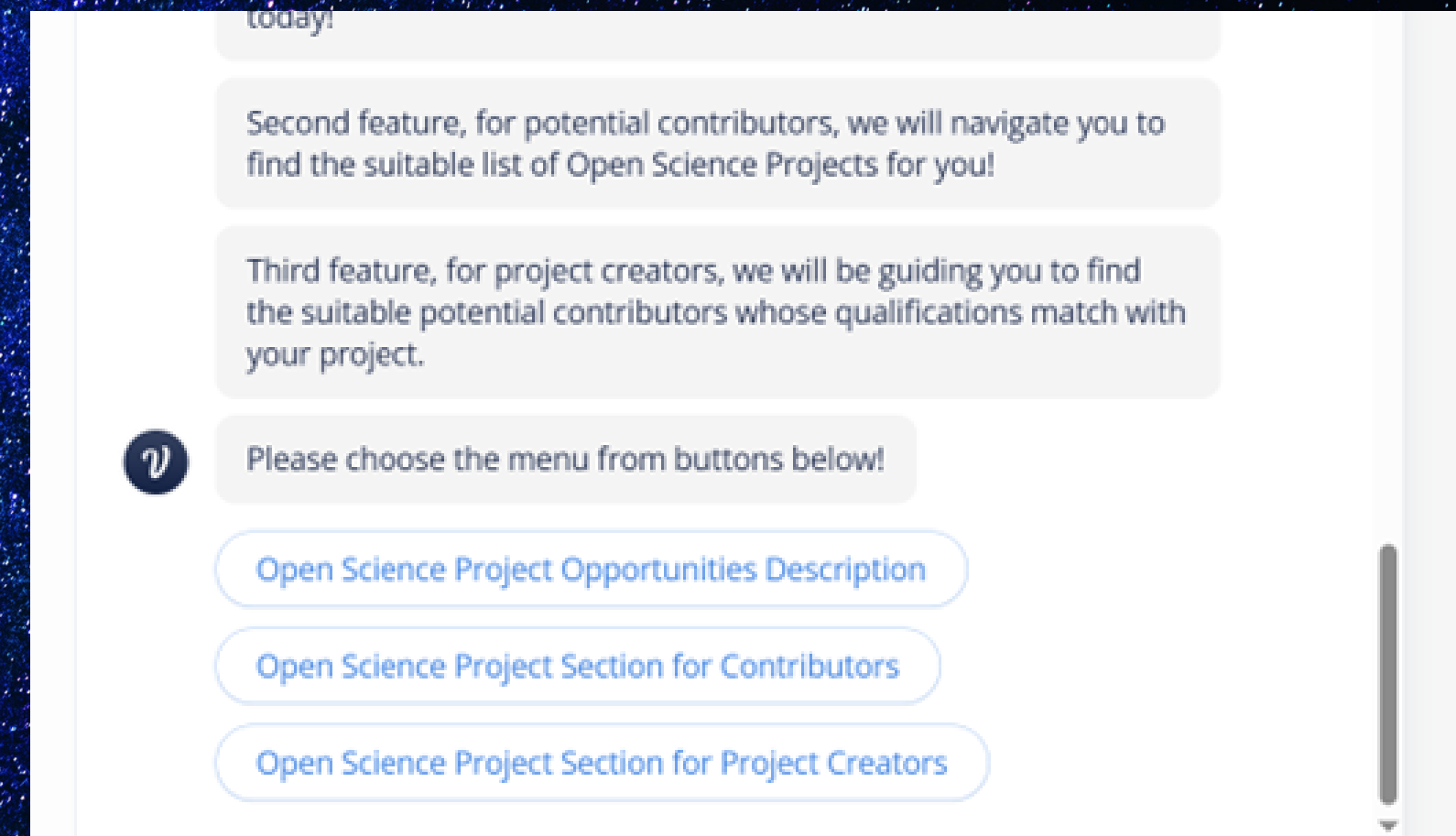
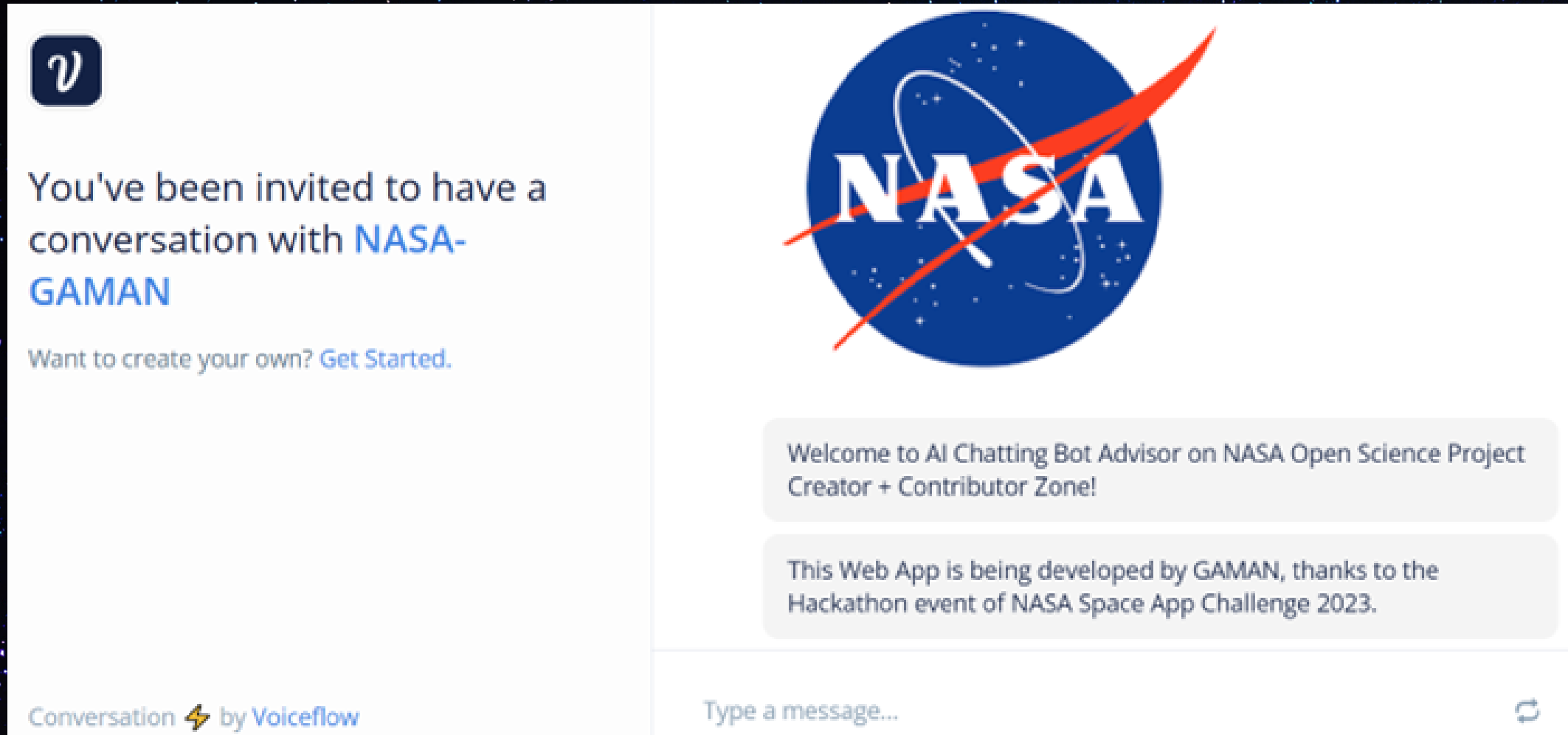
In this website, in Menu APP, we are going to show you how we use Machine Learning to build Recommendation approach for contributors to match their interest with the available open science projects.



< Manage app

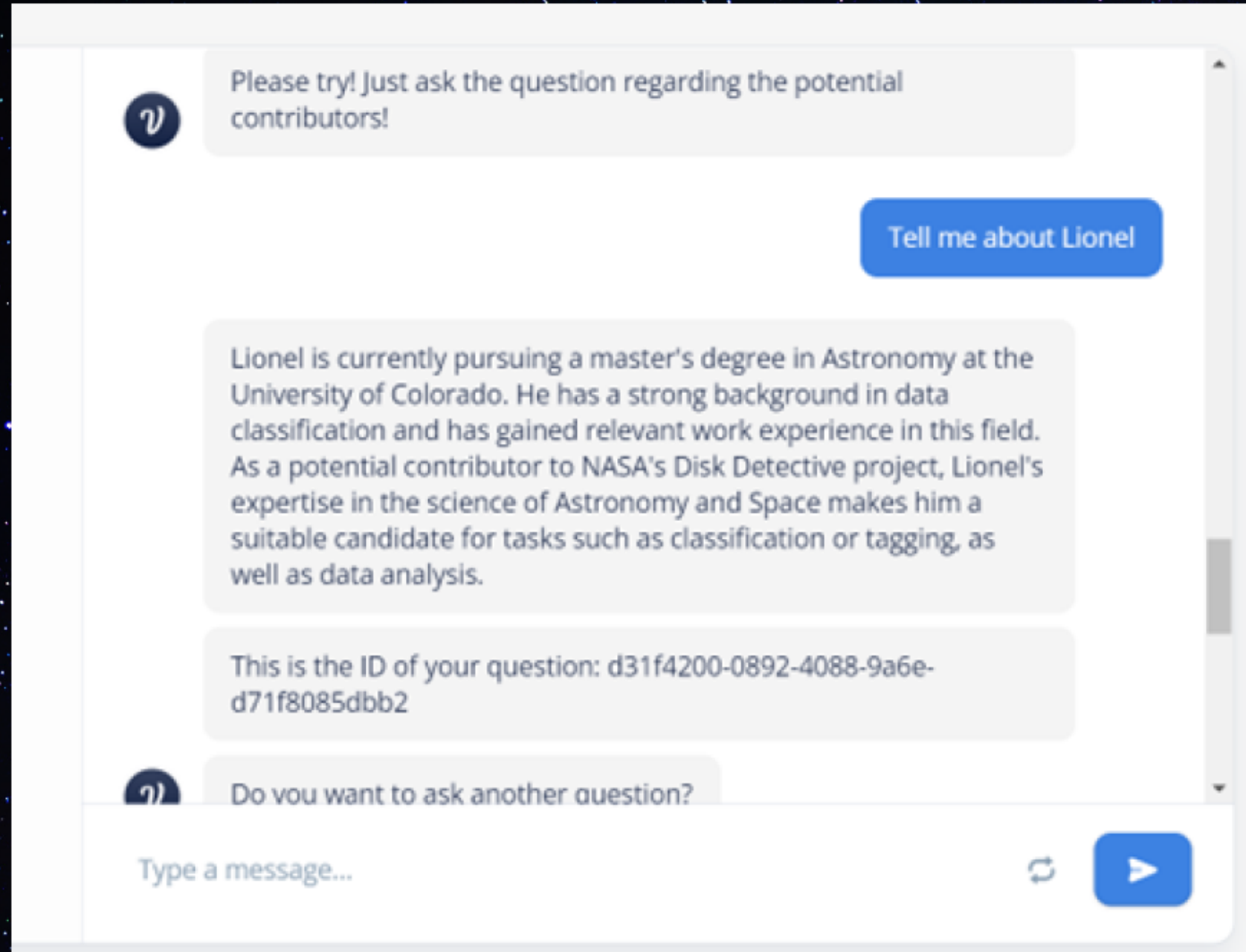
PROJECT SEARCH ENGINE - AI CHATTING BOT (FEATURE NO.2)

Our second feature, we have developed AI Chatting Bot as a complement to what we already built in feature no.1



In our AI Chatting Bot, we facilitate project creators and contributors to look for all information such as what the latest update for open science projects that available in certain region, or how many potential contributors are ready to fill the position of particular fields of science and so on.

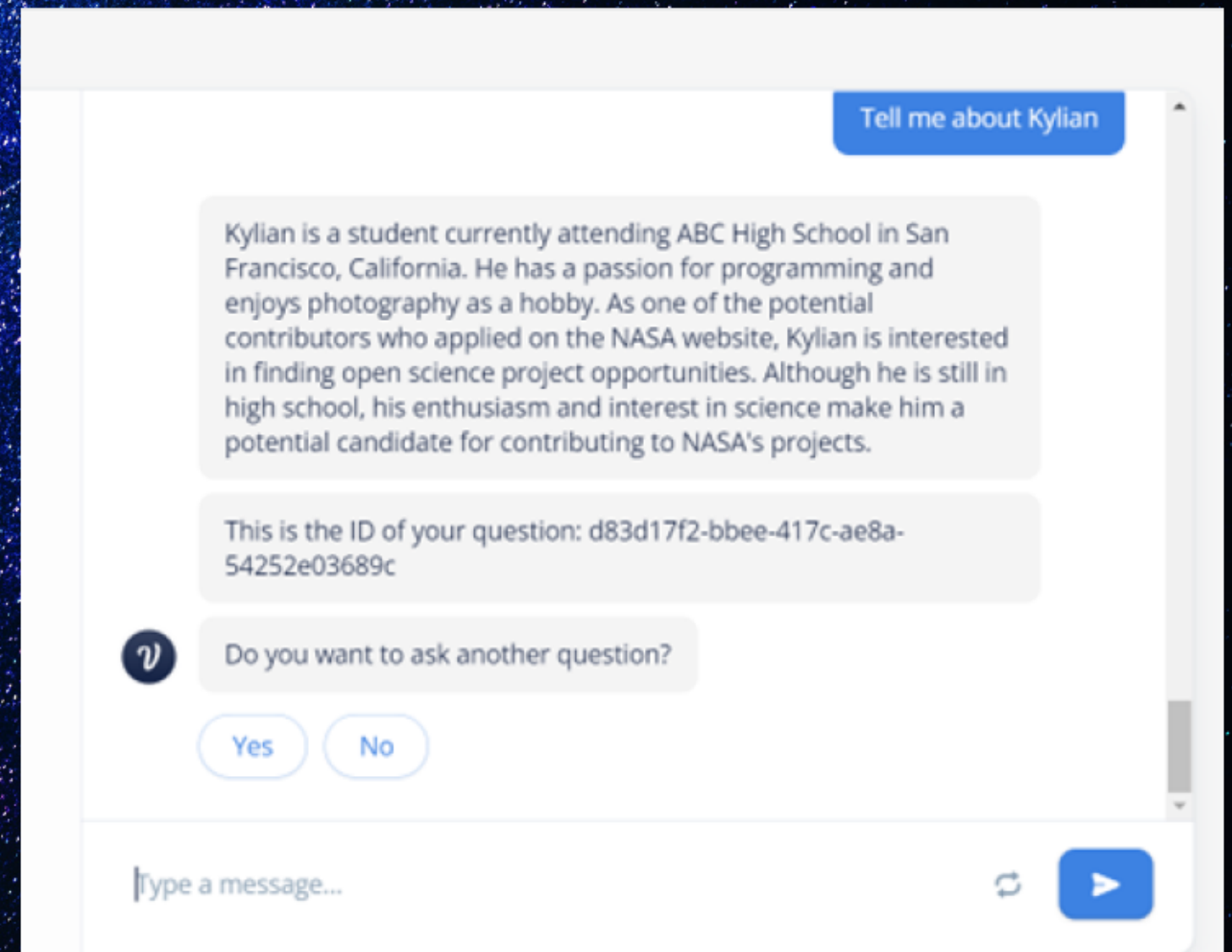
PROJECT SEARCH ENGINE - AI CHATTING BOT (FEATURE NO.2)



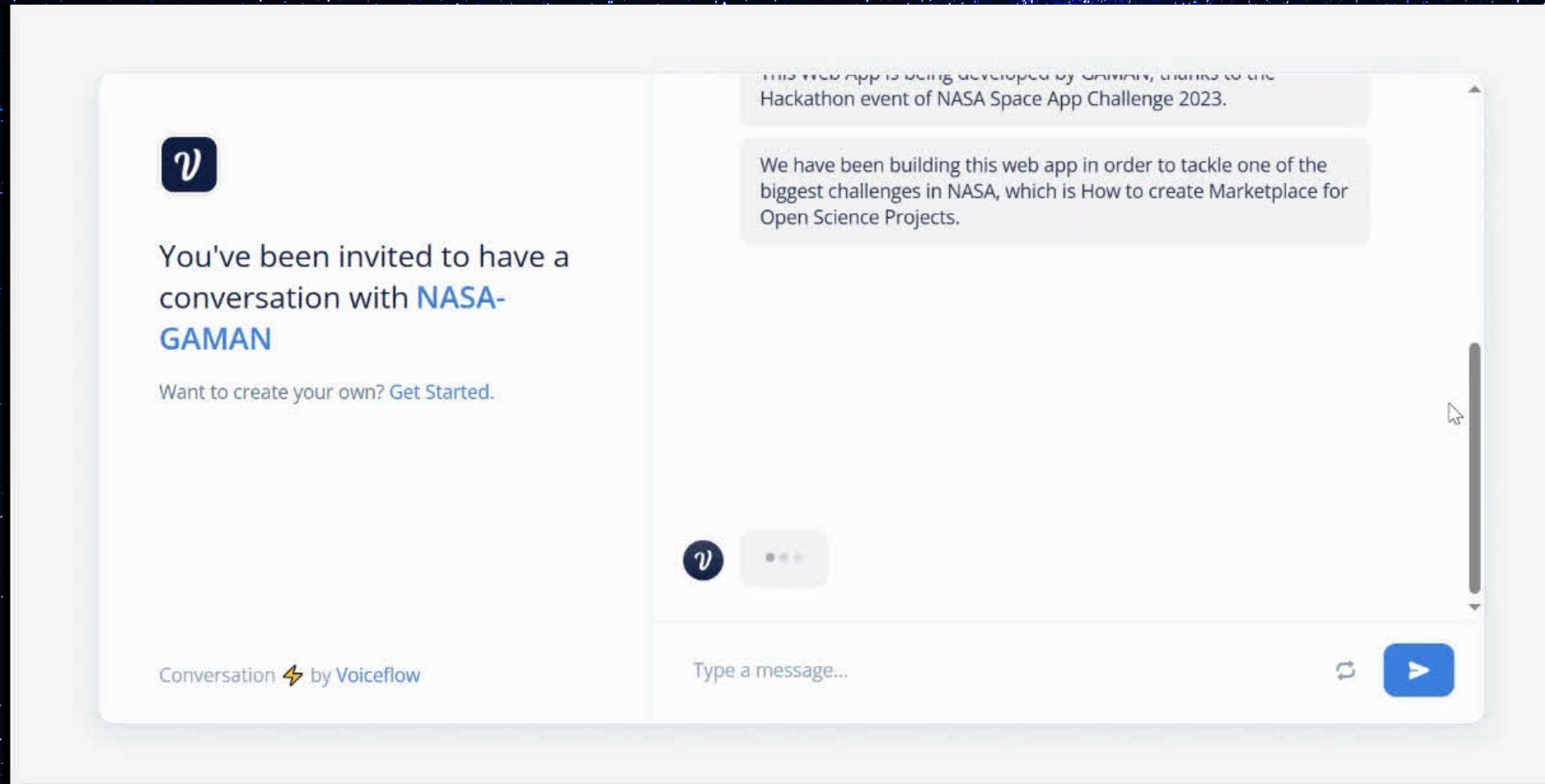
Let say there are 3 potential contributors, Lionel, Cris and Kylian.

Our AI Chatting Bot will help the project creators to do profiling to those three potential contributors.

AI Chatting Bot has also encrypted every chat in order to prevent or minimize the leak of data



DEMO FEATURE NO.2 (PLEASE CLICK BELOW)



The screenshot shows a chat interface with a white background. On the left side, there is a profile card for 'NASA-GAMAN' featuring a dark blue square icon with a white 'V' logo. The text on the card reads: 'You've been invited to have a conversation with **NASA-GAMAN**'. Below this, it says 'Want to create your own? [Get Started.](#)' At the bottom left of the card, it says 'Conversation ⚡ by Voiceflow'. The right side of the interface is a chat window with a scrollable message history. The messages are in light gray bubbles. The top message says: 'This web app is being developed by GAMAN, thanks to the Hackathon event of NASA Space App Challenge 2023.' The second message says: 'We have been building this web app in order to tackle one of the biggest challenges in NASA, which is How to create Marketplace for Open Science Projects.' Below the messages is a dark blue circular icon with a white 'V' logo and a gray bubble containing three dots. At the bottom of the chat window is a text input field with the placeholder 'Type a message...'. To the right of the input field are a refresh icon (two circular arrows) and a blue button with a white right-pointing triangle.

V

You've been invited to have a conversation with **NASA-GAMAN**

Want to create your own? [Get Started.](#)

Conversation ⚡ by Voiceflow

This web app is being developed by GAMAN, thanks to the Hackathon event of NASA Space App Challenge 2023.

We have been building this web app in order to tackle one of the biggest challenges in NASA, which is How to create Marketplace for Open Science Projects.

V ...

Type a message...

↻

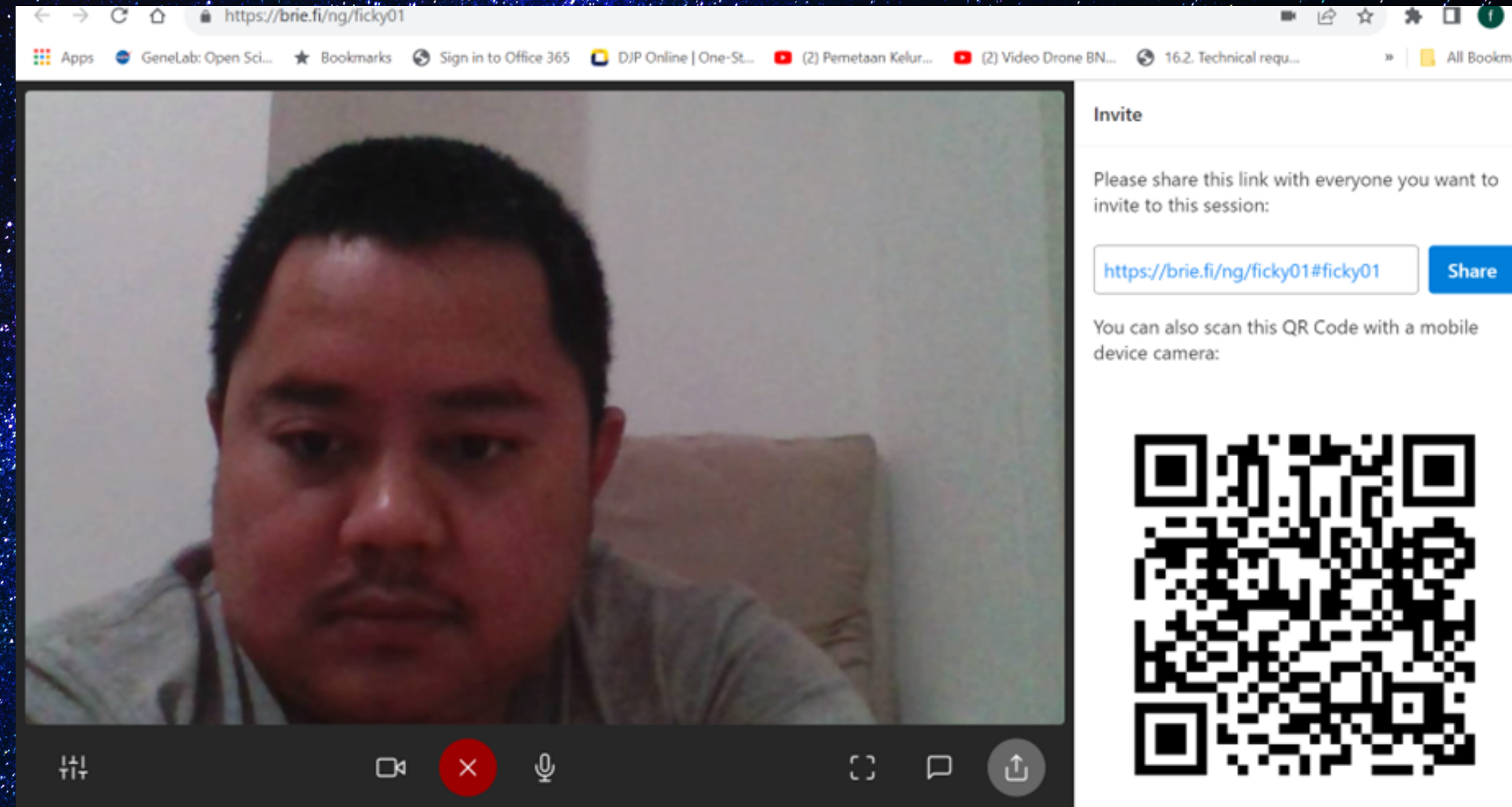
▶

COMMUNICATION TOOLS BETWEEN CREATORS & CONTRIBUTORS (FEATURE NO.3)

Last feature, we provide communication tools for chat and video call between project creators and contributors.

This is the Link of App for Feature 03:
Just Click

<https://brie.fi/ng/ficky01#ficky01>



The screenshot displays a web browser window with the URL <https://brie.fi/ng/ficky01>. The main content area shows a video call in progress, featuring a man's face. Below the video is a control bar with icons for full screen, video off, a red 'X' for ending the call, microphone off, chat, and share. To the right of the video is an 'Invite' section with the text 'Please share this link with everyone you want to invite to this session:' followed by a text input field containing the link <https://brie.fi/ng/ficky01#ficky01> and a 'Share' button. Below this, it says 'You can also scan this QR Code with a mobile device camera:' and displays a large QR code.

DEMO FEATURE NO.3 (PLEASE CLICK BELOW)



Invite



Please share this link with everyone you want to invite to this session:

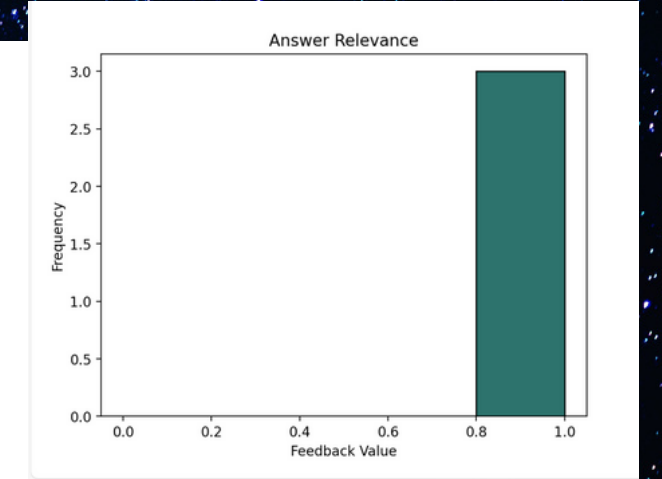
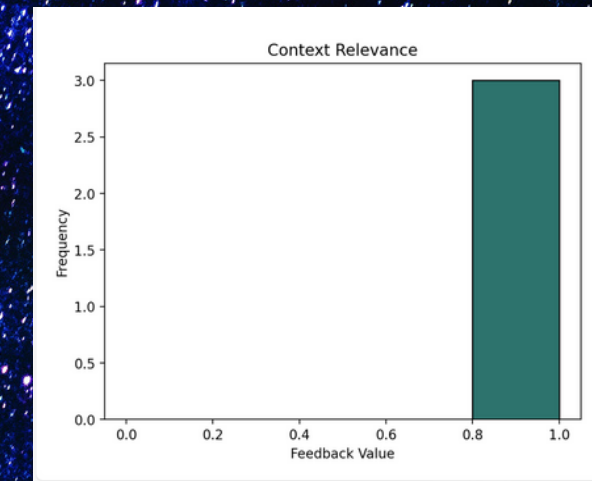
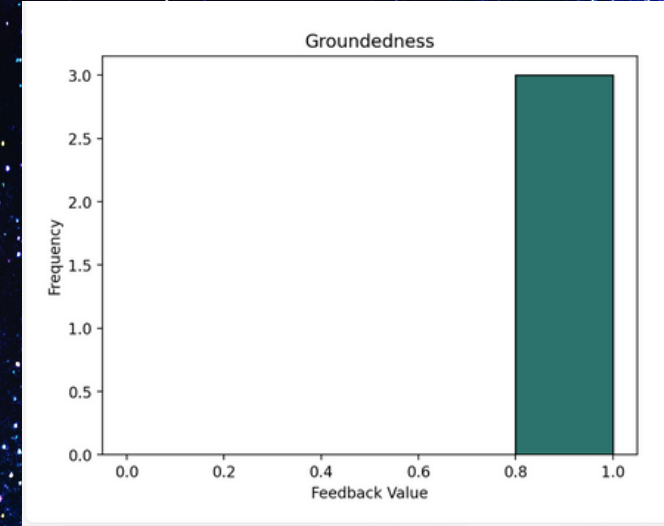
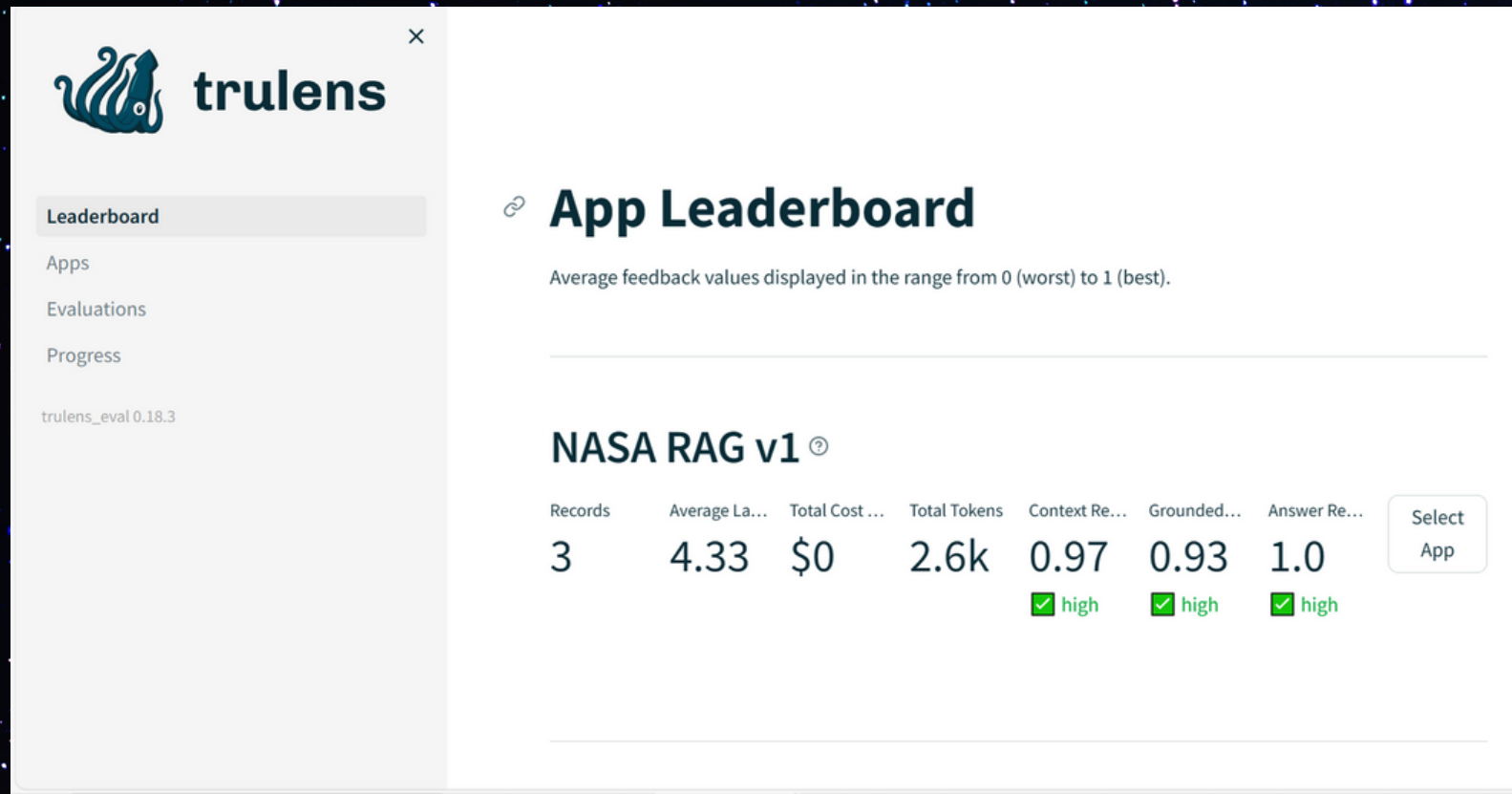
<https://brie.fi/ng/ficky01#ficky01>

Share

You can also scan this QR Code with a mobile device camera:

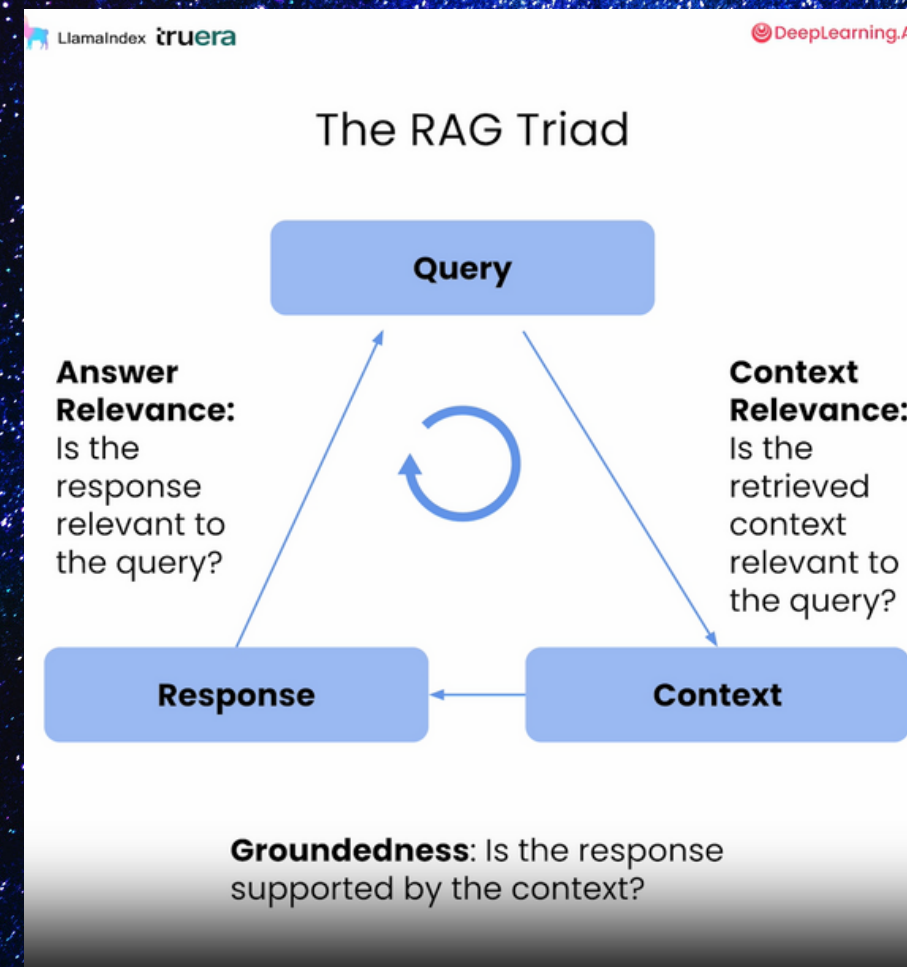


IMPLEMENTING GOOGLE VERTEX AI & TRULENS (1)



We already evaluated the RAG with TruLens. We got the several results with regard to The RAG Triad:

Answer: Relevance: 1.0
Context Relevance: 0.97
Groundedness: 0.93



Feedback Metadata

Context Relevance = 0.9

Statement	result	reason
There are some open science projects in NASA. Project Number 1: NOAA NWS SKYW	0.9	Criteria: The statement

Groundedness = 0.8

source	statement
0	There are some open science projects in NASA. Project Number 1: NOAA NWS SKYW Based on the provided

Answer Relevance = 1

Statement	result	reason
nce projects available in f	1.0	Criteria: The response provides relevant information about open science pro

THE TECHNOLOGIES THAT WE HAVE USED (1)



The screenshot shows the Vertex AI chat interface. On the left is a navigation menu with sections for TOOLS (Dashboard, Model Garden, Pipelines), NOTEBOOKS (Colab Enterprise, Workbench), and GENERATIVE AI STUDIO (Overview, Language, Vision, Speech, Marketplace). The main area is titled 'Untitled prompt' and shows a conversation. The user asks: 'Please elaborate Which US government agency that sponsoring Florida Keys Water Watch project!'. The AI responds: 'The Florida Keys Water Watch project is sponsored by the U.S. Environmental Protection Agency (EPA). The EPA is a federal agency responsible for protecting human health and the environment. It works to reduce pollution, enforce environmental regulations, and promote sustainable practices. The EPA's involvement in the Florida Keys Water Watch project aligns with its mission to protect water resources and ensure the quality of our environment. By supporting this program, the EPA empowers communities to actively participate in monitoring and improving water quality in the Florida Keys region.' On the right, there are controls for Region (us-central1 (Iowa)), Model (chat-bison (latest)), Temperature (0.8), and Token limit (1024). There are also buttons for 'SAVE', 'VIEW CODE', 'RESET PARAMETERS', and 'REPORT INAPPROPRIATE RESPONSES'.

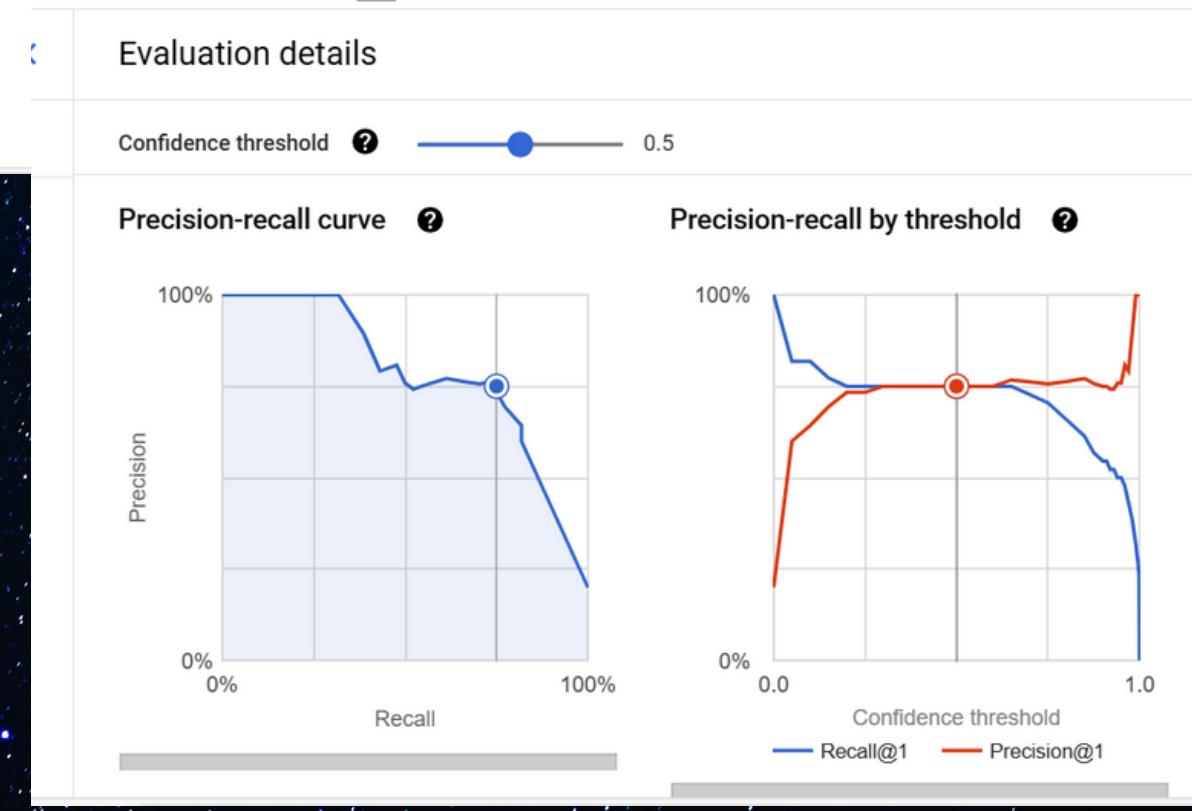
The screenshot shows the Vertex AI evaluation dashboard for a dataset named 'nasa_open_science_project_01'. The interface includes a search bar, navigation menu, and tabs for 'EVALUATE', 'DEPLOY & TEST', 'BATCH PREDICT', and 'VERSION DETAILS'. The 'EVALUATE' tab is active, showing a table of labels and their corresponding scores. The 'Evaluation details' section on the right provides a summary of performance metrics.

Labels	Score
All labels	0.792
complete	0.802
active	0.747
hiatus	0.531
pending	0.518
active_but_seasonal	0.443

Metric	Value
Confidence threshold	0.5
Average precision	0.792
Precision	75%
Recall	75%
F1 score	0.75
Created	Dec 8, 2023, 11:42:00 PM
Total items	428
Training items	340
Validation items	44
Test items	44

We have used several features on Vertex AI :

1. Generative AI Studio - to create prompting & explore Bison Model
2. Model Development - to establish Training dataset & Test Dataset (Vertex AI AutoML)
3. Workbench & Google Colab - to run Notebooks for coding



THE TECHNOLOGIES THAT WE HAVE USED (2)



We have used OpenAI to create Embedding & Vector Store



We have used Trulens to evaluate our LLM Model in RAG Application

```
# Define a groundedness feedback function
f_groundedness = (
    Feedback(grounded.groundedness_measure_with_cot_reasons, name = "Groundedness")
    .on(Select.RecordCalls.retrieve.rets.collect())
    .on_output()
    .aggregate(grounded.grounded_statements_aggregator)
)

# Question/answer relevance between overall question and answer.
f_qa_relevance = (
    Feedback(fopenai.relevance_with_cot_reasons, name = "Answer Relevance")
    .on(Select.RecordCalls.retrieve.args.query)
    .on_output()
)

# Question/statement relevance between question and each context chunk.
f_context_relevance = (
    Feedback(fopenai.qs_relevance_with_cot_reasons, name = "Context Relevance")
    .on(Select.RecordCalls.retrieve.args.query)
    .on(Select.RecordCalls.retrieve.rets.collect())
    .aggregate(np.mean)
)

In Groundedness, input source will be set to __record__.app.retrieve.rets.collect() .
In Groundedness, input statement will be set to __record__.main_output or `Select.RecordOutput` .
In Answer Relevance, input prompt will be set to __record__.app.retrieve.args.query .
```

```
import chromadb
from chromadb.utils.embedding_functions import OpenAIEmbeddingFunction

embedding_function = OpenAIEmbeddingFunction(api_key=os.environ.get('OPENAI_API_KEY'),
                                             model_name="text-embedding-ada-002")

chroma_client = chromadb.Client()
vector_store = chroma_client.get_or_create_collection(name="NASA",
                                                    embedding_function=embedding_function)

vector_store.add("open_science_project", documents=nasa_open_science_info)

from trulens_eval import Tru
from trulens_eval.tru_custom_app import instrument
tru = Tru()

Tru initialized with db url sqlite:///default.sqlite .
```

```
Gaman_NASA_project_Trulens_VertexAi.ipynb
File Edit See Insert Runtime Feature Help Last edited on December 8

+ Code + Text

In Context Relevance, input question will be set to __record__.app.retrieve.args.query .
In Context Relevance, input statement will be set to __record__.app.retrieve.rets.collect() .

[ ] from trulens_eval import TruCustomApp
tru_rag = TruCustomApp(rag,
    app_id = 'NASA RAG v1',
    feedbacks = [f_groundedness, f_qa_relevance, f_context_relevance])

[ ] with tru_rag as recording:
rag.query("Please elaborate which open science projects available in NASA recently!")

boto3,botocore is/are required for using BedrockEndpoint. You should be able to install it/them with
pip install boto3 botocore
WARNING:chromadb.segment.impl.vector.local_hnsw:Number of requested results 2 is greater than number of elements i

[ ] tru.get_leaderboard(app_ids=["NASA RAG v1"])

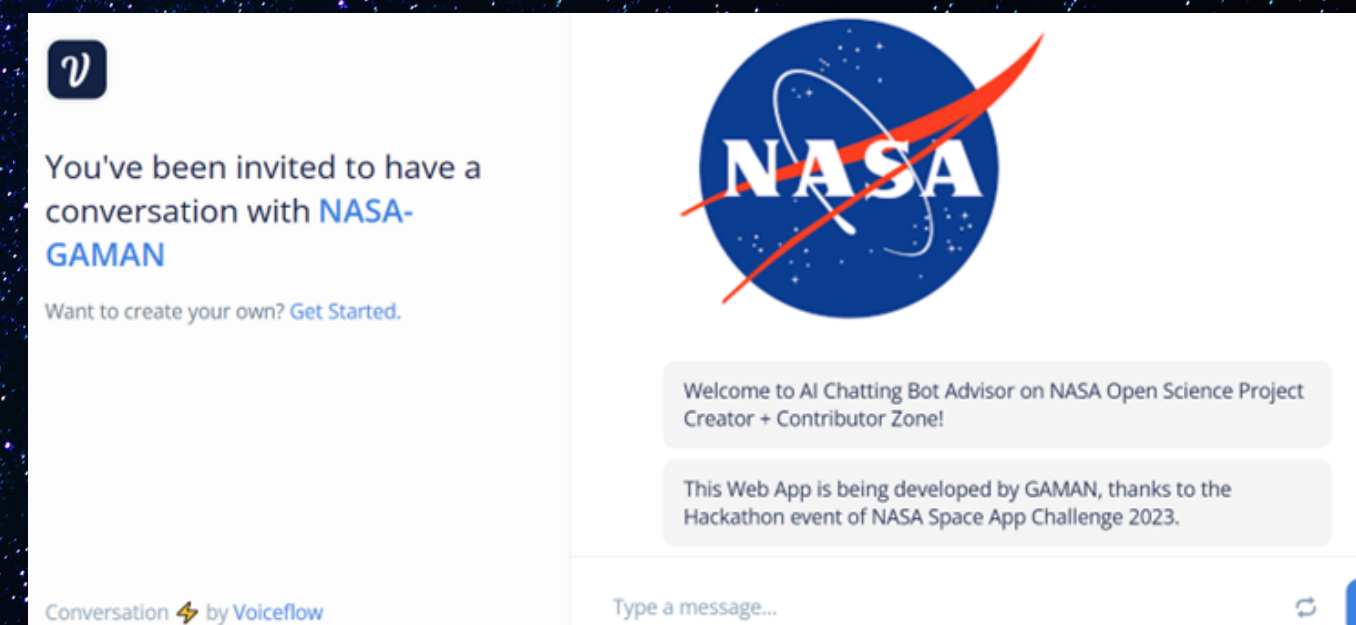
Answer Relevance Groundedness Context Relevance latency total_cost
app_id
NASA RAG v1 1.0 0.8 0.9 8.0 0.001693

with tru_rag as recording:
rag.query("Please elaborate Which US government agency that sponsoring Florida Keys Water Watch project!")

WARNING:chromadb.segment.impl.vector.local_hnsw:Number of requested results 2 is greater than number of elements i
```



We have used Voiceflow to build AI ChatBot Page



This is How Project Creators and Contributors Can Use Gaman Apps to Participate in Open Science Projects

Let say, John, a scientist, who have PhD in ocean marine, is looking for a job. He opens Gaman NASA Recommendation Dashboard



Project	Description	Location	keywords	email
Florida Keys Water Watch	The University of Florida IFAS Monroe County Extension developed Florida Keys Water Watch, a community-based volunteer water quality monitoring program to promote awareness of the importance of water quality, reduce nonpoint source pollution, and involve students and citizens to monitor coastal habitats from man-made canals to bays to beaches. Florida Keys Water Watch teaches volunteers to test a site of their choice for dissolved oxygen, pH, salinity and temperature after attending a 4-hour workshop.	Florida	water quality, dissolved oxygen, pH, salinity, Florida Keys, South Florida, coastal	shellykrueger@ufl.edu

GAMAN NASA Recommendation Dashboard



After searching for a project, John found out that a NASA project which very relatable to his PhD background was Florida Water Watch Project



John emailed Jay. Within couple hours, Jay replied the email and both of them decided to communicate through Gaman NASA Video Call Platform. With this communication tool, John and Jay had a very productive conversation. Since John background match with the skill requirements needed in Florida Water Watch project, he got hired by Jay.

After that, John clicked Gaman NASA AI Chatbot. This AI ChatBot guided John into Jay, the recruiter in Florida Water Watch Project. John got the Jay email address from the ChatBot.

Invite

Please share this link with everyone you want to invite to this session:

<https://brie.fi/ng/ficky01#ficky01> Share

You can also scan this QR Code with a mobile device camera:

GAMAN NASA Video Call Platform

You've been invited to have a conversation with NASA-GAMAN

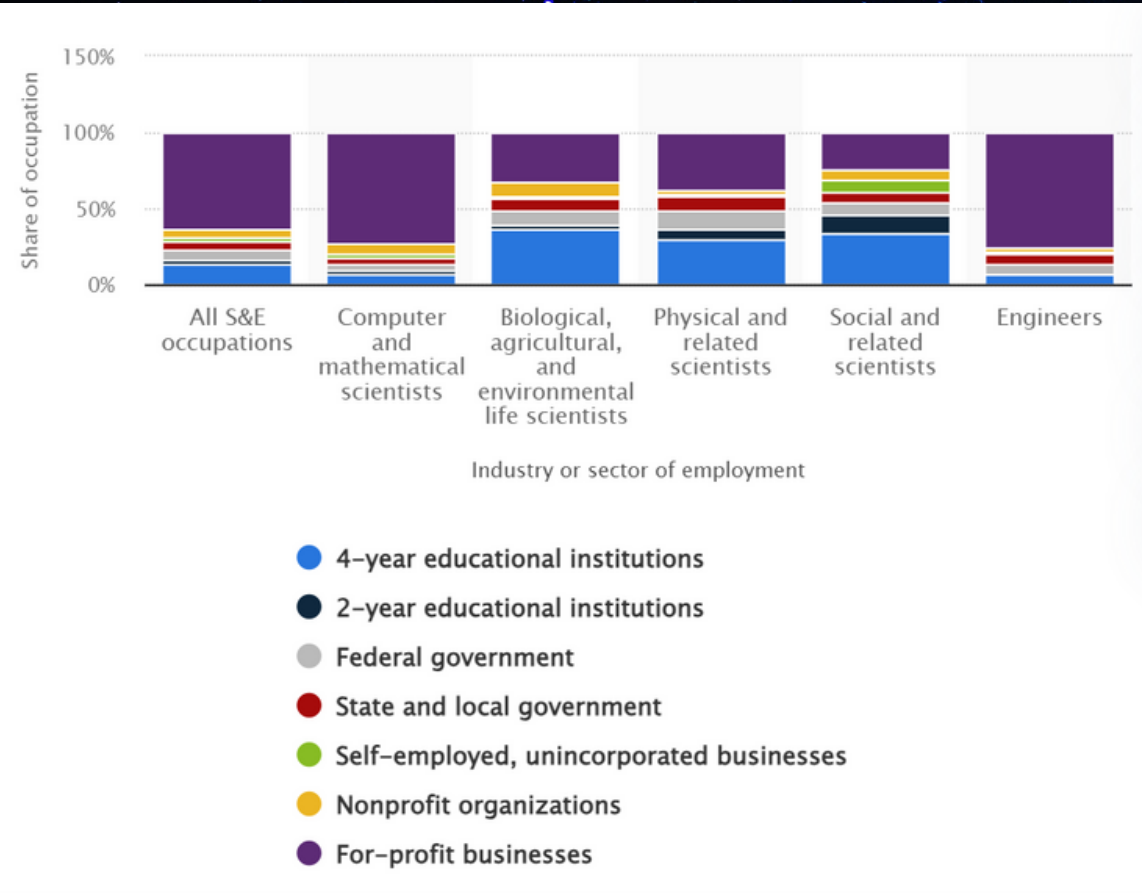
Welcome to AI Chatting Bot Advisor on NASA Open Science Project Creator + Contributor Zone!

This Web App is being developed by GAMAN, thanks to the Hackathon event of NASA Space App Challenge 2023.

GAMAN NASA Chatting Bot



MARKET SCOPE



All S&E occupations	
• 4-year educational institutions	13%
• 2-year educational institutions	3.2%
• Federal government	6.1%
• State and local government	5.5%
• Self-employed, unincorporated businesses	3.1%
• Nonprofit organizations	5.3%
• For-profit businesses	63.8%

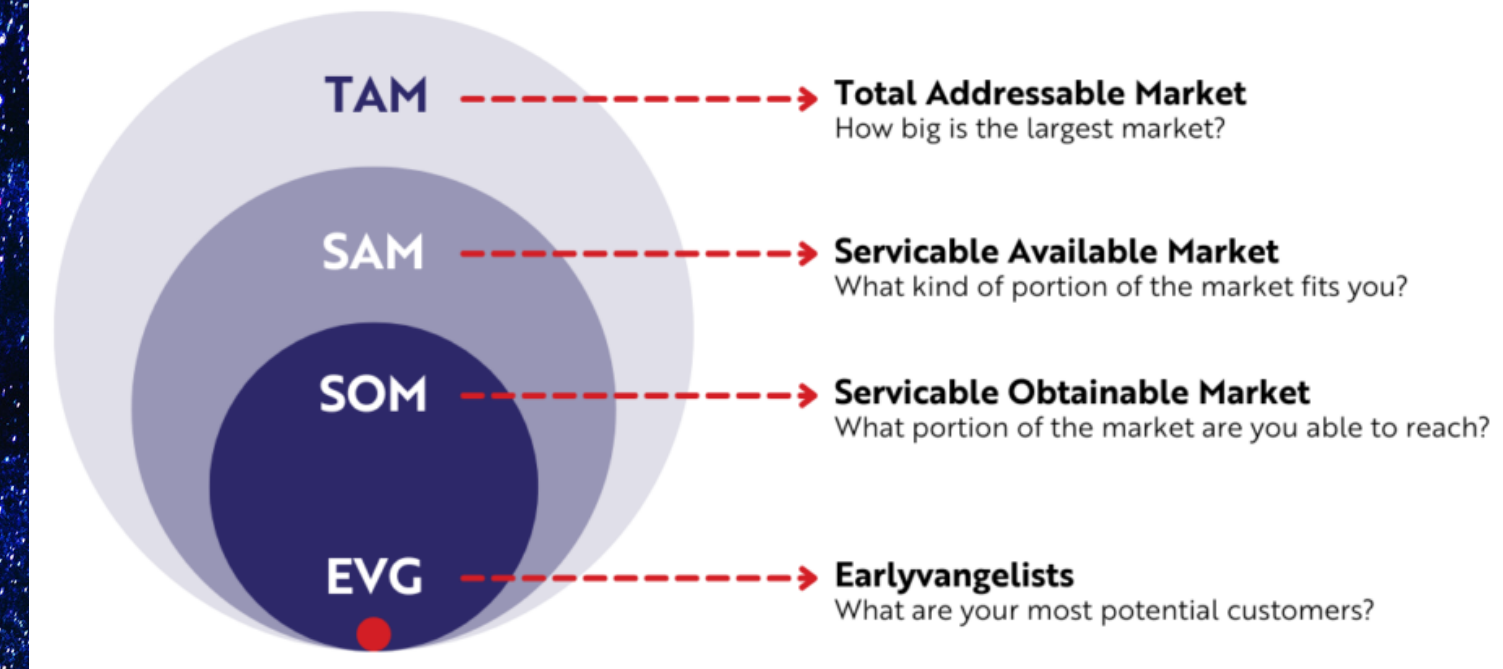
Source: Statista

TAM

Approx. 43,000
Scientists in USA
(Source: Zippia 2021)

SAM

Let say, we will be targeting scientists or researchers which are self-employed, unincorporated businesses, and are still in 2-4 year educational institutions. Thus it will be approx. $(19\% * 43,000 = 8,170)$ scientists & researchers can be our first target market.



SOM

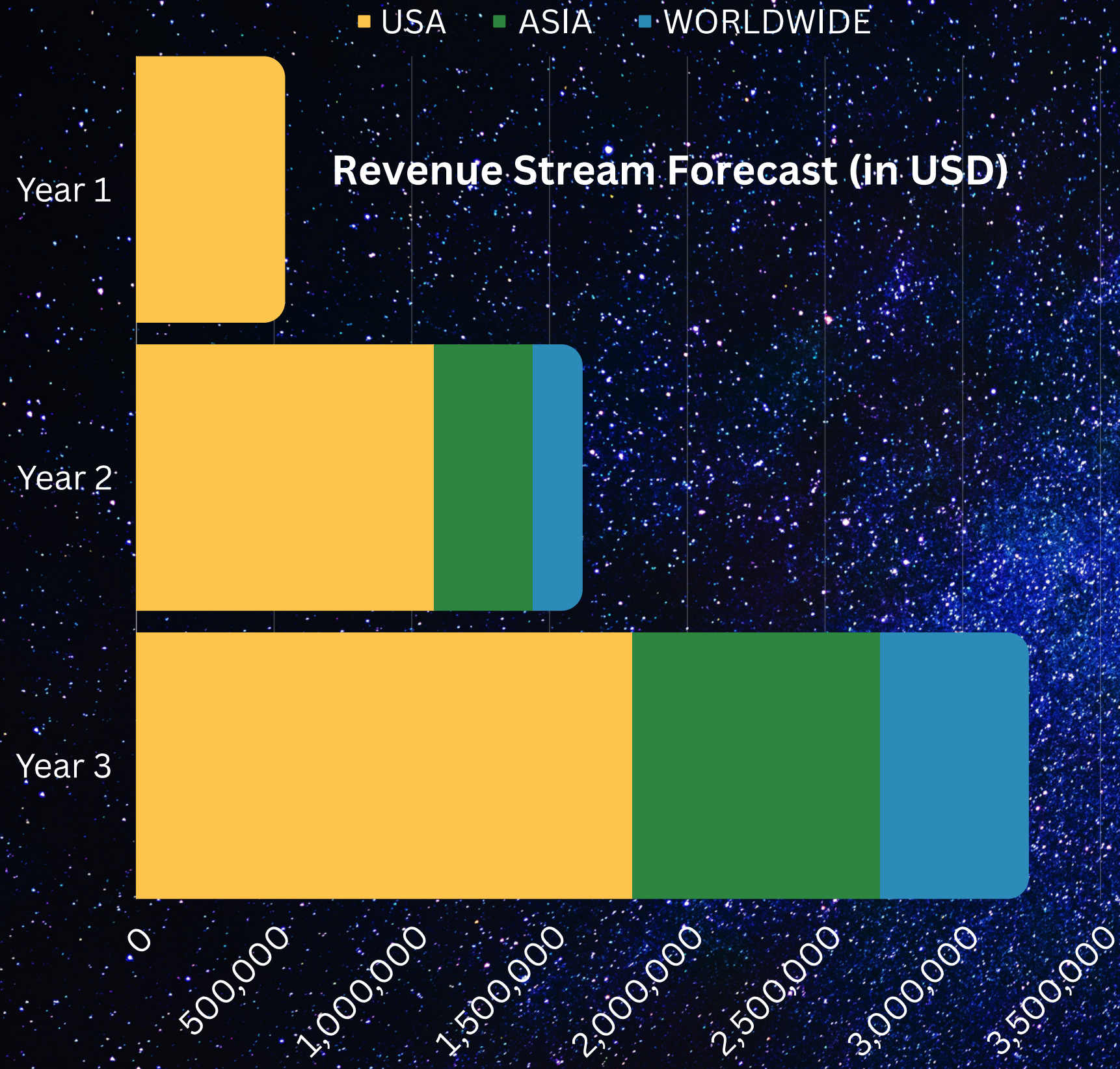
We might narrow down our target market to self-employed scientists or researchers who look for projects or jobs

EARLYVANGELISTS

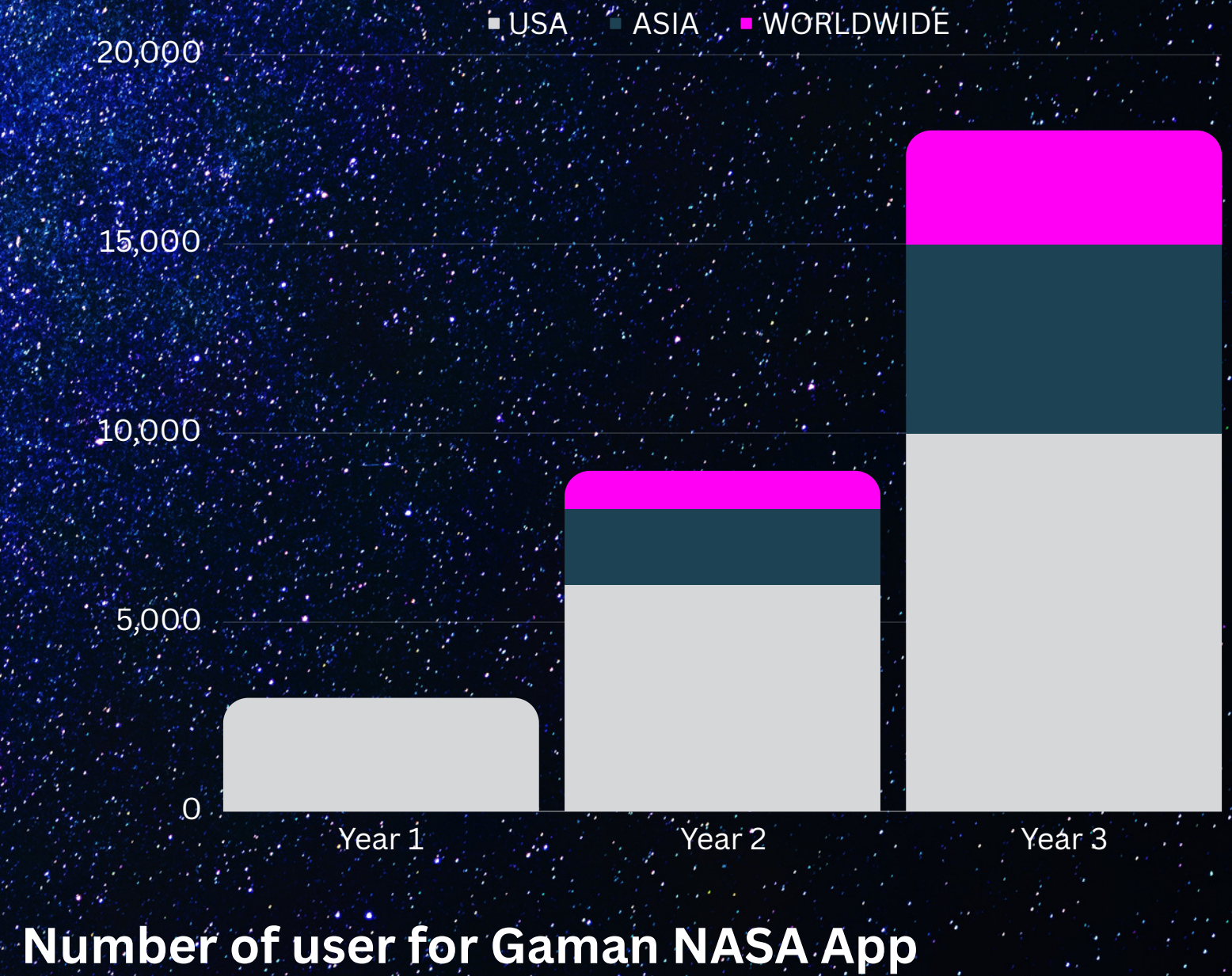
Self-Employed Scientists & Researchers in US

PROJECTED REVENUE STREAM

SUBSCRIBE



We are planning to monetize our app into
SUBSCRIPTION
Price will be approx. **15\$/month**





FICKY ALKARIM

**CTO & COO GAMAN AI
COFOUNDER**

<https://www.linkedin.com/in/ficky-alkarim-a89353a9/>

- MicroMasters Program in Statistics and Data Science in Massachusetts Institute of Technology (MIT), January 2022 - January 2024 (Expected)
- Master of Science in Supply Chain Management, Naveen Jindal School of Management, University of Texas at Dallas, USA, 2013-2015
- Berkeley Summer Sessions, Haas School of Management, University of California Berkeley, USA, 2010
- Bachelor of Management, School of Business and Management, Bandung Institute of Technology (ITB), 2006-2009
- *Certified Machine Learning – Amazon Web Services (AWS)*
- *Front-End & Back-End Development - META*
- *Machine Learning with Python & Linear Models to Deep Learning – MIT*
- *Fundamental of Statistics & The Science of Uncertainty Data –MIT*
- *Data Science: Inference & Modeling - Harvard*
- *Data Science: Probability - Harvard*
- *Data Science: R Basic - Harvard*
- *Blockchain Technology Certificate - University of California Berkeley*
- *Solana Blockchain Developer Program - Solana Labs*
- *Foundations of Project Management Certificate - Google*
- *Developing & Deploying an Internet of Things (IoT) – AWS*
- *Python for Natural Language Processing (NLP) – Udemy*
- *Practical Hands-on Guide for Alexa Skill Development – AWS*
- *Google Voice Assistant Development - Google*



**Thank you
very much!**