

APP-SOLUTELY FABULOUS

A Data Scientist's Guide to Choosing Python Web Tools
Wisely

Alan Feder

WHO AM I?



- Alan Feder
- Staff LLM Data Scientist at Magnifi / TIFIN
- Freelance Consultant
- NYC Expat  → 
- R Expat  → 

DATA SCIENTISTS IN R LOVE SHINY

- How do I show interactive analysis?
- Am I stuck with copying and pasting my graphs to PowerPoint?



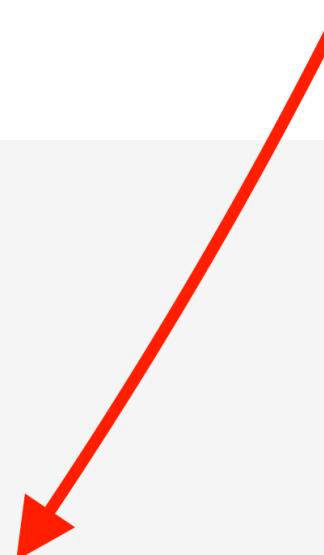
THE ZEN OF PYTHON

Abstract

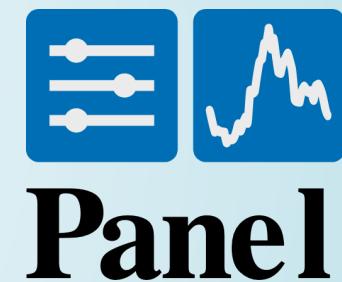
Long time Pythoner Tim Peters succinctly channels the BDFL's guiding principles for Python's design into 20 aphorisms, only 19 of which have been written down.

The Zen of Python

Beautiful is better than ugly.
Explicit is better than implicit.
Simple is better than complex.
Complex is better than complicated.
Flat is better than nested.
Sparse is better than dense.
Readability counts.
Special cases aren't special enough to break the rules.
Although practicality beats purity.
Errors should never pass silently.
Unless explicitly silenced.
In the face of ambiguity, refuse the temptation to guess.
There should be one-- and preferably only one --obvious way to do it.
Although that way may not be obvious at first unless you're Dutch.
Now is better than never.
Although never is often better than *right* now.
If the implementation is hard to explain, it's a bad idea.
If the implementation is easy to explain, it may be a good idea.
Namespaces are one honking great idea -- let's do more of those!



ONLY ONE??



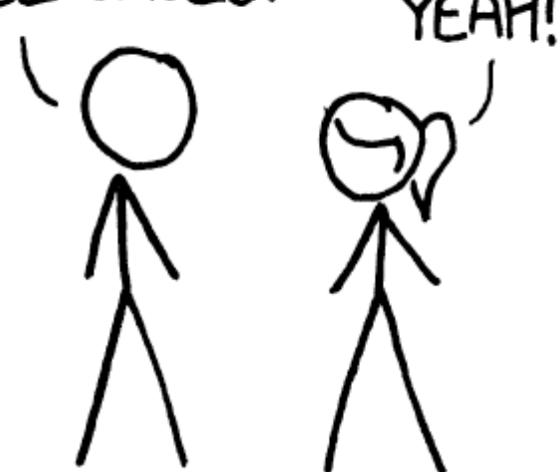
TOO MANY OPTIONS

HOW STANDARDS PROLIFERATE:

(SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC)

SITUATION:
THERE ARE
14 COMPETING
STANDARDS.

14?! RIDICULOUS!
WE NEED TO DEVELOP
ONE UNIVERSAL STANDARD
THAT COVERS EVERYONE'S
USE CASES.



YEAH!

SOON:

SITUATION:
THERE ARE
15 COMPETING
STANDARDS.

TRY(SOME OF) THEM!



- Let's try to make one app with three different methods
- RAG Chatbot referencing a specific data set
 - Previous RGov Talks



SERVER VS. SERVERLESS

- The computations need to happen *somewhere*
- `localhost:8000` - doesn't help with sharing
 - Setting up servers can be difficult without software experience
- In-browser versions can be helpful
 - Not all Python packages work
 - BE CAREFUL ABOUT YOUR PASSWORDS AND API KEYS

STREAMLIT



Streamlit

STREAMLIT

- Launched in 2019
- Purchased by Snowflake in 2022
for \$800m 
- I've been using it a lot for the past 2 years - and even used in a (NY) R Conference talk!



thiago Streamlit Team Member

Oct 2019

The original name was “Streamlet”, but then someone mistyped it as “Streamlit” and we liked it better that way. The “lit” evokes the ideas of something lighting up the path ahead, or bringing clarity, or illuminating a solution, and so on.

Now as to why it was Streamlet to begin with, depends on who you ask 😊

My version of the story is that we *stream* data to your app lightning-fast, so we wanted to have “stream” in the name. I like that version so much that I eventually forgot the real origin story, which at this point only @Adrien_Treille knows!

STREAMLIT EXAMPLE

Streamlit RAG on R/Gov Talks x +

https://rgov-2024-ajf.streamlit.app

Incognito (2)

Fork ⌂ ⌄

Which GPT model do you want to use?

Cheaper

More Accurate

Use Streamlit to Run RAG on the previous R/Gov Talks ↗

Enter your question:

How does the US Government use R?

Submit!

Wendy Martinez discusses the use of R within the U.S. government, particularly at the Bureau of Labor Statistics (BLS). She highlights the journey to get R approved for production use at BLS, emphasizing the challenges faced due to the need for stable and reliable software to produce official statistics. R is used for research and innovative applications, such as creating dynamic maps and automating text generation for publications. Despite initial resistance due to concerns about support and stability, R has been approved for production use at BLS, marking a significant step in integrating open-source software into government processes.

Kimberly Kreiss from the Federal Reserve Board of Governors explains how R is used to automate production processes for the Survey of Household Economics and Decision Making. R is employed to generate tables and charts for reports, replacing manual data entry and reducing human error. The use of R allows for flexibility and efficiency in handling large datasets and producing visualizations, which are essential for the survey's annual report. This automation process is part of a broader effort to improve data handling and presentation within the government using R.

RAG-identified relevant videos

The Rocky Road to Using R at a U.S. Government Agency	The Journey Continues: Using R at a U.S. Government Agency	SHED-ing Light on Survey Data
Wendy Martinez	Wendy Martinez	Kimberly Kreiss
Year: 2020	Year: 2021	Year: 2019
Similarity Score: 57/100	Similarity Score: 53/100	Similarity Score: 49/100



The Rocky Road to Using R at a U.S. Government Agency

Wendy Martinez

Year: 2020

Similarity Score: 57/100

Lander Analytics Presents
The Government & Public Sector R Conference

Speaker
Wendy Martinez

The Journey Continues: Using R at a U.S. Government Agency

Wendy Martinez

Year: 2021

Similarity Score: 53/100

RSTATS.AI #RSTATSNYC

Dr. Wendy Martinez

SHED-ing Light on Survey Data

Kimberly Kreiss

Year: 2019

Similarity Score: 49/100

Kimberly Kreiss - SHED-ing Light on Surv...

STREAMLIT - 😊

- Lots of elements to use
 - New ones all the time
- Very Pythonic
 - Simple to transition from script

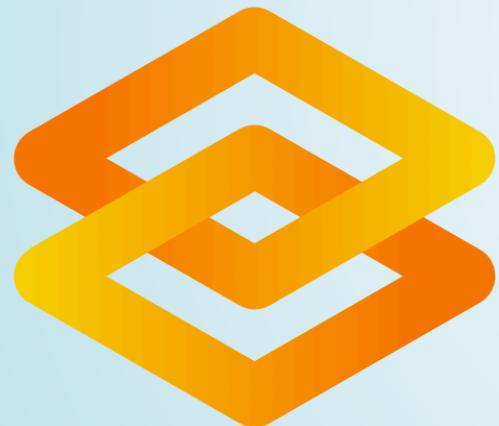
STREAMLIT - 😐

- Runs top to bottom - not at all reactive
- Can get very slow if it gets big
 - Has caching, `session_state`
- All Streamlit apps kinda look the same
 - Hard to tweak UI

STREAMLIT-LITE

- [Stlite](#)
 - Not officially supported by Streamlit
- Pretty easy interface to copy your Streamlit apps
- I couldn't figure out how to do streaming
- Slower to load, slower to run

GRADIO



gradio

GRADIO

- Started by Stanford PhD Students in 2019
- Purchased by Hugging Face  in December 2021
- Powers most Huggingface Spaces

The name “Gradio” comes from a combination of the words “Graphical” and “Audio”, which reflects the library’s original focus on building interfaces for audio and speech processing models. However, the library has since expanded to support a wide range of input and output types, including text, images, and video, and the name “Gradio” has come to represent a more general-purpose interface builder for data processing functions.

GRADIO - SIMPLEST

Spaces | AlanFeder/RGov_Gradio_App_-_Basic_Chat | like 0 | Running | ...

Use Gradio to Run RAG on the previous R/Gov Talks - Chat Interface 1

Chatbot

Type a message...

Send

This Gradio app was created for Alan Feder's talk at the 2024 R/Gov Conference

GRADIO - MORE COMPLEX

RAG on R/Gov Talks

Use Gradio to Run RAG on the previous R/Gov Talks

Model
Choose the model to use
 Cheaper More Accurate

Enter your question:
How does the US Government use R?

Response

Wendy Martinez discusses the use of R within the U.S. government, particularly at the Bureau of Labor Statistics (BLS). She highlights her efforts to get R approved for production use at BLS, emphasizing the importance of using R for producing official statistics. Martinez notes that R is favored for its cost-effectiveness and the ability to implement new methodologies quickly, which are often first available in R before commercial software. She also mentions the challenges faced, such as security concerns, the need for stability, and the preference for commercial support. Despite these challenges, R has been officially approved for production use at BLS, marking a significant achievement in integrating open-source software into government operations.

Kimberly Kreiss from the Federal Reserve Board of Governors explains how R is used to automate production processes for the Survey of Household Economics and Decision Making (SHED). R is utilized to generate tables and charts for the survey report, replacing manual data entry and reducing human error. Kreiss highlights the efficiency and flexibility of using R, which allows for easy updates and changes to the report's visuals. The use of R has significantly improved their workflow, and there are plans to develop a public package to facilitate data access and analysis.

The Rocky Road to Using R at a U.S. Government Agency

Wendy Martinez

Year: 2020

Similarity Score: 57/100



► Transcript

The Journey Continues: Using R at a U.S. Government Agency

Wendy Martinez

Year: 2021

Similarity Score: 53/100



► Transcript

SHED-ing Light on Survey Data

Kimberly Kreiss

Year: 2019

Similarity Score: 49/100



► Transcript

GRADIO - 😊

- Very focused on AI
- Even simpler than Streamlit
- Built in interfaces for Chat
 - Easy to add in features like chat history, Thumbs Up/ Thumbs Down
- Has reactivity

GRADIO - 😐

- Less flexible than Streamlit
 - Complicated to do multiple things at once
- Fewer components available
 - Sometimes need to use HTML to do what you want
- All Gradio apps look the same – at least as much as Streamlit

GRADIO - DEPLOYMENT POSSIBILITIES

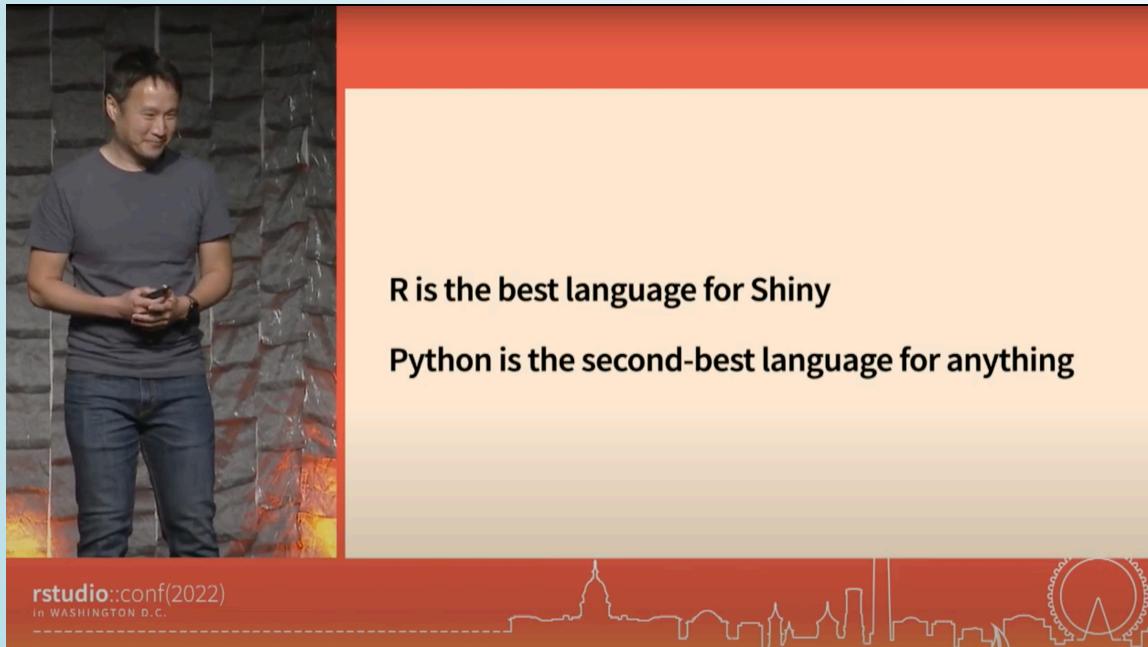
- Local / Self-hosted
- Public Link, local compute (`share=True`)
- Through Huggingface Spaces (free with basic CPU)
- In-browser (Gradio Lite)
 - I couldn't figure it out for complex app

SHINY



SHINY

- 2012: Developed for R by the company formerly known as RStudio 
- 2022: 



TWO TYPES OF SHINY

SHINY CORE

- More similar to RShiny
- Split Server vs. UI sections
 - Only one file
- Nested Functions

SHINY EXPRESS

- Kinda Streamlit-ish
- No Split between Server and UI sections
- Heavy use of context managers

```
1 with ui.row():  
2     xxxx
```

-
- Use of decorators

```
1 @reactive.text  
2 def function():  
3     ...
```

SHINY EXAMPLE

Use Shiny to Run RAG on the previous R/Gov Talks

What question do you want to ask?

What is the tidyverse?

Submit!

SHINY - 😊

- All Shiny is reactive - great for complex Apps
- Very flexible, can make almost any change you want
- Lots of components
- Documentation gives really good background/philosophy on how it works

SHINY - 😐

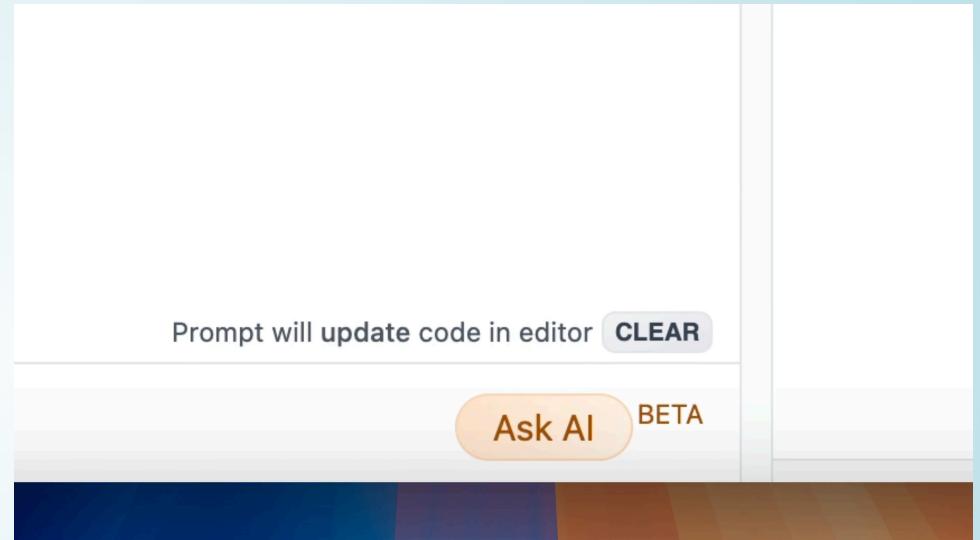
- I found it confusing, not as intuitive/"Pythonic"
- Default UI is pretty bland
- Not as many components as Streamlit
- You kinda need to learn the philosophy of Shiny before you can do much complicated

HELP IS ON THE WAY!

SHINY



GRADIO



SHINYLIVE

SHINYLIVE IS SO EASY

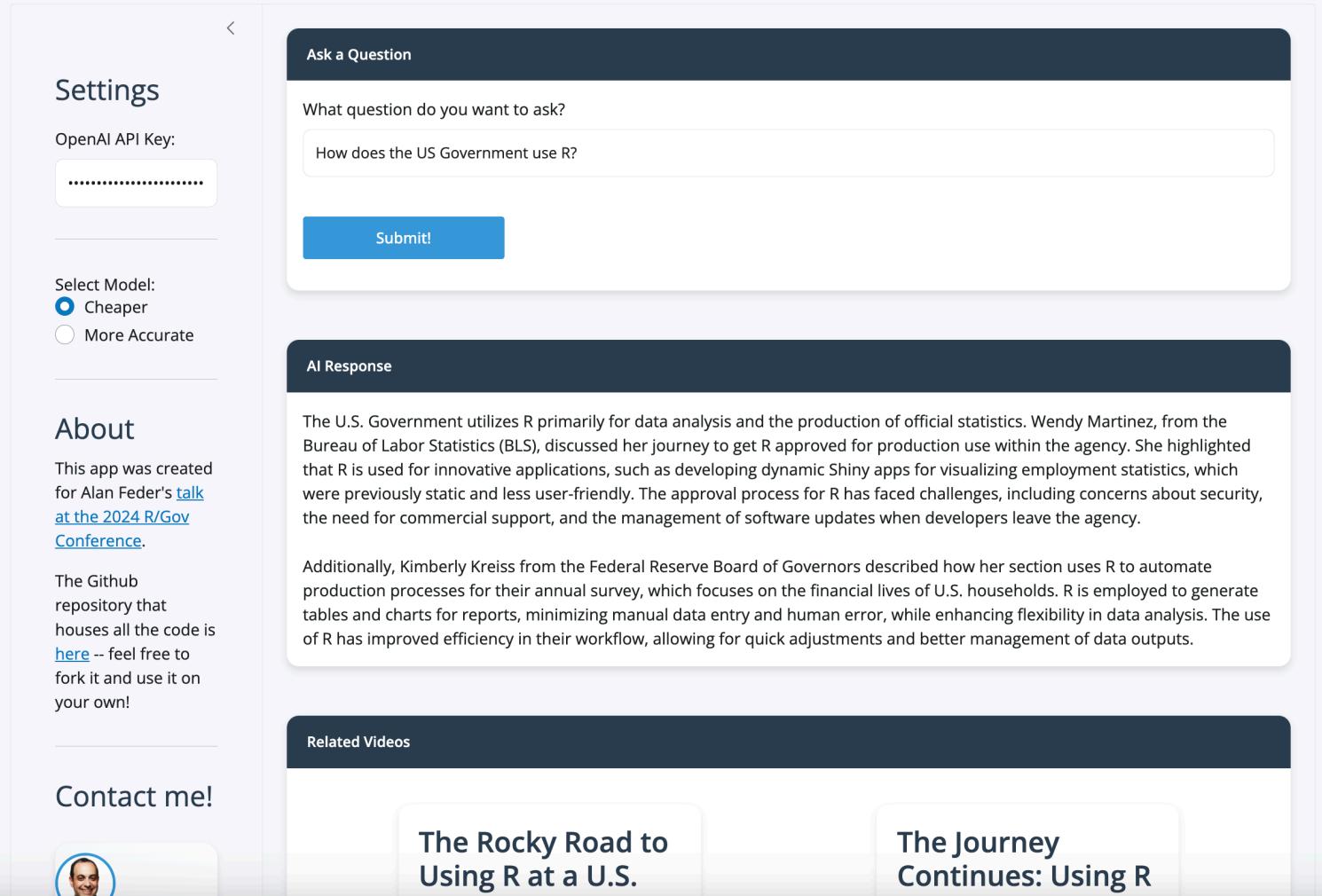
shinylive is an extension to the Quarto open-source scientific and technical publishing system. It enables shiny applications to run locally, without a shiny server using WebAssembly. I'll show examples and discuss the limitations of using shinylive.

[Slides](#)



SHINYLIVE REALLY IS EASY!

Use Shiny to Run RAG on the previous R/Gov Talks
[RAG on R/Gov Talks](#)



The screenshot displays a Shiny application interface. On the left, there's a sidebar with 'Settings' (OpenAI API Key, Select Model - Cheaper selected), 'About' (app created for Alan Feder's talk at the 2024 R/Gov Conference, GitHub repository link), and 'Contact me!' (with a small profile picture). The main area has three sections: 'Ask a Question' (a form with a question input field containing "How does the US Government use R?" and a 'Submit!' button), 'AI Response' (a text block about R usage in government statistics), and 'Related Videos' (two video thumbnails: 'The Rocky Road to Using R at a U.S.' and 'The Journey Continues: Using R').

Ask a Question

What question do you want to ask?

How does the US Government use R?

Submit!

AI Response

The U.S. Government utilizes R primarily for data analysis and the production of official statistics. Wendy Martinez, from the Bureau of Labor Statistics (BLS), discussed her journey to get R approved for production use within the agency. She highlighted that R is used for innovative applications, such as developing dynamic Shiny apps for visualizing employment statistics, which were previously static and less user-friendly. The approval process for R has faced challenges, including concerns about security, the need for commercial support, and the management of software updates when developers leave the agency.

Additionally, Kimberly Kreiss from the Federal Reserve Board of Governors described how her section uses R to automate production processes for their annual survey, which focuses on the financial lives of U.S. households. R is employed to generate tables and charts for reports, minimizing manual data entry and human error, while enhancing flexibility in data analysis. The use of R has improved efficiency in their workflow, allowing for quick adjustments and better management of data outputs.

Related Videos

The Rocky Road to Using R at a U.S.

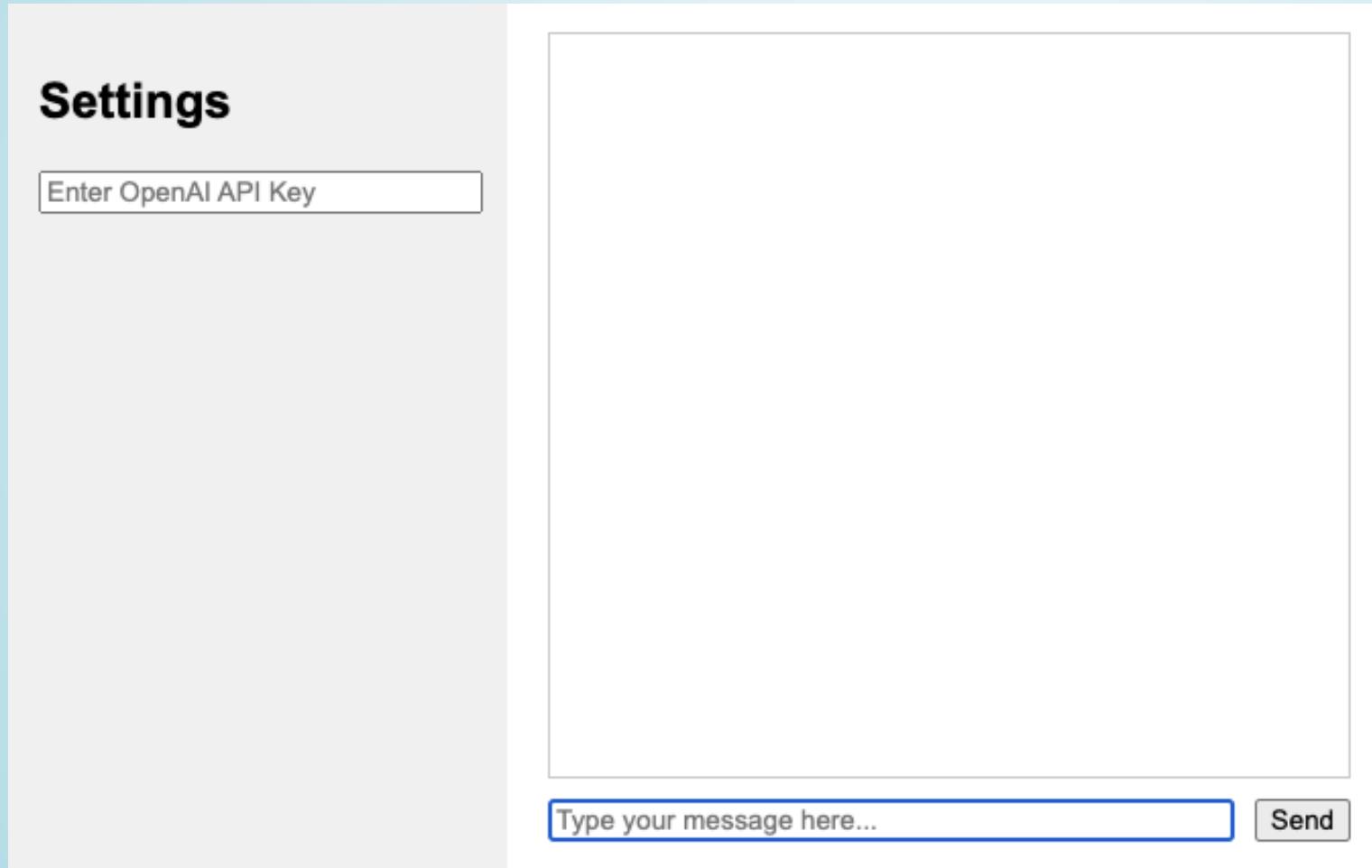
The Journey Continues: Using R

ONE OTHER OPTION...

If you're going to go in-browser, and what you're doing isn't that complex...

... AI might be able to build it for you without any of these apps

ASK CLAUDE/CHATGPT TO BUILD YOU AN IN-BROWSER TOOL



IT WILL THEN IMPROVE IT NICELY

I've got an HTML based RAG app. It works entirely in-browser successfully. Can you make it prettier?

great! now make it even prettier

Settings

Select model:

Cheaper
 More Accurate

Contact me!


[Email](#)
[Website](#)
[LinkedIn](#)
[GitHub](#)

LLM Generated RAG - Entirely HTML/JS in-browser

How does the US Government use R?

The U.S. Government, particularly the Bureau of Labor Statistics (BLS), uses R primarily for research and is currently in the process of getting it approved for production purposes. Wendy Martinez, a speaker from the BLS, outlined her journey in advocating for R within the agency, emphasizing its potential for producing innovative applications and approaches to official statistics, such as the monthly employment situation and consumer price index. Despite initial approvals for research, R faced considerable

Type your message here... Send

Retrieved Videos:

Video 1: The Rocky Road to Using R at a U.S. Government Agency

Video 2: The Journey Continues: Using R at a U.S. Government Agency

About

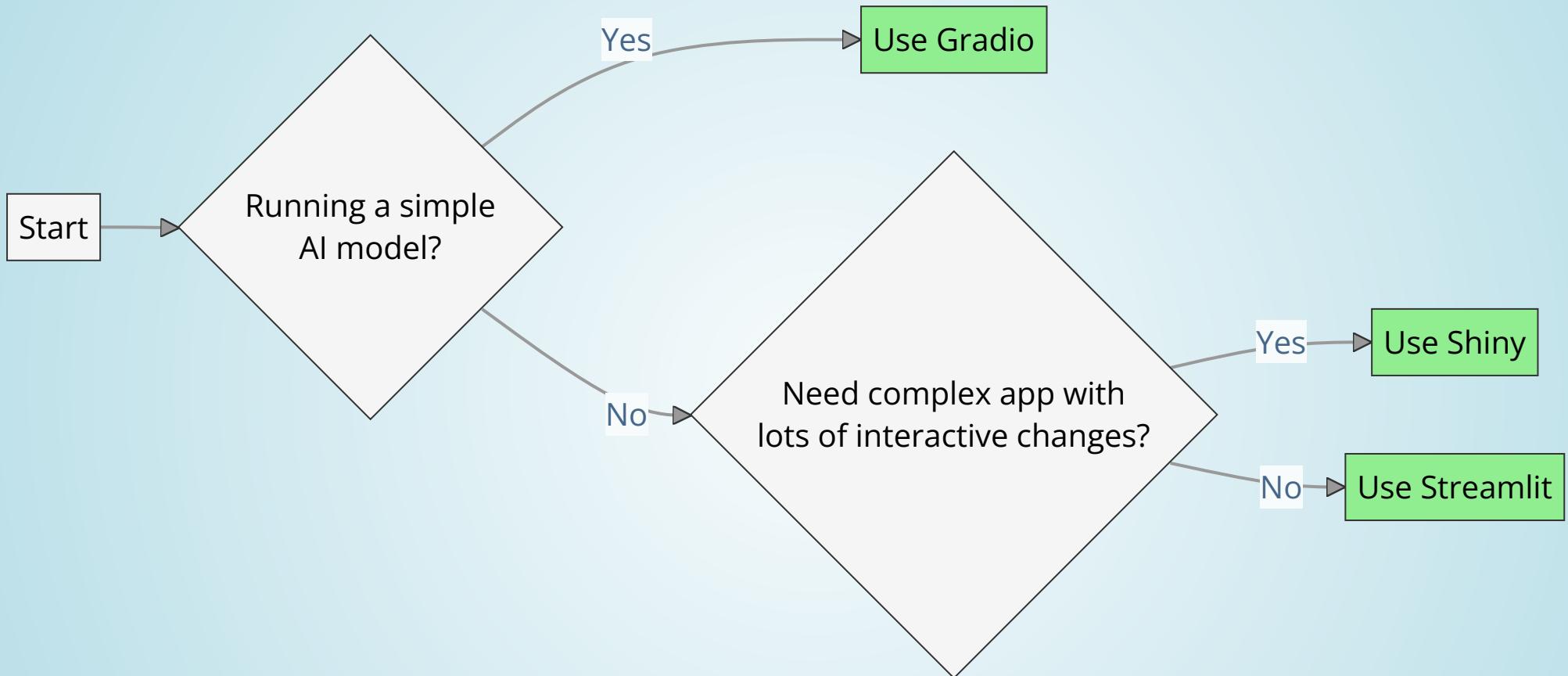
This app was created for Alan Feder's [talk at the 2024 R/Gov Conference](#).

The Github repository that houses all the code is [here](#) -- feel free to fork it and use it on your own!

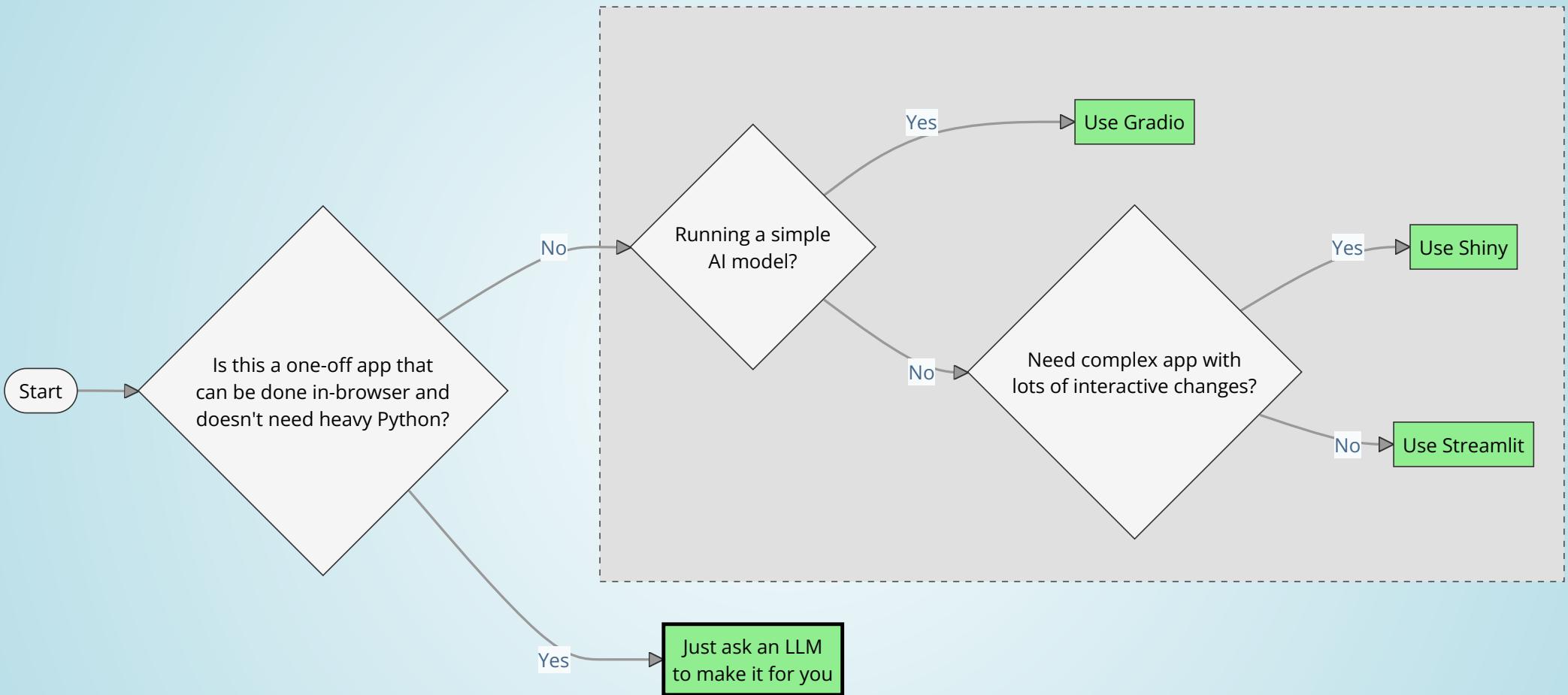
SO WHAT SHOULD I DO?

If you know one tool well, you can use it for most things
- they overlap a lot.

REALLY, WHAT SHOULD I DO?



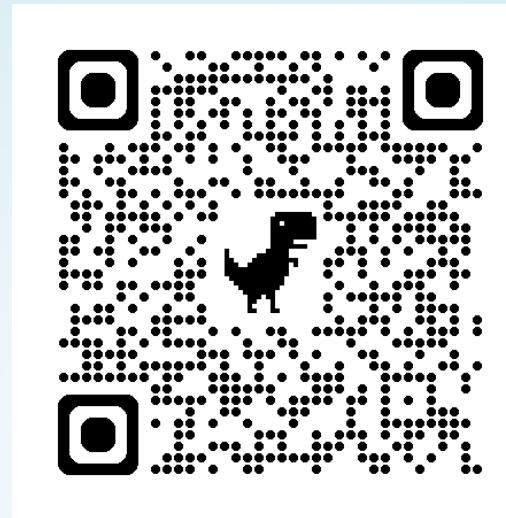
BUT...



ALAN FEDER



- Code [AlanFeder/rgov-2024](https://github.com/AlanFeder/rgov-2024)
- Slides https://www.alanfeder.com/dcr_rag/



✉️ AlanFeder@gmail.com
🌐 <https://www.alanfeder.com>
linkedin [alanfeder](https://www.linkedin.com/in/alanfeder)

Tool	Basic	Complex	Lite
Streamlit	Link	Link	Link
Gradio	Link	Link	
Shiny	Link	Link	Link
LLM/HTML			Link