

Single Day Event

Activity Report of PHUSE Japan Open Source Technologies Working Group
2023/12/08
Open-source Technology Working Group

Topics

- About Open Source Technology Working Group (OST WG)
- OST WG Activities 2023
 - Summary of PHUSE R Workshop
 - Presentation at DIA Japan 2023
 - Learn and use a generative AI tool (ChatRWKV) to increase our productivity



Disclaimer

The speaker is a paid employee of Astellas. This presentation is intended for informational purposes only and does not replace independent professional judgment. This presentation is not intended to provide medical or legal advice. Statements of fact, positions taken and opinions expressed are those of the speaker individually and, unless expressly stated to the contrary, do not necessarily reflect the opinion or position of the speaker's employer, Astellas, or any of its subsidiaries and/or related entities.

Topics

- About Open Source Technology Working Group (OST WG)
- OST WG Activities 2023
 - Summary of PHUSE R Workshop
 - Presentation at DIA Japan 2023
 - Learn and use a generative AI tool (ChatRWKV) to increase our productivity

About OST WG

Goal:

To broaden the acceptance and general level of comfort with open source technologies in our industry to assist in increasing their level of adoption in Japan.

About OST WG

Members:

- Ippei Akiya, A2 Healthcare Corporation
- Kazuyuki Miyashita, Boehringer Ingelheim Co., Ltd.
- Masashi Aonuma, Astellas Pharma Inc.
- Masashi Tanaka, EPS Corporation
- Masataka Higashijima, Eli Lilly and Company
- Naoko Izumi, Novartis Pharma K.K.
- Tomoki Nishikawa, Kyowa Kirin Co.,Ltd.
- Tomoyuki Namai, Chugai Pharmaceutical Co., Ltd.
- Yasutaka Moriguchi, GlaxoSmithKline K.K.
- Yuichi Nakajima, Novartis Pharma K.K.
- Yutaka Morioka, EPS Corporation

Topics

- About Open Source Technology Working Group (OST WG)
- **OST WG Activities 2023**
 - Summary of PHUSE R Workshop
 - Presentation at DIA Japan 2023
 - Learn and use a generative AI tool (ChatRWKV) to increase our productivity



Summary of PHUSE R Workshop

- 2023-11-24 14:00-17:30 hosted by Novartis.
- Number of participants: 13

時間	内容
14:00 – 14:15	はじめに • PHUSEとは • Ground rule
14:15 – 14:45	Posit Cloudの使い方について
14:45 – 15:15	• R言語の基礎
15:15 – 15:30	休憩
15:30 – 17:30	• 基本的なデータハンドリング • Hands-on Exercise: ADSLを作成しよう



Summary of PHUSE R Workshop





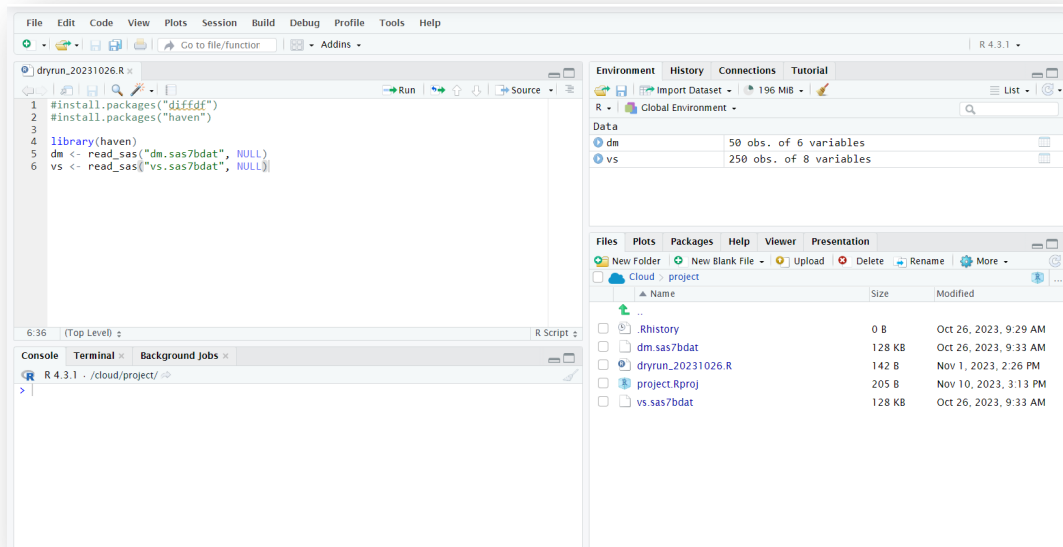
Summary of PHUSE R Workshop

- Programming environment: Posit cloud (R 4.3.1)



New Name,
Same DNA

RStudio is now Posit.



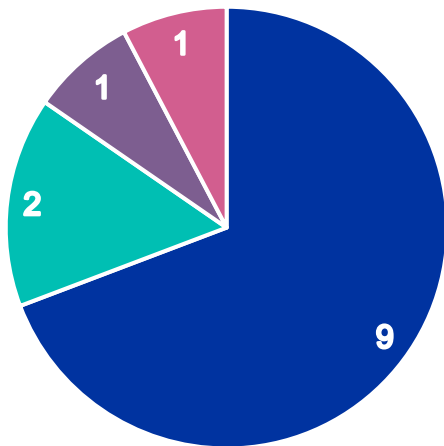


Summary of PHUSE R Workshop

- Contents summary
 - R basics / Differences between SAS and R
 - R data handling
 - Import SAS datasets with “haven”
 - Data manipulation with “tidyverse”
 - Hands-on
 - Create ADSL from SDTM datasets (DM / VS)

Survey results – Background

Attendee's role
N = 13



■ Programmer ■ Statistician ■ Data Manager ■ Other

	N = 13 n (%)
Programming experience (years)	
< 1	1 (7.7)
1 <= < 5	4 (30.8)
>= 5	8 (61.5)
Reason for participation (Multiple)	
- Business use	6 (46.2)
- Learning	9 (69.2)
- Other	2 (15.4)
Do you have an opportunity to use R ?	
- No	8 (61.5)
- Yes	5 (38.5)

Survey results – Workshop

4.62

Understanding

2.23

Difficulties

3.62

Length of workshop

Survey results – Future plan

- What sort of hands-on do you expect in future?
 - R packages: tidyverse / dplyr / pharmaverse (admiral, ...)
 - R TFL creation
 - R markdown
 - Analysis Results Standard
 - dataset.json



Topics

- About Open Source Technology Working Group (OST WG)
- **OST WG Activities 2023**
 - Summary of PHUSE R Workshop
 - **Presentation at DIA Japan 2023**
 - Learn and use a generative AI tool (ChatRWKV) to increase our productivity

- “Effective R Submission Readiness: Sponsor’s Point of View”.
 - Session 11: CHALLENGES AND TRANSFORMING INTERACTIVE COMMUNICATION IN DRUG DEVELOPMENT USING OPEN SOURCE SOFTWARE

Acknowledgement

PHUSE Japan OST Working Group

Masashi Aonuma (Astellas Pharma Inc.)

Ippei Akiya (A2 Healthcare Corporation)

Naoko Izumi (Novartis Japan)

Masashi Tanaka (EPS Corporation)

Tomoyuki Namai (Chugai Pharmaceutical Co., Ltd.)

Tomoki Nishikawa (Kyowa Kirin Co., Ltd.)

Masataka Higashijima (Eli Lilly Japan K.K.)

Kazuyuki Miyashita (Nippon Boehringer Ingelheim Co., Ltd.)

Yutaka Morioka (EPS Corporation)

Yasutaka Moriguchi (GlaxoSmithKline K.K.)

- Presentation outline

To understand HA requirements

- Understand FDA / PMDA statement and feedback from pilot projects.



To know actual cases

- Learn from others. Some organizations will be a good example.



To settle package managements

- To ensure “Reproducibility”, “Reliability” and “Reusability”.



HA requirements

- **“PMDA has not expressed any restrictions on the software used for new drug applications.”**, Yuki Ando, 19th DIA Japan Annual Meeting 2022
- Whether sponsors provide the SAS program or the R program, what we do as authorities is the same. However, it is necessary to share the execution environment with PMDA in more detail.
- The credibility of Open Source is secured by sponsors.



R submission pilots – R consortium

	Purpose	Status	Success criteria
Pilot 1	- Common Analyses. - Example R based submission contains common analyses, submission through the current FDA e-submission system and process.	Completed	For Submitter • Submission of proprietary R packages • Preparation of R-based submission materials For FDA Staff • Reproduce analysis results by reconstructing submitted proprietary packages / retrieving open source packages
Pilot 2	- Shiny Submission. - Example R based submission contains a shiny app, submission through the current FDA e-submission system and process.	Completed	For Submitter • Submission of a shiny app through the eCTD portal • The shiny codes will be provided as a proprietary R package For FDA Reviewer • Reproduce analysis results by deploying the shiny app on a FDA laptop/server

R submission pilots – R consortium

	Purpose	Status	Notes
Pilot 3	- Generation of ADaM Data Sets. - Example R based submission contains ADaM data sets generated using R based on the content of the Pilot 1 submission.	Ongoing	Extension of pilot 1. Updated on 28August, 2023. Submitted pilot 3 package through the FDA eCTD gateway. • M5 • tabulation • analysis • datasets (contains ADaM xpt derived by R) • programs • ADaM and TFL programs by R <pre>library(haven) library(admiral) library(dplyr) library(tidyr) library(metacore) library(metatools) library(pilot3) library(xportr) library(janitor)</pre> ◀ Example, libraries in adsl.r.
Pilot 4	- Alternative Submission Formats. - Utilized alternative methods of distributing a self-contained submission bundle with container technology.	Ongoing	Linux containers / Web Assembly software to bundle a Shiny application.



R submission actual – What to submit

Request to submit R code for all TFL of ISS /
main safety,

Unsuccessful in restoring environment

No dependency to internal R packages

Programs

All ADaM datasets
Selected TFLs (Primary /
Secondary)

All ADaM datasets
All TFLs

All ADaM datasets
All TFLs

Packages

Package from CRAN /
NN package

Package from CRAN /
Re-packaged NN
package

Package from CRAN /
Re-packaged NN
package **with setup
information**

ADRG

1 page

3 pages

24 pages

Package Management in Novartis.

HIGH

MIDDLE

LOW

- **High trust & reputable sources**
 - base, recommended, tidyverse, ...
 - background check of package's author/institution
- **Community usage**
 - download rate & reverse dependencies
- **Display/visual or data only package**

- **Test coverage**
 - Threshold min. 60%
- **Trusted publications/ citation**
 - publications, peer reviews in accepted/approved journals

- **Maintenance practice**
 - documentation, release date, frequency of updates
- **PQ tests functions that intend to be used**
 - For regulatory submissions, **you may only use the tested functions**
 - In adhoc-installs **the project team must approve and assume the risk**

Topics

- About Open Source Technology Working Group (OST WG)
- **OST WG Activities 2023**
 - Summary of PHUSE R Workshop
 - Presentation at DIA Japan 2023
 - Learn and use a generative AI tool (ChatRWKV) to increase our productivity



Generative AI Tool (ChatRWKV)

Background:

- Generative AI tool (like ChatGPT) might be a game changer in many industries
- Many pharmaceutical companies have introduced internal generative AI tools and been considering how to use them effectively
- We believe learning about this new technology is important to increase our productivity



Generative AI Tool (ChatRWKV)

How to learn generative AI tool?



Build a generative AI tool by myself
Customize the generative AI tool by fine-tuning



How about ChatRWKV?



Generative AI Tool (ChatRWKV)

What is ChatRWKV?

- Pronounced as "RwaKuv", from 4 major params: R W K V
- Like ChatGPT but powered by RWKV (100% RNN) language model
- Requires small GPU memory
- Open Source
- User can do fine tuning



Generative AI Tool (ChatRWKV)

What we have done:

- Chose Google Colaboratory as Development Environment
- Google Colaboratory:
 - Cloud-based development environment
 - Run in user's web browser
 - User can use GPU memory
 - Free with restrictions

Generative AI Tool (ChatRWKV)

← → ↺ colab.research.google.com/?hl=ja

Google Colaboratory

Colaboratory へようこそ
ファイル 編集 表示 挿入 ランタイム ツール ヘルプ

+ コード + テキスト ドライブにコピー 接続 ▾

Colab へようこそ

すでに Colab をよくご存じの場合は、この動画でインタラクティブなテーブル、実行されたコードの履歴表示、コマンド パレットについてご覧ください。



Colab とは

Colab（正式名称「Colaboratory」）では、ブラウザ上で Python を記述、実行できます。以下の機能を使用できます。

- 環境構築が不要
- GPU に料金なしでアクセス
- 簡単に共有

Colab は、**学生からデータサイエンティスト、AI リサーチャー**まで、皆さんの作業を効率化します。詳しくは、[Colab の紹介動画](#)をご覧ください。下のリンクからすぐに使ってみることもできます。

ref: <https://colab.research.google.com/?hl=ja>



Generative AI Tool (ChatRWKV)

What we have done: Built own ChatRWKV on Google Colaboratory

Ex.: Make ChatRWKV to create an original story.

* Model: RWKV-4-Raven-3B-v12-Eng98%-Other2%-20230520-ctx4096.pth

Bob: can you write an original story about SAS programmer?

Alice: Sure, here's an original story about a SAS programmer named Max:

Max was a talented SAS programmer who had been working for the same company for over 20 years. He was known for his exceptional programming skills and his ability to solve complex problems with ease.

One day, Max received a call from his boss, who told him that he had been promoted to the position of CEO. Max was overjoyed and immediately began planning his next project.

But as he worked on his next project, Max began to notice some strange things happening around him. He noticed that the company's servers were crashing more frequently than usual, and he noticed that the company's productivity was decreasing.

Max decided to investigate further and discovered that there was a rogue program running in the background of the server. The rogue program was causing all sorts of problems, from slowing down the server to corrupting data.

Max knew that he had to stop the rogue program before it caused any more damage. He started by disabling all unnecessary programs and services on his server, but still, the problem persisted.

Max then decided to take matters into his own hands and use his programming skills to find a solution. He spent hours analyzing the code and debugging it until he finally found a solution.

With the help of some other programmers, Max was able to fix the problem and restore order to the server. The company's productivity improved dramatically, and Max received a promotion for his efforts.

From that day forward, Max continued to work as a SAS programmer, always striving to improve his skills and make a positive impact on his colleagues and customers.



Generative AI Tool (ChatRWKV)

What we have done: Developed a beginner's guide about how to build ChatRWKV on Google Colaboratory

+ Code + Text

▼ ChatRWKVの試しに試してみるためのチュートリアル

各種ファイルが消えないように、Google Driveをマウントする。

▶

```
from google.colab import drive
drive.mount('/content/drive')
```

ChatRWKV自体のダウンロード

```
[ ] import shutil
#shutil.rmtree('./ChatRWKV')

!git clone https://github.com/BlinkDL/ChatRWKV
```

ライブラリのインストール

PyTorchはGoogle Clabolatoryの環境にインストール済みのため、インストールは不要。

```
[ ] !pip install rwkv
!pip install ninja
!pip install numpy
!pip install -r ./ChatRWKV/requirements.txt
```

+ Code + Text All changes saved

Connect 14 ▾

👤 ⚙️ ▾

モデルのダウンロード。ここでは3B/パラメータのモデルをダウンロードします。保存先はdrive/MyDrive/Colab Notebooks/modesに保存されます。すでにモデルをダウンロード済みの場合は、この手順をスキップしてください。

```
[ ] !wget -P ./drive/MyDrive/Colab\ Notebooks/models https://huggingface.co/BlinkDL/rwkv-4-raven/resolve/main/RWKV-4-Raven-3B-v12-Eng98%25-Other2%25-20230520-ctx4096.pth
```

ダウンロードしたモデル使用するようにchat.pyを書き換える。

```
[ ] with open("./ChatRWKV/v2/chat.py") as f:
    chat_py_str = f.read()

    new_chat_py_str = chat_py_str.replace('\./fsx/BlinkDL/HF-MODEL/rwkv-4-raven/RWKV-4-Raven-14B-v12-Eng98%-Other2%-20230523-ctx8192\'',
    '\./drive/MyDrive/Colab Notebooks/models/RWKV-4-Raven-3B-v12-Eng98%-Other2%-20230520-ctx4096\'' )

    with open("./ChatRWKV/v2/chat.py", mode='w') as f:
        f.write(new_chat_py_str)
```

ターミナルが表示されたら、ターミナルにpython ./ChatRWKV/v2/chat.pyと入力し、Enterを押します。モデルのロードが少し遅いですが、しばらく待ちましょう。

Bob:と表示されたら、準備完了です。

ちなみに、キーレスポンスはメチャメチャ悪いです。

▶

```
!pip install colab-xterm
%load_ext colabxterm
%xterm
```

Generative AI Tool (ChatRWKV)



Build a generative AI tool by myself



Customize the generative AI tool by fine-tuning





Generative AI Tool (ChatRWKV)

What we have done: fine-tuning by PMDA FAQ

Files

- README.md
- RWKV_ChatRWKV.ipynb
- RWKV_LangChain.ipynb
- RWKV_LoRA.ipynb
- RWKV_v2_RNN_Pile_Fine_Tuning....
- RWKV_v3_RNN_Pile_Fine_Tuning....
- RWKV_v4_RNN_Pile_Fine_Tuning....
- RWKV_v4neo_Fine_Tuning.ipynb

Preview Code Blame 311 lines (311 loc) · 10.6 KB

[Open in Colab](#)

RWKV-v4-RNN-Pile Fine-Tuning

RWKV is an RNN with transformer-level performance

This notebook aims to streamline fine-tuning RWKV-v4 models

Setup

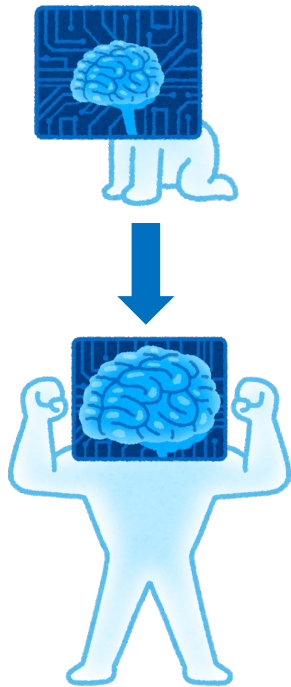
```
In [ ]:
##@title Google Drive Options { display-mode: "form" }
save_models_to_drive = True #@param {type:"boolean"}
drive_mount = '/content/drive' #@param {type:"string"}
output_dir = 'rwkv-v4-rnn-pile-tuning' #@param {type:"string"}
tuned_model_name = 'tuned' #@param {type:"string"}

import os
from google.colab import drive
if save_models_to_drive:
    from google.colab import drive
    drive.mount(drive_mount)

output_path = f"{drive_mount}/MyDrive/{output_dir}" if save_models_to_drive else f"/content/{output_dir}"
os.makedirs(f"{output_path}/{tuned_model_name}", exist_ok=True)
os.makedirs(f"{output_path}/base_models/", exist_ok=True)

print(f"Saving models to {output_path}")

In [ ]: !nvidia-smi
```



[Ref: RWKV-v4-RNN-Pile Fine-Tuning](#)



Generative AI Tool (ChatRWKV)

What we have done: fine-tuning by PMDA FAQ

Original data ([ref: https://www.pmda.go.jp/files/000264582.pdf](https://www.pmda.go.jp/files/000264582.pdf))

Q1-1: In the gateway application, if the application form is to be replaced and FD application data for replacement are submitted using the gateway system, is it necessary to separately submit in writing a replacement request form and the application form to be replaced?

A: Yes. In case of replacement, please also submit the application notice receipt form as with the case of new drug applications.

Q1-1-1: In reference to Q1-1, it is stated that if the replacement request form and the application form to be replaced are brought or mailed to the PMDA window, the application notice receipt form must also be submitted. Is it permissible if a brand name, etc. is changed during review and it is consequently not consistent with information at the time of giving the application notice printed on the application notice receipt form?

A: Yes.

Generative AI Tool (ChatRWKV)

What we have done: fine-tuning by PMDA FAQ

Learning data (txt file)

```
# Instruction:↓
Q1-1: In the gateway application, if the application form is to be replaced and FD application data for replacement are submitted using the gateway system, is it necessary to separately submit in writing a replacement request form and the application form to be replaced?↓
↓
# Response:↓
A: Yes. In case of replacement, please also submit the application notice receipt form as with the case of new drug applications.↓
<|endoftext|>↓
↓
# Instruction:↓
Q1-1-1: In reference to Q1-1, it is stated that if the replacement request form and the application form to be replaced are brought or mailed to the PMDA window, the application notice receipt form must also be submitted. Is it permissible if a brand name, etc. is changed during review and it is consequently not consistent with information at the time of giving the application notice printed on the application notice receipt form?↓
↓
# Response:↓
A: Yes.↓
<|endoftext|>↓
.
```



Generative AI Tool (ChatRWKV)

What we have done: Fine-tuning by PMDA FAQ

Q1-15: When seeking advice on the scope of electronic study data to be submitted at the clinical trial consultation, what kind of consultation document should an applicant submit?

PMDA FAQ:

A: In addition to the documents used for the consultation on the application data package, applicants are required to give an explanation of the appropriateness of the scope of electronic study data to be submitted based on the notification on electronic study data.

Before: Model: RWKV-4-Pile-430M-20220808-8066.pth

Bob: When seeking advice on the scope of electronic study data to be submitted at the clinical trial consultation, what kind of consultation document should an applicant submit?

Alice:: You may contact me if you want to know the detailed procedure for submitting an electronic clinical trial consultation document.

After:

Bob: When seeking advice on the scope of electronic study data to be submitted at the clinical trial consultation, what kind of consultation document should an applicant submit?

Alice:: Consultation on the scope of electronic study data to be submitted is required to be completed, when electronic study data in a format other than the CDISC standards is submitted.

Generative AI Tool (ChatRWKV)



Build a generative AI tool by myself



Customize the generative AI tool by fine-tuning





Generative AI Tool (ChatRWKV)

Actual examples in healthcare companies

Company	Can you use internal generative AI tool?	Are you allowed to input your company information into the internal generative AI tool?	Were your company information used as learning data?	Actual use cases	Comments
A pharma	Yes	Yes	No	translation, spell and grammar check of English writing, support for writing, support for programming	
B pharma	Yes	Yes with restrictions	No	summarize writing, get ideas, support for writing, spell and grammar check of English writing, support for programming	If Japanese is used, the reply are not accurate sometimes.
C pharma	No Plan to introduce				
D pharma	Yes	Yes with restrictions	Yes. But Partially	Search company information, get ideas, support for writing, summarize writing	



Generative AI Tool (ChatRWKV)

Actual examples in healthcare companies

Company	Can you use internal generative AI tool?	Are you allowed to input your company information into the internal generative AI tool?	Were your company information used as learning data?	Actual use cases	Comments
E pharma	Yes	Yes	No	support for writing, summarize writing	
F pharma	Yes	Yes	No	Unknown	

Generative AI Tool (ChatRWKV)

Ideas for future utilization:

- Deliverable Auto-generation:
 - Statistical analysis plan and TFL specs from study protocol
 - TLFs from statistical analysis plan and TFL specs
 - Programming source code for non-programmer
- Faster access to information we want to know:
 - PMDA FAQ
 - SOP
 - Internal documents about expense claim or IT systems
- Automatic judgment of adverse events whether 'known' or 'unknown' in pharmacovigilance





Generative AI Tool (ChatRWKV)

What we have learned:

- Generative AI Tool is very powerful if we can use it well
- The replies are not accurate when a small model is being used. Large model requires large GPU memory
- Necessity to learn new knowledge/technologies which we do not use for our daily work so as not to fall behind

Reference

- <https://github.com/BlinkDL/ChatRWKV>
- [Google Colaboratory](#)
- [Google Colab で RWKV を試す - ChatRWKV版](#)
- [Google Colab で RWKV のファインチューニングを試す](#)
- [RWKV-v4-RNN-Pile Fine-Tuning](#)
- [PMDA FAQ](#)

Thank you for your attention!



**The Global Healthcare
Data Science Community**

Contact Channels

📞 UK +44 1843 609600
✉ office@phuse.global
🌐 phuse.global

Social Media

🐦 [@phusetwitta](https://twitter.com/phusetwitta)
📘 [/phusebook](https://facebook.com/phusebook)
▶ [/phusetube](https://youtube.com/phusetube)
🌐 [/company/phuse](https://linkedin.com/company/phuse)

