

開源軟體介紹 - OSS Introduction

Hsin @ 信誠金融科技(Footprint AI)

2024/09/10



Download Slides

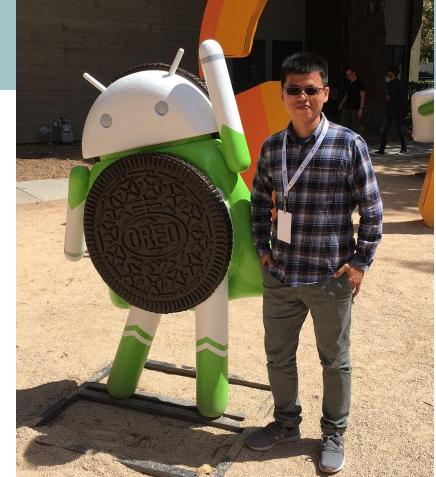
<https://bit.ly/4d7u85i>

<https://github.com/FootprintAI/talks/tree/main/slides>



About Me

- 2020 - Present at 信誠金融科技 (Footprint AI)
- 2016 - 2020 at IglooInsure (16M+ in series A+ 2020)
 - Provide digital insurance for e-economic world
 - Funded in KUL, Headquartered in Singapore
 - First employee/ Engineering Lead / Regional Head/ Chief Engineer
- 2013 - 2016 at Studio Engineering @ hTC
 - Principal Engineer on Cloud Infrastructure Team
- 2009 - 2012 at IIS @ Academia Sinica
 - Computer vision, pattern recognition, and data mining
- CS@CCU, CS@NCKU alumni



NVIDIA and TEAM AI In Taiwan



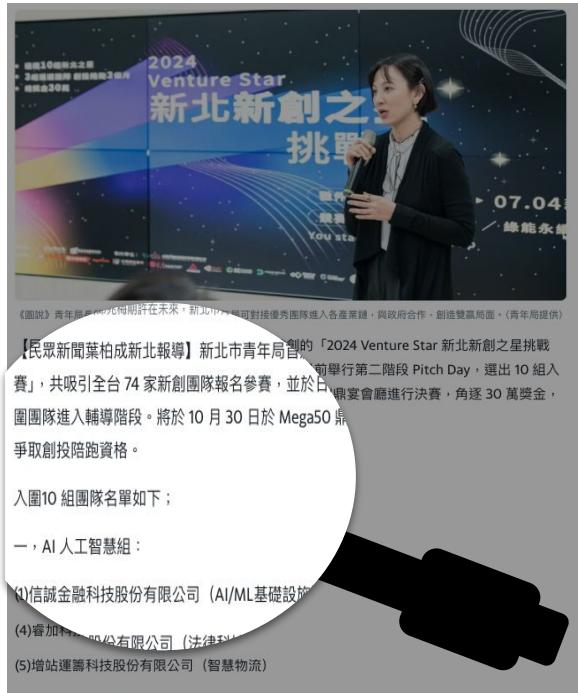
- Recognized by global events and competitions



2024 ComputeX/InnoVEX@Taiwan



2024 Techsauce@Thailand

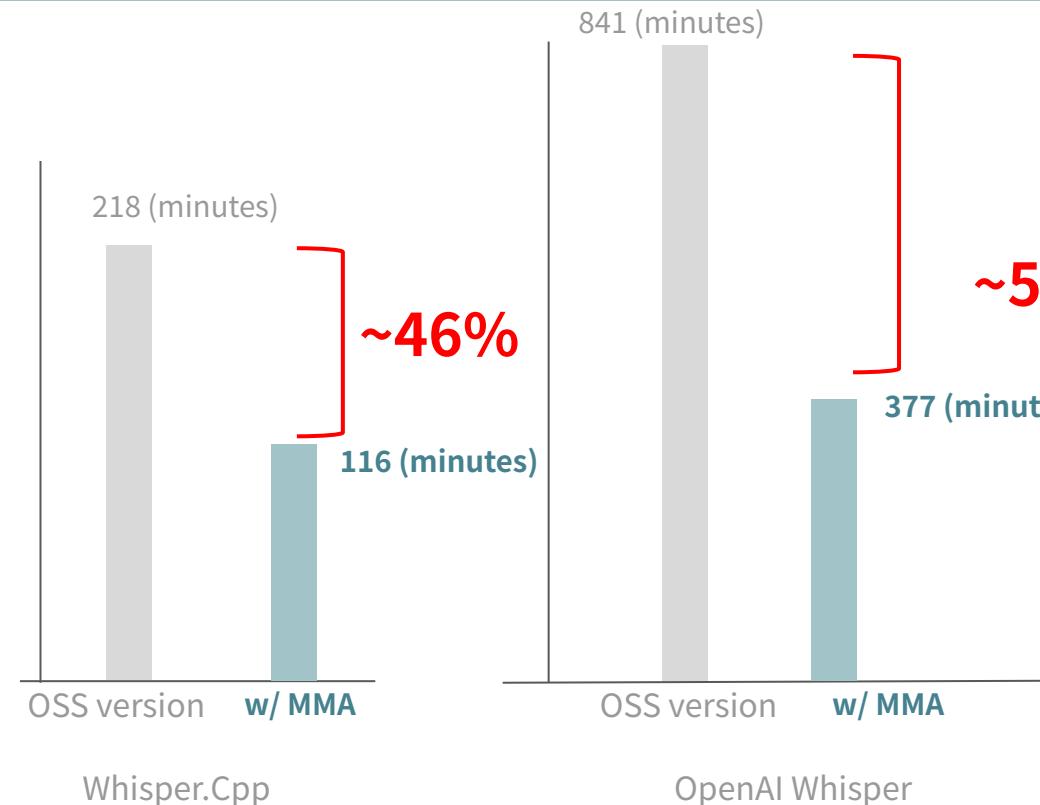


2024 Venture Start Competition

- Backed by Accelerators and Partners



Unlock a substantial increase in efficiency, secured by our cutting-edge intellectual property (TW/US).



And YES, we are **open source** (24Q4).

Follow our [linkedin](#)/[facebook](#) to get latest information.

Benchmarking with whisper.cpp(a c/cpp port version of whisper) and OpenAI Whisper (the native one) - using the same model to transcribe 72 hrs audios(in mp3), running on the accelerating resource RTX 3090 GPU(*) with Intel 13700k (10c), and DDR4 3200 24G.

Build for the NEXT big thing

- We are hiring interns for developing AI system, building AI models to solving unknown tasks.
- Feel free to connect if you're eager to embrace the **fail fast** ethos.



<https://www.bing.com/images/create/c...>

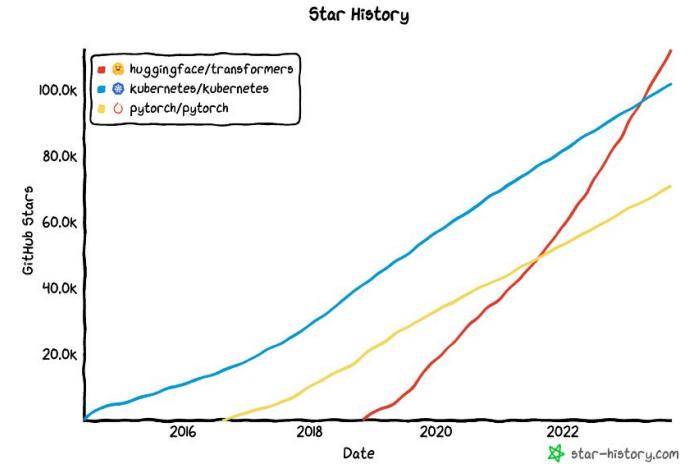
What is Open Source?



Huggingface sets the standard as an open-source startup role model



- An open source data science and machine learning platform, aka **github for AI**.
- Notable with its **open source transformer library**, which simplifies the process of working with pre-trained models.
- Huggingface Hub [1]
 - A central hub for community to share and explore machine learning.



year	2021		2023
Fundraising status	40M (B)		235M(D)

https://en.wikipedia.org/wiki/Hugging_Face
<https://github.com/huggingface/transformers>
<https://collabnix.com/what-is-hugging-face-and-why-it-is-damn-popular/> 9

Flower: A friendly federated learning framework

- Flower accelerates **the adoption of federated** and decentralized AI, providing a framework for training AI on distributed data, challenging traditional GPU-based methods, and enabling more secure, efficient, and **regulation-compliant AI** solutions, with notable backing from major investors and support from Fortune 500 companies and top universities.

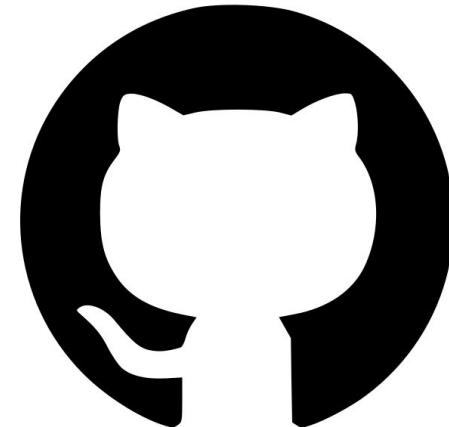
year	2024	
Fundraising status	20M (A)	



<https://flower.ai/>

Github: Let's Build From Here

- In April 2005, Linus Torvalds began developing Git after the free license for BitKeeper, the proprietary SCM system used for Linux kernel development since 2002, was revoked.
- GitHub founded in 2007, its name was chosen as a compound of *Git* and *hub*.
- Microsoft buys Github for \$7.8B in 2018.



What is Open Source Software? Definition

“Open-source software is a type of computer software ... the copyright holder grants users the rights to **use, study, change, and distribute** the software to anyone and for any purpose.” [1]

[1] https://en.wikipedia.org/wiki/Open-source_software

What is Open Source Software? In Plain Text

Source Code is publicly available.

Open to contribution.

Source code available with a license that permits users to freely run,
study, modify and redistribute.

Comparison

	Open Source Software	Proprietary Software
Ownership	Everyone	A group of limited people
Usability	Difficult use for unskilled users Ex: Desktop Linux	Feature neat to end-user Ex: Microsoft Windows
Security	Better (?) Ex: <u>PHP source code is injected backdoors by hackers</u> XZ-backDoor	Less
Stability	Less	Better
Opacity	Transparent	Closed

XZ-backDoor - Background

Screenshot of a GitHub profile page for user JiaT75.

Pinned:

- STest** (Public) - Multi-testing framework for C/C++. Easy to use by simply dropping stest.c and stest.h into your project!
C 13 ⚡ 8
- libarchive/libarchive** (Public) - Multi-format archive and compression library
C 3k ⚡ 762
- keithn/seatest** (Public) - Seatest - Simple C based Unit Testing
Shell 70 ⚡ 19

Jia Tan
JiaT75

Follow

589 followers · 1 following

jiat0218@gmail.com

Achievements:

- x3
- x2

Block or Report

Contribution activity:

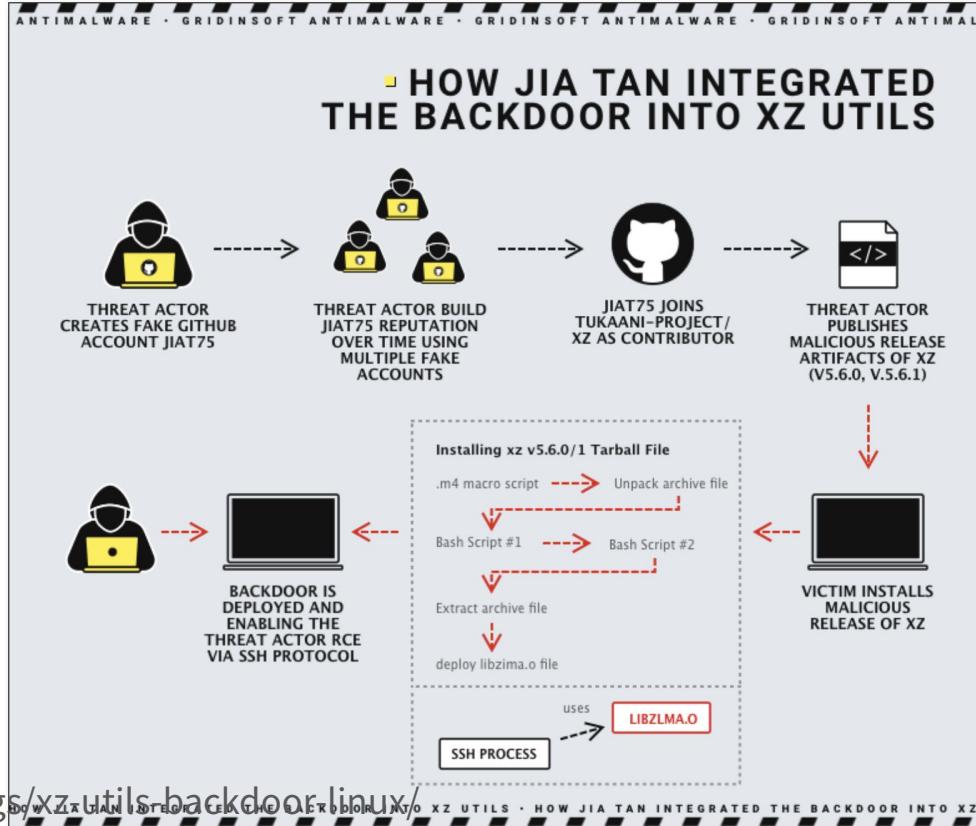
September 2024

JiaT75 has no activity yet for this period.

Timeline:

- 2024
- 2023
- 2022
- 2021
- 2020
- 2019
- 2018
- 2017

XZ-backDoor - Trajectory



XZ-backDoor - Discovered

Message-ID: <20240329155126.kjjfduxw2yrlxgzm@awork3.anarazel.de>
Date: Fri, 29 Mar 2024 08:51:26 -0700
From: Andres Freund <andres@...razel.de>
To: oss-security@...ts.openwall.com
Subject: backdoor in upstream xz/liblzma leading to ssh server compromise

Hi,

After observing a few odd symptoms around liblzma (part of the xz package) on Debian sid installations over the last weeks (logins with ssh taking a lot of CPU, valgrind errors) I figured out the answer:

The upstream xz repository and the xz tarballs have been backdoored.

At first I thought this was a compromise of debian's package, but it turns out to be upstream.

== Compromised Release Tarball ==

One portion of the backdoor is *solely in the distributed tarballs*. For easier reference, here's a link to debian's import of the tarball, but it is also present in the tarballs for 5.6.0 and 5.6.1:

https://salsa.debian.org/debian/xz-utils/-/blob/debian/unstable/m4/build-to-host.m4?ref_type=heads#L63

That line is *not* in the upstream source of build-to-host, nor is build-to-host used by xz in git. However, it is present in the tarballs released upstream, except for the "source code" links, which I think github generates directly from the repository contents:

<https://github.com/tukaani-project/xz/releases/tag/v5.6.0>
<https://github.com/tukaani-project/xz/releases/tag/v5.6.1>

Ref: <https://www.openwall.com/lists/oss-security/2024/03/29/4>

License

The misunderstanding on open source software

- Open source software is **Free** in terms of Freedom, not Monetary.
- Open source developer is **NOT** a free labor [1]
 - Sponsorship by companies
 - Add values(a.k.a layers) on the existing project
 - Professional consult
- Open source means **Pool** quality
 - Good quality when the project gained enough eyeballs
 - On average the density of code defect is smaller than proprietary software [2]



[1] <https://softwareengineering.stackexchange.com/questions/100685/making-money-with-open-source-as-a-developer>

[2] <https://blogs.worldbank.org/opendata/quality-open-source-software-how-many-eyes-are-enough>

Business Model of OSS

Where open source communities lives?

- Github
 - Free for any public repositories
- Github Sponsor
- SaaS business model
 - Hugging Face
- LLM foundational model
 - Grok-1 by xAI.

Key Files to focus

Key files to understand a github project

- README.md
 - How to build and use the project
- LICENSE
 - Terms and Conditions
- CONTRIBUTING.md
 - Rules on contributing

The screenshot shows a GitHub repository named "grok-1". The repository has 445 stars, 6.6k forks, and 40.1k issues. The main branch has 2 branches and 0 tags. The repository page displays a list of files with their commit history:

File	Description	Age
.gitignore	Create .gitignore for checkpoints (#149)	2 days ago
CODE_OF_CONDUCT.md	Add initial code	3 days ago
LICENSE.txt	Add initial code	3 days ago
README.md	Update huggingface link	2 days ago
checkpoint.py	Add initial code	3 days ago
model.py	Add initial code	3 days ago
pyproject.toml	Add initial code	3 days ago
requirements.txt	Corrected name of package "cuda12-pip" (#194)	yesterday
run.py	Add initial code	3 days ago
runners.py	Add initial code	3 days ago
tokenizer.model	Add initial code	3 days ago

A red box highlights the first three files: LICENSE.txt, README.md, and CONTRIBUTING.md.

About
Grok open release
Readme
Apache-2.0 license
Code of conduct
Activity
Custom properties
40.1k stars
445 watching
6.6k forks
Report repository

Releases
No releases published

Packages
No packages published

Contributors 7

How to contribute Open Source?



How to contribute ?

<https://opensource.guide/>

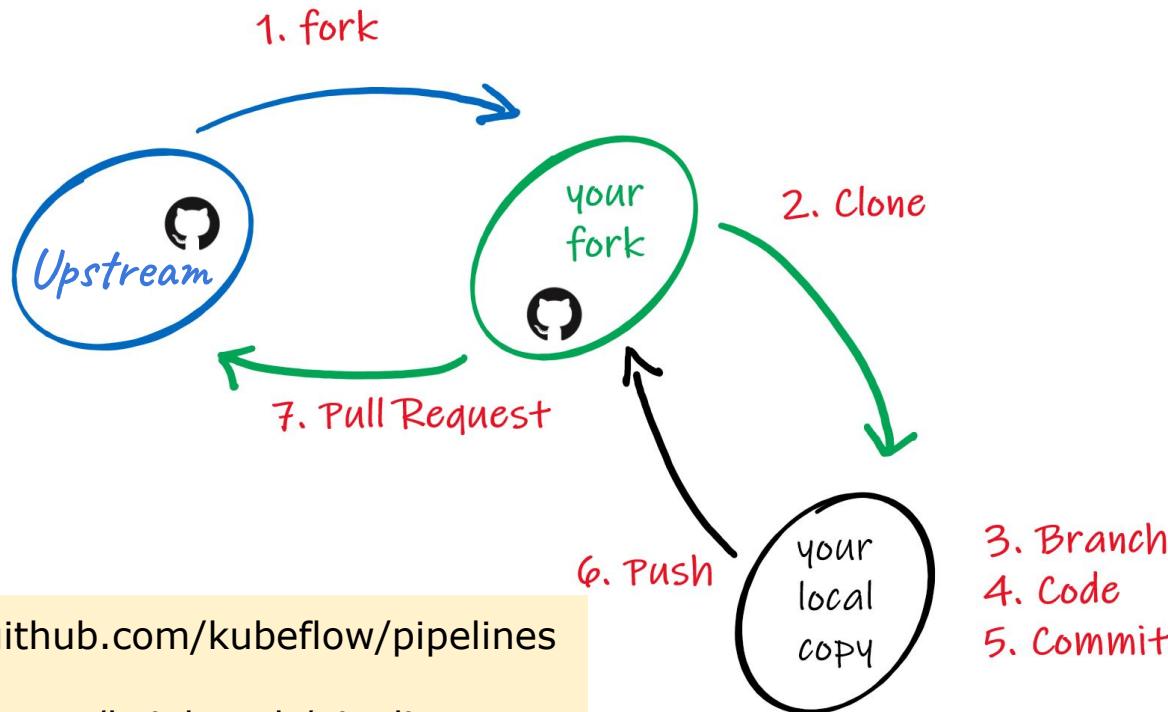
The two most popular way to contribute

1. Raising an issue where you can feel the project could be improved
 - a. Report a problem/bug (kind/bug)
 - b. Suggest new features (kind/feature)
 - c. Gap in documentation (doc)
2. Contribute to code
 - a. Pull-Request in Github

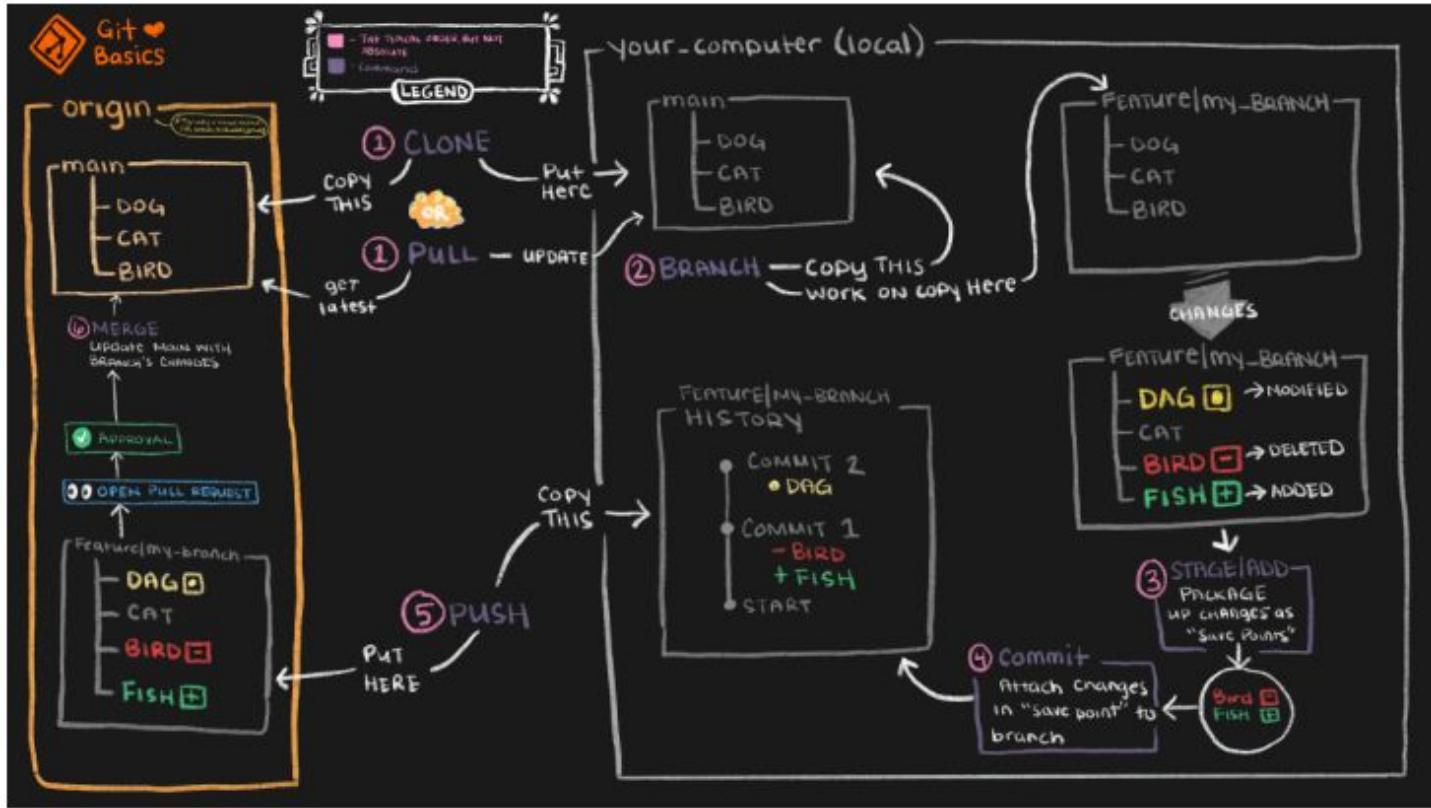
Issue list @kubeflow/pipeline

Author ▾		Label ▾	Projects ▾	Milestones ▾	Assignee ▾	Sort ▾
415 Open	✓ 1,720 Closed					
[frontend] The sort, filter of artifact does not work	#5565 opened 20 hours ago by capri-xiyue	area/frontend kind/bug				1
[Doc] Update "Visualize Results in the Pipelines UI" page	#5564 opened yesterday by Bobgy				Bobgy	2
[backend] OutputArtifact path is None	#5560 opened 2 days ago by willbry	area/backend kind/bug				
[feature] Configurable KFP Artifact types while storing to Minio	#5558 opened 2 days ago by rado-parrak	area/sdk kind/feature				2
[Multi User] Kubeflow Pipelines Missing Headers	#5556 opened 3 days ago by RoyerRamirez	area/backend kind/bug				
[feature] Allow jobs to be scheduled on AWS Fargate	#5555 opened 3 days ago by yuhuishi-convect	area/backend kind/feature				2
[sdk] mlpipelines-metrics path is wrong when using kfp.components.load_component_from_text	#5548 opened 4 days ago by santiagoolivar2017	area/sdk kind/bug				
pipeline_conf.set_ttl_seconds_after_finished(seconds=10) NOT WORKING	#5545 opened 4 days ago by prasadkyp7	area/backend kind/bug				
[sdk] input parameter created without default value	#5544 opened 5 days ago by amitza	area/sdk kind/bug				1

Pull-Request flow in Github



Git Basic



Explain Git

<https://onlywei.github.io/explain-git-with-d3/>

Visualizing Git Concepts with D3

This website is designed to help you understand some basic git concepts visually. This is my first attempt at using both SVG and D3. I hope it is helpful to you.

Adding/staging your files for commit will not be covered by this site. In all sandbox playgrounds on this site, just pretend that you always have files staged and ready to commit at all times. If you need a refresher on how to add or stage files for commit, please read [Git Basics](#).

Sandboxes are split by specific git commands, listed below.

Basic Commands

[git commit](#)
[git branch](#)

[git checkout](#)
[git checkout -b](#)

Undo Commits

[git reset](#)
[git revert](#)

Combine Branches

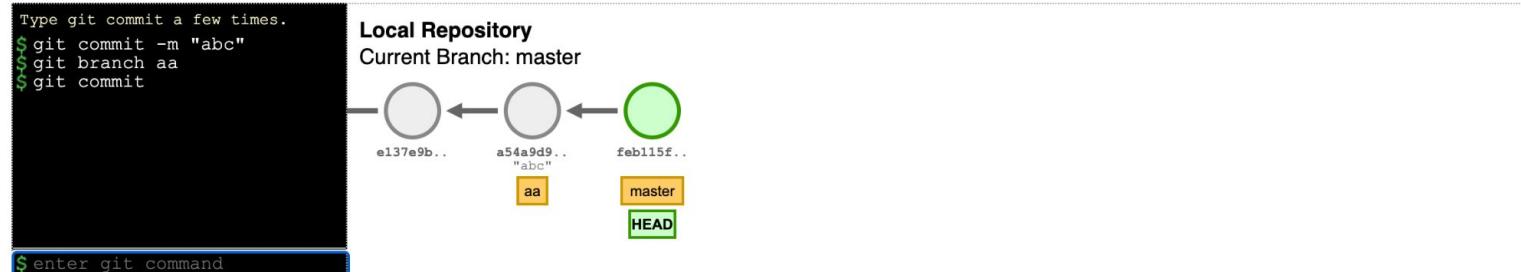
[git merge](#)
[git rebase](#)

Remote Server

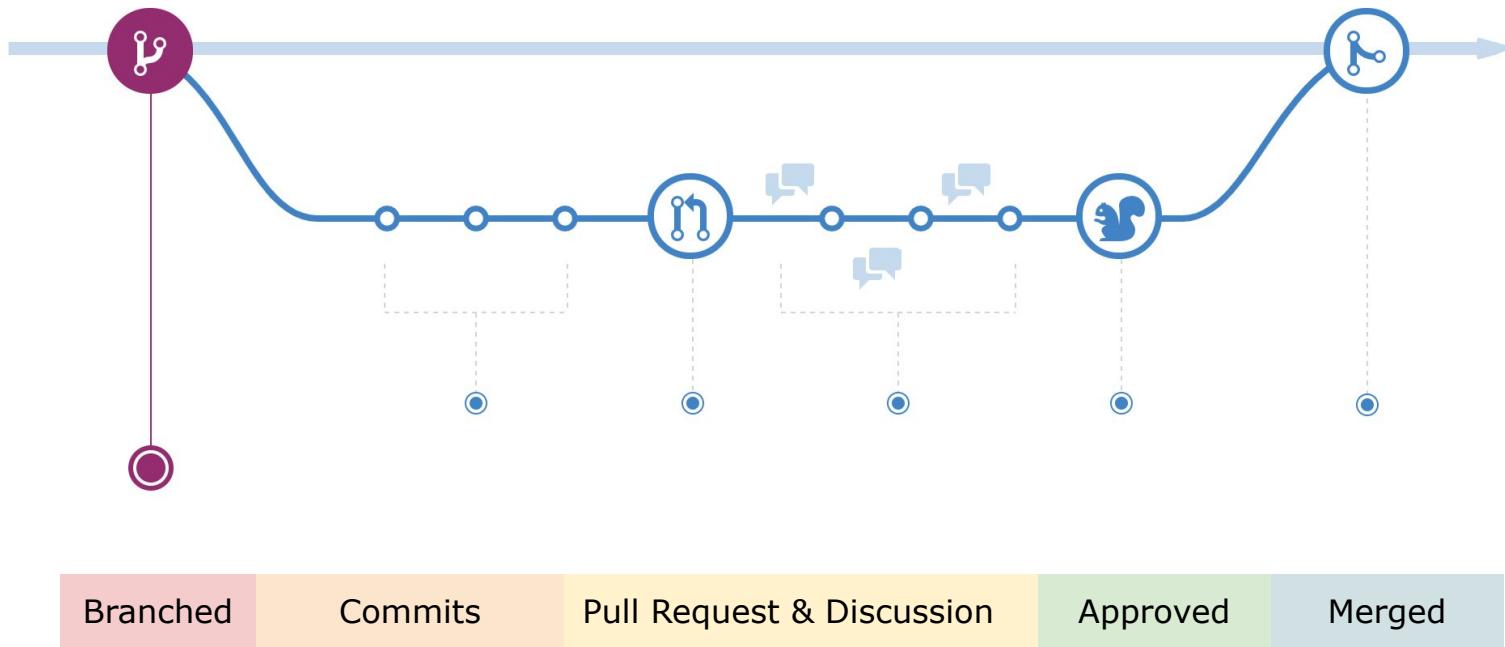
[git push](#)
[git pull](#)
[git tag](#)

We are going to skip instructing you on how to add your files for commit in this explanation. Let's assume you already know how to do that. If you don't, go read some other tutorials.

Pretend that you already have your files staged for commit and enter `git commit` as many times as you like in the terminal box.



Github Flow: branch-based workflow



Pull-Request list @kubeflow/pipeline

81 Open ✓ 3,342 Closed

Author ▾ Label ▾ Projects ▾ Milestones ▾ Reviews ▾ Assignee ▾ Sort ▾

feat(sdk.v2): reimplement custom job support ● cla: yes size/XL

#5572 opened 4 hours ago by ChenSun 1 of 1

Add OWNERS for v2 folder ● cla: yes lgtm size/XS

#5571 opened 10 hours ago by Bobgy 0 of 1

Add zijianjyu to frontend/OWNERS ● approved cla: yes size/XS

#5570 opened 10 hours ago by Rohan 0 of 1

chore(deps): bu

#5569 opened 11 ho

chore(google_c

size/M
#5554 opened 3 day

feat(sdk): Add a

#5552 opened 3 day

chore(v2): fix i

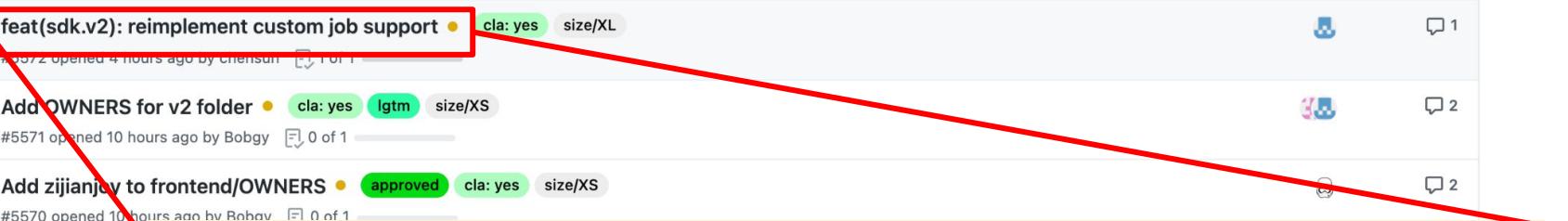
size/M
#5543 opened 5 day

chore(samples): v2 compatible exit handler sample ● cla: yes size/L

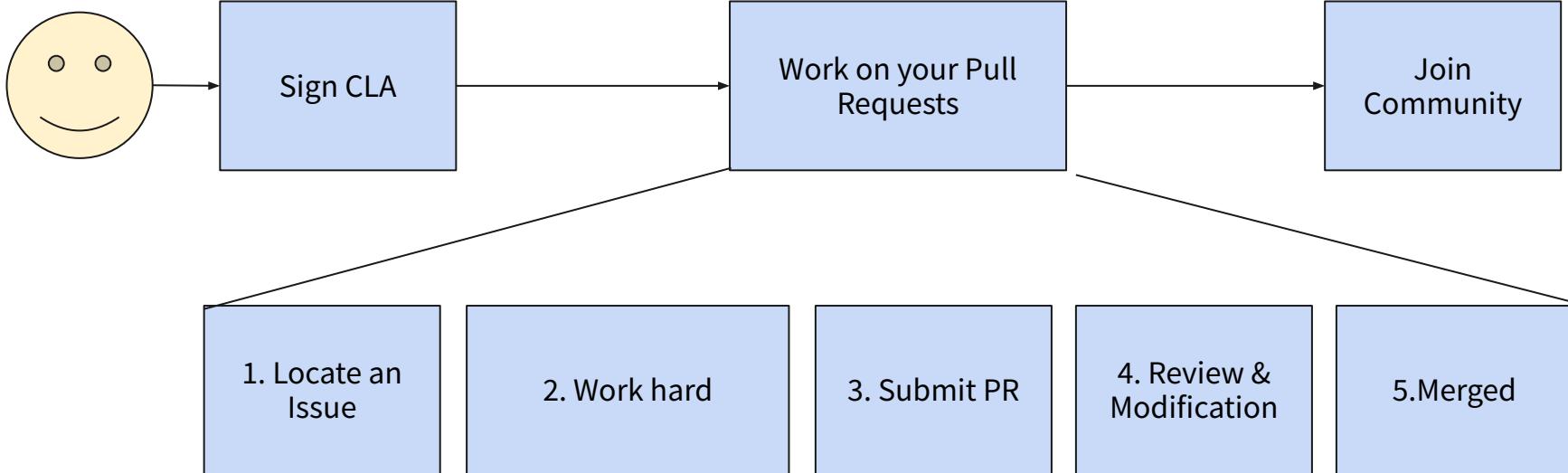
Ref: <https://www.conventionalcommits.org/en/v1.0.0/>

<type>(<scope>): <short summary>

- Summary in present tense. Not capitalized. No period at the end.
- Scope: common|compiler|authentication|core|
- Type: chore, docs, feat, fix, refactor, style, or test.

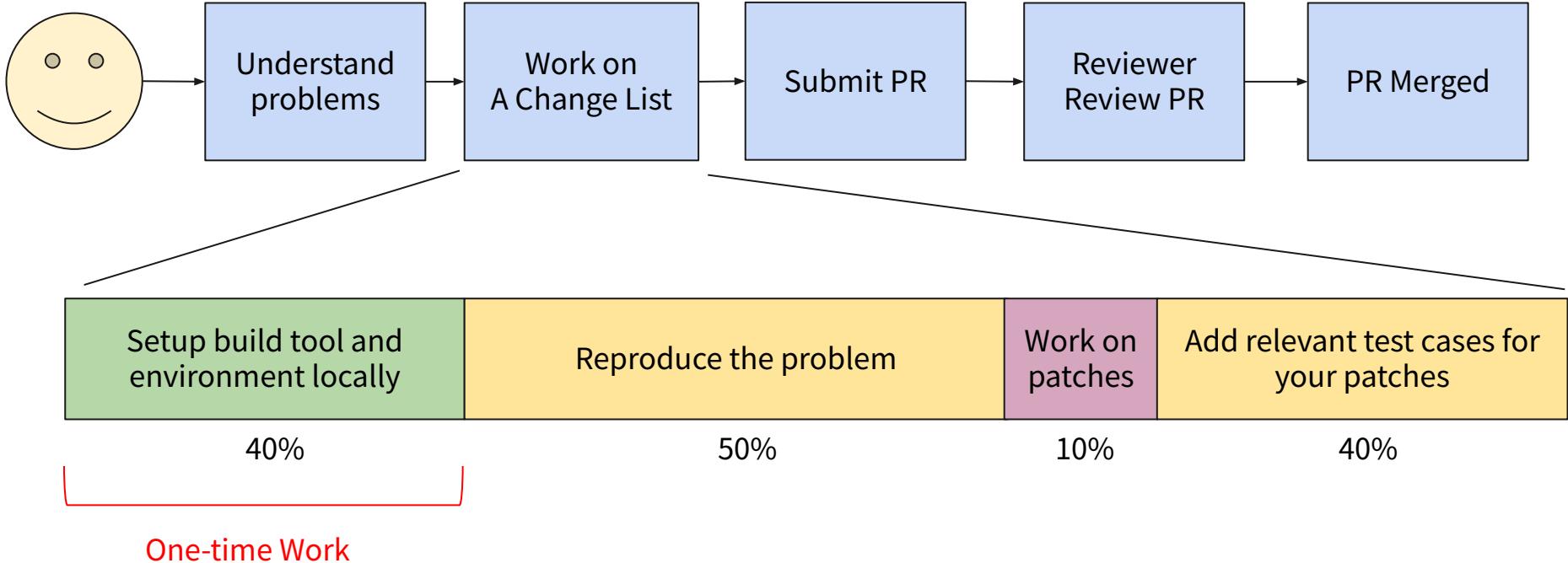


How to be a Contributor?



Ref: <https://www.kubeflow.org/docs/about/contributing/>

Why does the first contribution be important?



Case Study

<https://github.com/ggerganov/whisper.cpp/pull/2350>

ggerganov / whisper.cpp

Type ⌘ to search

Code Issues 643 Pull requests 64 Discussions Actions Projects 1 Wiki Security Insights

feat(go binding): add beamsize/entropyhold/maxcontext to context interface #2350

Merged ggerganov merged 6 commits into ggerganov:master from hsinhoyeh:feat---add-args---gobinding 2 weeks ago

Conversation 13 Commits 6 Checks 47 Files changed 6 +36 -3

hsinhoyeh commented last month

fixes: #2349

feat(go binding): add beamsize/entropyhold/maxcontext to context int... a8e9686

hsinhoyeh added a commit to FootprintAI/cockburn that referenced this pull request last month a26118b

feat: add flags and sync them with params a26118b

hsinatfootprintai mentioned this pull request last month

feat: add flags and sync them with params FootprintAI/cockburn#16 Merged

ggerganov approved these changes last month

View reviewed changes

Reviewers: ggerganov, hsinatfootprintai

Assignees: No one assigned

Labels: None yet

Projects: None yet

Milestone: No milestone

nats-io/nats.go#22 Case Study

 **Closed** goroutine leaking when calls BindRecvChan #22
hsinhyeh opened this issue on Jul 26, 2014 · 1 comment

in the underlying implementation, nats defines a callback (i.e. cb) and then call the Subscribe from a conn
<https://github.com/apcera/nats/blob/master/netchan.go#L81>

```
_ , err := c.Conn.subscribe(subject, queue, cb)
```

in the conn.subscribe, nats spawns a go routine to send the message:
<https://github.com/apcera/nats/blob/master/nats.go#L1198>

```
go nc.deliverMsgs(sub.mch)
```

and that go routine will be terminated "only if sub.mch is closed", which is called via conn.Unsubscribe().
<https://github.com/apcera/nats/blob/master/nats.go#L1244>

```
func (nc *Conn) unsubscribe(sub *Subscription, max int) error {
```

The bug actually happened, when we look BindRecvChan again: even the caller has closed the channel, the go routine who serves deliverMsg will never be destroyed. We actually have an additional zombie go routine when we call BindRecvChan every time.

 hsinhyeh referenced this issue in hsinhyeh/nats on Jul 26, 2014

 workaround the goroutine leak problem in apcera/nats/netchan.go   0d117eb

 derekcollison added the **bug** label on Jul 26, 2014

 derekcollison self-assigned this on Jul 26, 2014

 derekcollison commented on Jul 26, 2014   

```
The proposed solution could affect ordering of messages due to the read. But great bug report. Will fix. Thanks.
```

Projects
None yet

Milestone
No milestone

Linked pull requests
Successfully merging a pull request may close this issue.
None yet

Notifications  
You're receiving notifications because you were assigned.

2 participants 

googleapis/google-cloud-go#170 Case Study (1/2)

 [bigtable] bttest needs UpdateColumnFamily support #170
hsinohoyeh opened this issue on Sep 15, 2015 · 3 comments

diff --git a/google.golang.org/cloud/bigtable/bttest/inmem.go b/google.golang.org/cloud/bigtable/bttest/inmem.go
index 5c4bc09..f4dbe03 100644
--- a/google.golang.org/cloud/bigtable/bttest/inmem.go
+++ b/google.golang.org/cloud/bigtable/bttest/inmem.go
+func (s *server) UpdateColumnFamily(ctx context.Context, req *bttdpb.ColumnFamily) (*bttdpb.ColumnFamily, error
+ return nil, errors.New("no op")
+}

we need it because we call SetGCPolicy to configure the maximum versions stored in a cell

```
adminClient.SetGCPolicy( ctx, tableName, columnFamily, bgt.MaxVersionsPolicy(1) )
```

and in the underlying implementation, it will first call GetTable and then update its column family.

```
func (ac *AdminClient) SetGCPolicy(ctx context.Context, table, family string, policy GCPolicy) error {  
    prefix := ac.clusterPrefix()  
    tbl, err := ac.tClient.GetTable(ctx, &bttspb.GetTableRequest{  
        Name: prefix + "/tables/" + table,  
    })  
    if err != nil {  
        return err  
    }  
    fam, ok := tbl.ColumnFamilies[family]  
    if !ok {  
        return fmt.Errorf("unknown column family %q", family)  
    }  
    fam.GcRule = policy.proto()  
    _, err = ac.tClient.UpdateColumnFamily(ctx, fam)  
    return err  
}
```

<https://github.com/GoogleCloudPlatform/gcloud-golang/blob/master/bigtable/admin.go#L161>

 hsinohoyeh commented on Sep 16, 2015

hi,
I think I have fixed this (#170) and #166. you can see more in below.
you can purify my CL and push to the repo or I can create a RP if you think the CL is ok

 Projects
None yet

 Milestone
No milestone

 Linked pull requests
Successfully merging a pull request may close this issue.
None yet

 Notifications  Customize
 Unsubscribe
You're receiving notifications because you authored the thread.

 4 participants


[googleapis/google-cloud-go#170 Case Study \(2/2\)](#)

Gerrit CHANGES DOCUMENTATION BROWSE Sign in

Merged 3471 bigtable/bttest: support GetTable/UpdateColumnFamily

Change Info SHOW ALL
Submitted Sep 23, 2015
Owner Hsin-Ho Yeh
Uploader David Symonds
Committer Gerrit Code Review
Reviewers David Symonds
Repo | Branch [gocloud](#) | [master](#)
Topic

bigtable/bttest: support GetTable/UpdateColumnFamily
Fixes #166.
Fixes #170.
Change-Id: [1c2b3698583923042253d591b137c7192605ecf0](#)

Comments 25 resolved

Submit requirements
✓ Code-Review +2 David Symonds

Files	Comments	Findings	DOWNLOAD	INCLUDED IN	EXPAND ALL
Base → Patchset 4 bcca3e9					
File			Comments	Size	Delta
Commit message					
bigtable/bttest/inmem.go			+79 -9		
			+79 -9		

Change Log

Hsin-Ho Yeh	Uploaded patch set 1.	Patchset 1 Sept 19, 2015 16:32
David Symonds	16 comments Thanks for doing this. Please check the CLA details first. If you're working for a company then you might need to get them to approve your contribution.	Patchset 1 Sept 22, 2015 13:10
Hsin-Ho Yeh	Uploaded patch set 2.	Patchset 2 Sept 22, 2015 23:05
Hsin-Ho Yeh	16 comments the updated patch address the comments.	Patchset 2 Sept 22, 2015 23:19
David Symonds	10 comments Nearly there.	Patchset 2 Sept 23, 2015 07:12

[torvalds/linux#661 Case Study - Not all CLs are valuable](#)

[Open](#) Fixed typo: extra space. #661
jacob-macleod wants to merge 1 commit into `torvalds:master` from `jacob-macleod:master`

jacob-macleod commented on Mar 8, 2019

There was a double space.

10 2 13 1 6

-o Fixed typo: extra space. [...](#) Verified ee9ca9c

KernelPRBot commented on Mar 8, 2019

Hi @jacob-and-nathan!

Thanks for your contribution to the Linux kernel!

Linux kernel development happens on mailing lists, rather than on GitHub – this GitHub repository is a read-only mirror that isn't used for accepting contributions. So that your change can become part of Linux, please email it to us as a patch.

Sending patches isn't quite as simple as sending a pull request, but fortunately it is a well documented process.

Here's what to do:

- Format your contribution according to kernel requirements
- Decide who to send your contribution to
- Set up your system to send your contribution as an email
- Send your contribution and wait for feedback

How do I format my contribution?

The Linux kernel community is notoriously picky about how contributions are formatted and sent. Fortunately, they have documented their expectations.

Firstly, all contributions need to be formatted as patches. A patch is a plain text document showing the change you want to make to the code, and documenting why it is a good idea.

You can create patches with `git format-patch`.

Secondly, patches need 'commit messages' which is the human-friendly documentation explaining what the change is

Reviewers
 ssbarnea ✓
 StackNeverFlow ✓

Assignees
No one assigned

Labels
None yet

Projects
None yet

Milestone
No milestone

Notifications Customize [Subscribe](#)
You're not receiving notifications from this thread.

5 participants

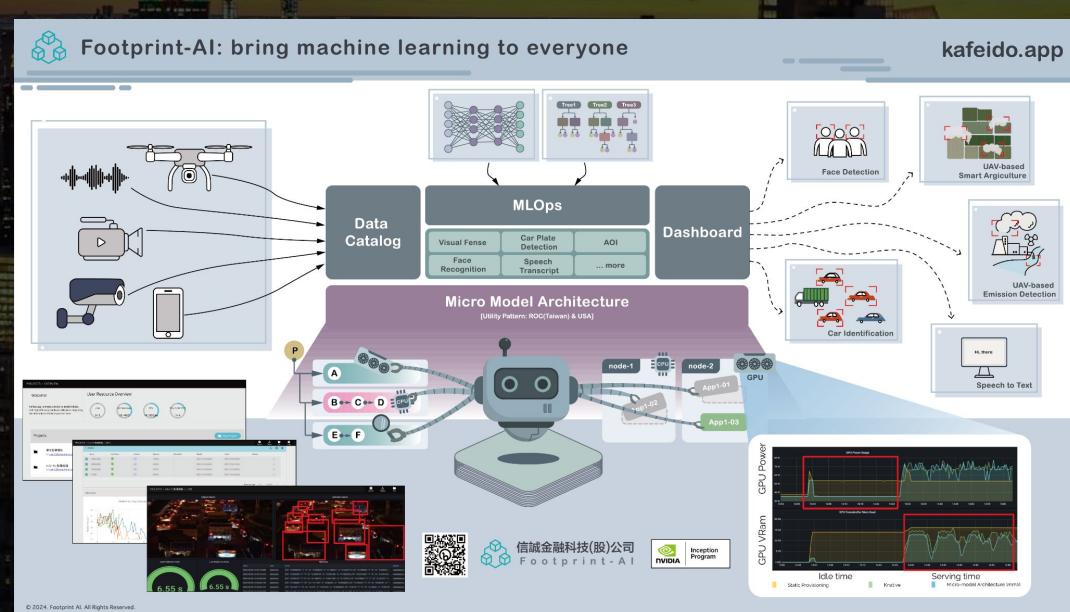
Take away

- Open source is a public artifacts that help you grow a reputation and a career.
- The work to contribute to Open Source could take days or weeks, but the rewards could keep years.
- Let's **consistently** make **small** changes.

Thank You

and follow us on Facebook /

LinkedIn



Contact Us: partnership@footprint-ai.com