Vision of project GUNDAM

Wenzhi Fu

Outline

- Vision:
  - Allow complicated program to be developed *within 1000 lines of C++ codes*.
  - With the *state-of-the-art performance*.
- For our group
  - Reduce the *meaningless low-quality* reconstruction
  - Each student can left something to the group after they have graduated
- For the entire field
  - Make it economics for the fundamental library to be continuous optimized.
  - Provide solid open-sourced code for others to follow.
  - Provides an efficient, sustainable, industrialized development model for ever growing complexity.

Vision of project GUNDAM

- Graph computing is a complex task and needs well-optimized algorithm, data structure and easy to use interface.
  - Same (dual simulation) algorithm, same index can still leads to ~3x performance gap

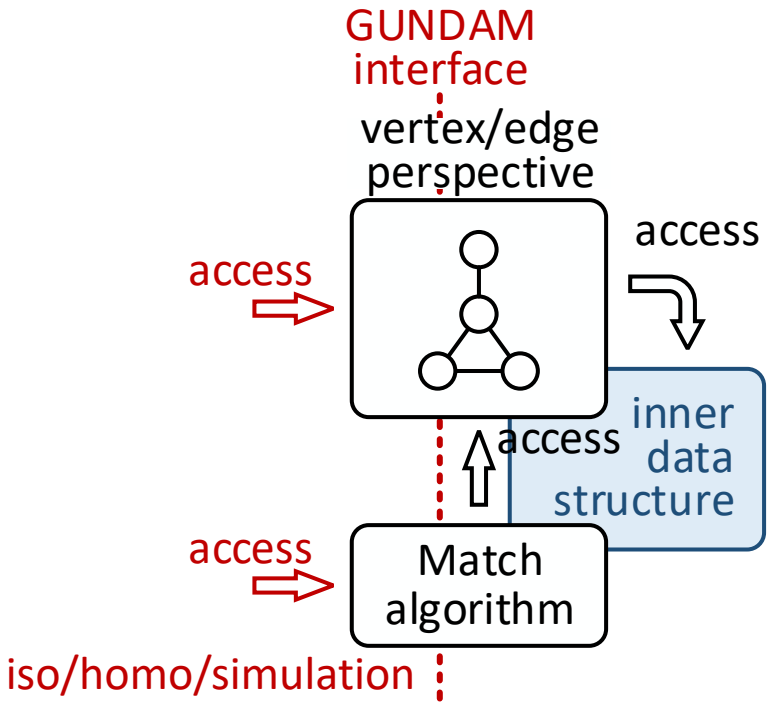
gar	log file	Graph	log file	LargeGraph
gar_918	./gor_match_new1_dbpedia_1_4_120_nb_inc_wenzhi.timelog	784.13516	./gor_match_new1_dbpedia_1_4_120_nb_inc_wenzhi_918_large_graph.timelog	2181.6447
gar_255	./gor_match_new1_dbpedia_1_4_120_nb_inc_wenzhi_255.timelog	585.4411	./gor_match_new1_dbpedia_1_4_120_nb_inc_wenzhi_255_large_graph.timelog	1737.9837

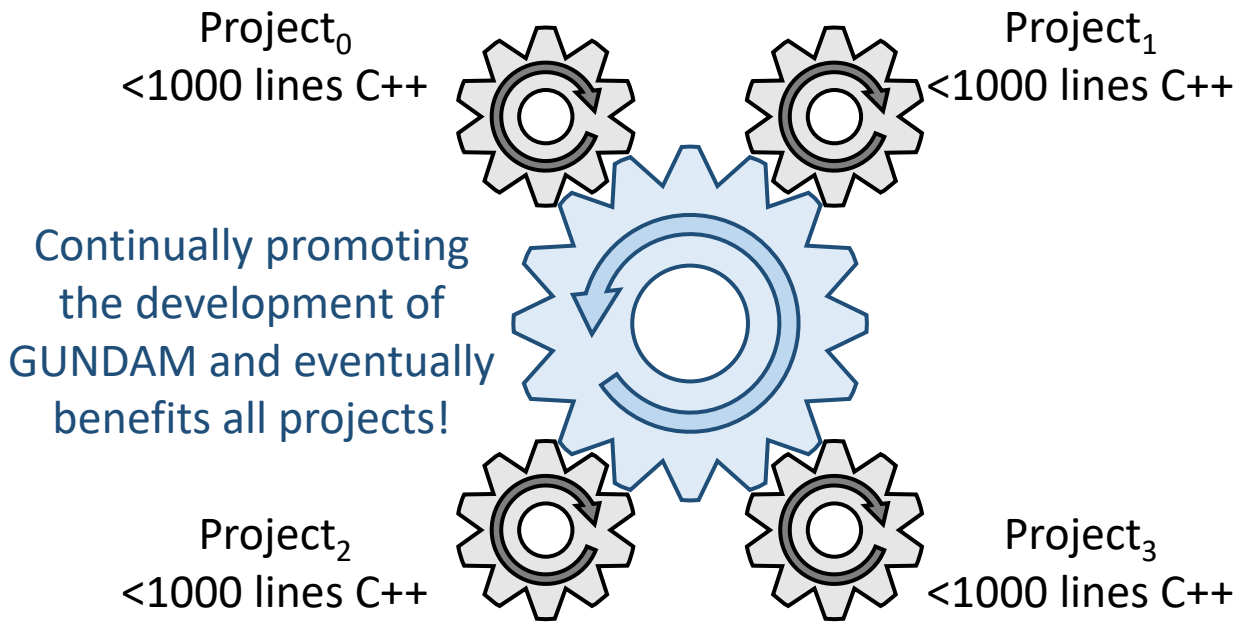
  - A self-implemented version can easily bring performance lost of several orders of magnitude.

once had >40x performance gap on pattern matching

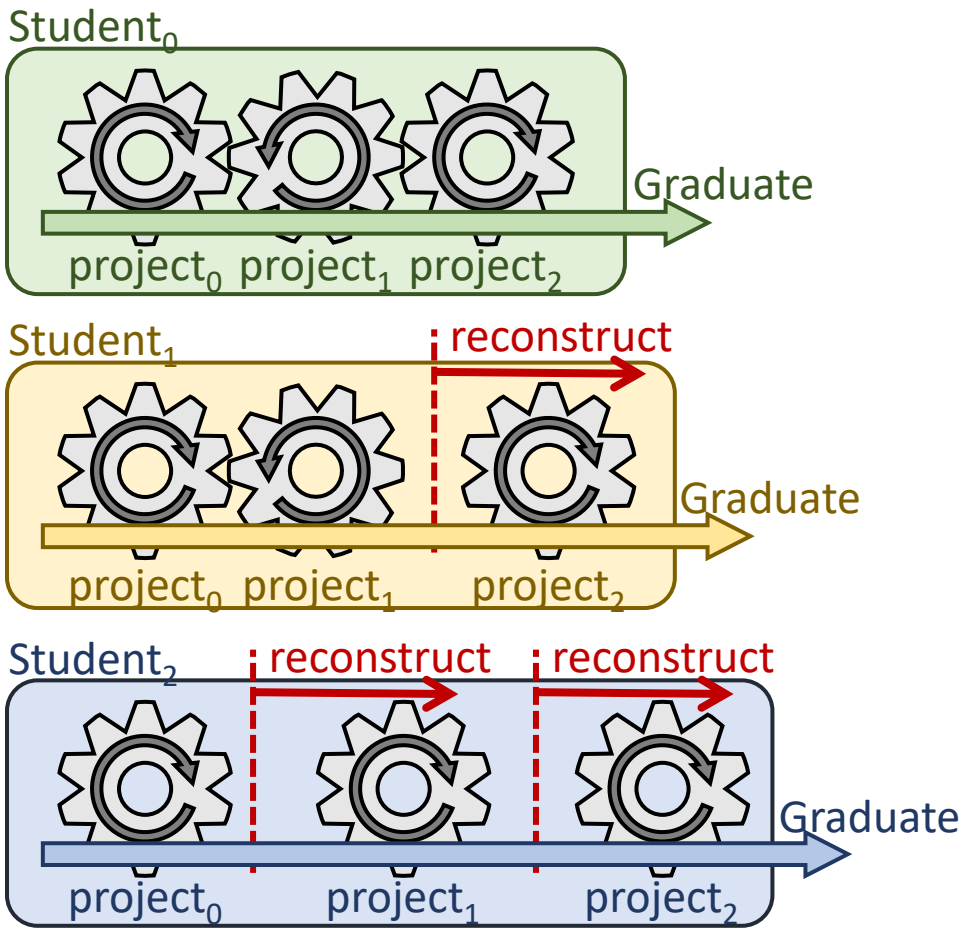
Vision of project GUNDAM

- Allow graph program to be developed *within 1000 lines of C++ codes*.
- With the *state-of-the-art performance*.

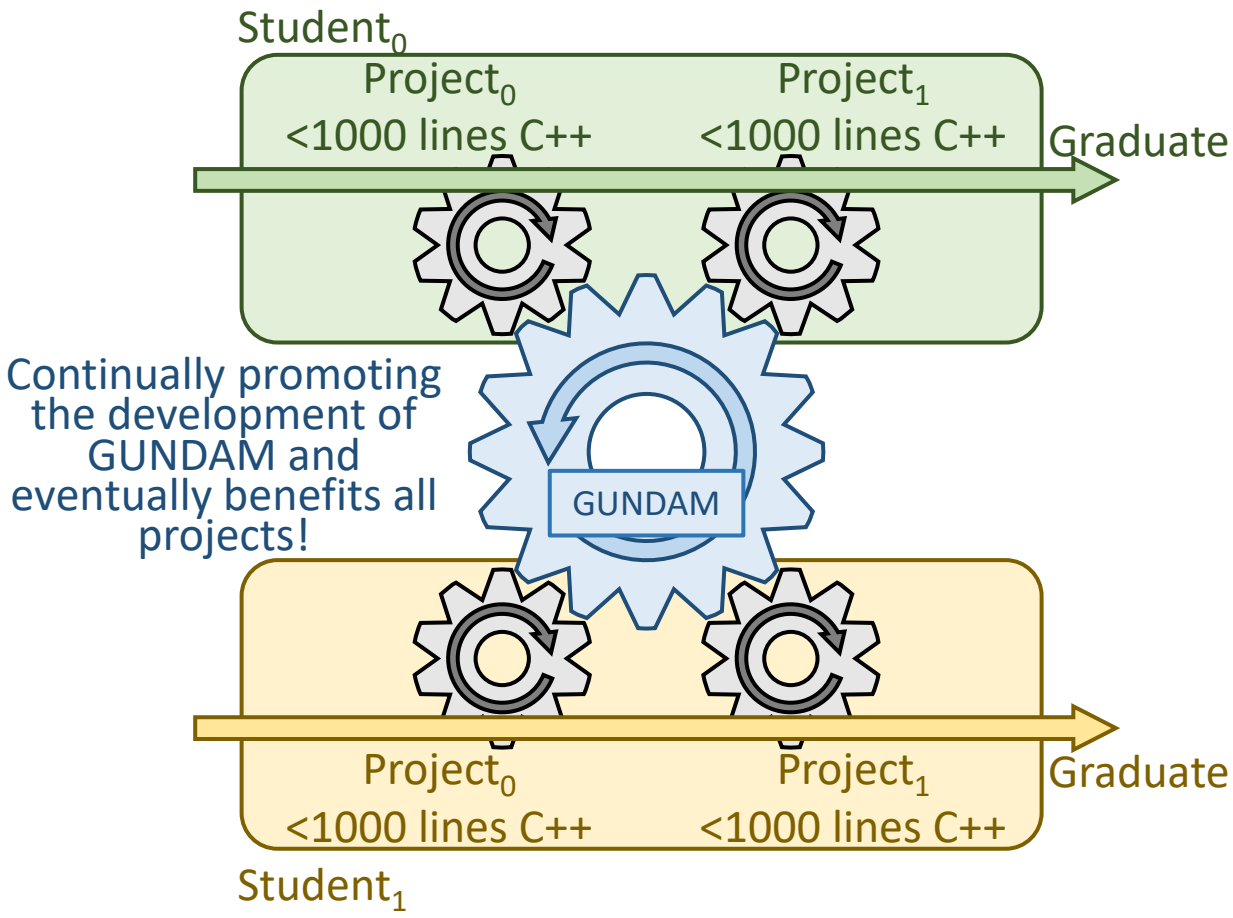




GUNDAM for the development of our group



Without GUNDAM

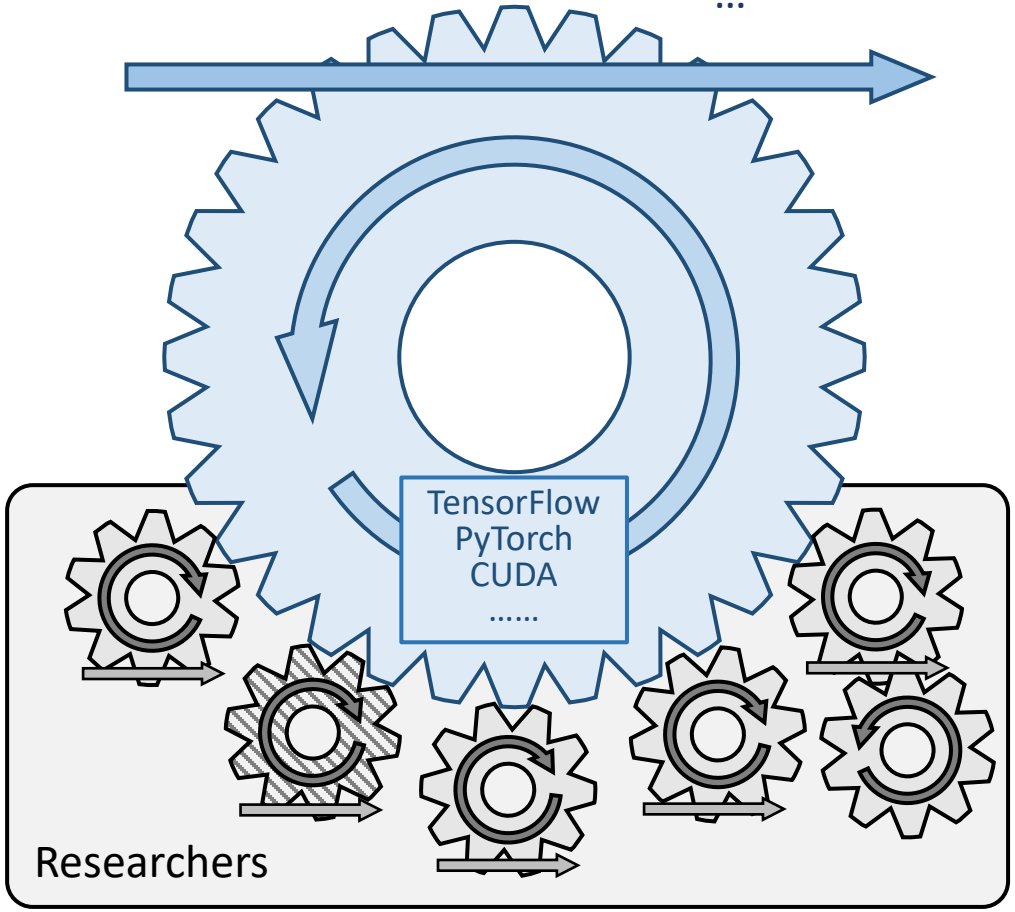


With GUNDAM

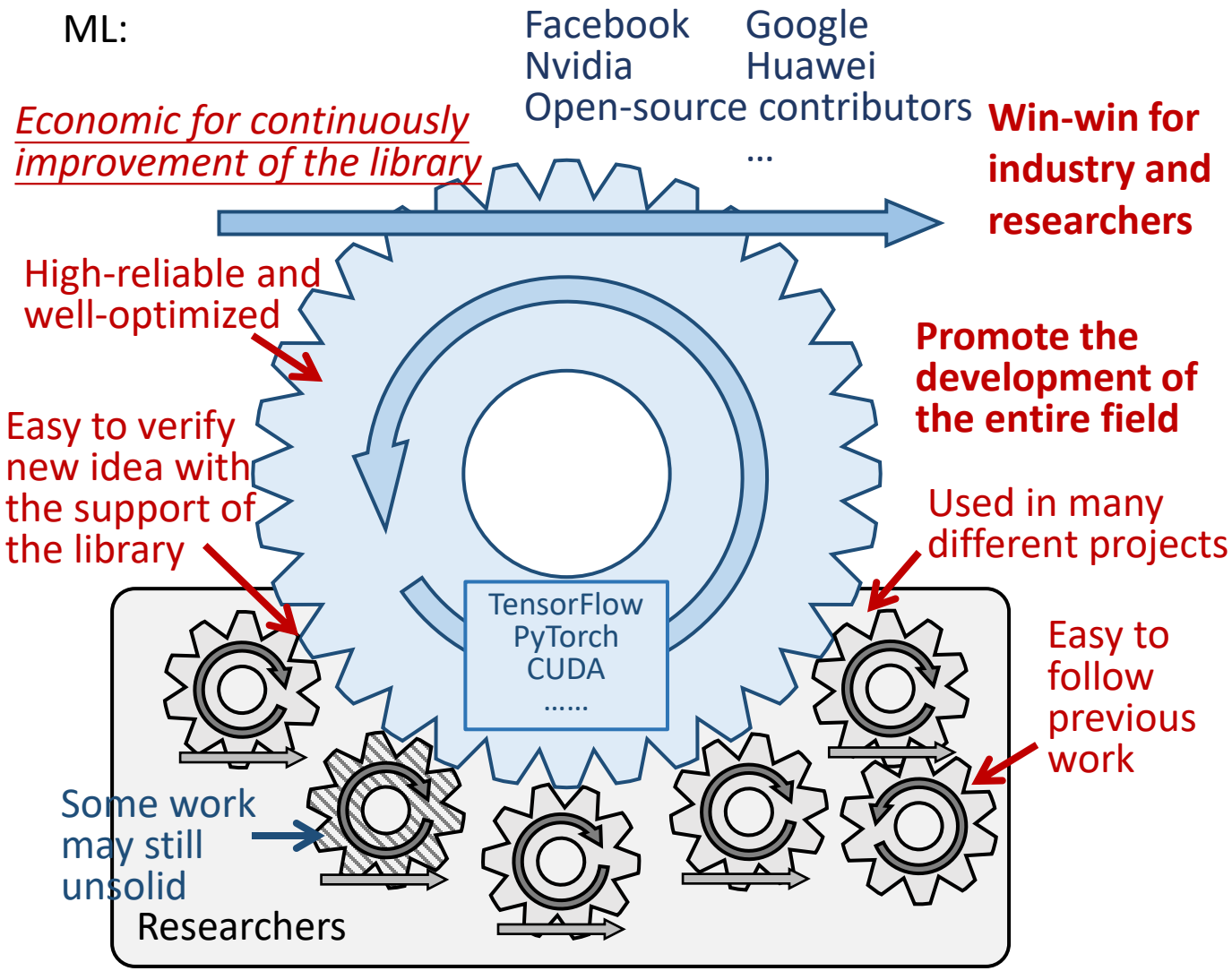
GUNDAM, provides an efficient, sustainable, industrialized development model of this field

ML:

Facebook Google
Nvidia Huawei
Open-source contributors
…



GUNDAM, provides an efficient, sustainable, industrialized development model of this field



GUNDAM, provides an efficient, sustainable, industrialized development model of this field

ML:

Facebook Google
Nvidia Huawei
Open-source contributors
…

Economic for continuously improvement of the library

Win-win for industry and researchers

High-reliable and well-optimized

Promote the development of the entire field

Easy to verify new idea with the support of the library

Used in many different projects

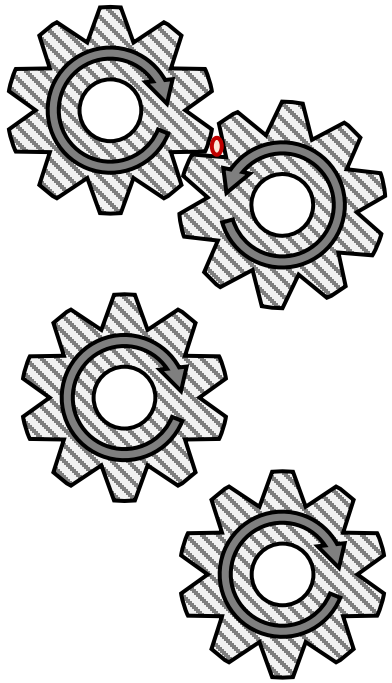
TensorFlow
PyTorch
CUDA
……

Easy to follow previous work

Some work may still unsolid

Researchers

GAR:



GUNDAM, provides an efficient, sustainable, industrialized development model of this field

ML:

Facebook Google
Nvidia Huawei
Open-source contributors
…

Economic for continuously improvement of the library

Win-win for industry and researchers

High-reliable and well-optimized

Promote the development of the entire field

Easy to verify new idea with the support of the library

Used in many different projects

TensorFlow
PyTorch
CUDA
……

Easy to follow previous work

Some work may still unsolid

Researchers

GAR:

Self-entertainment in a small group

Unsolid code make it hard to follow previous work even with the source code

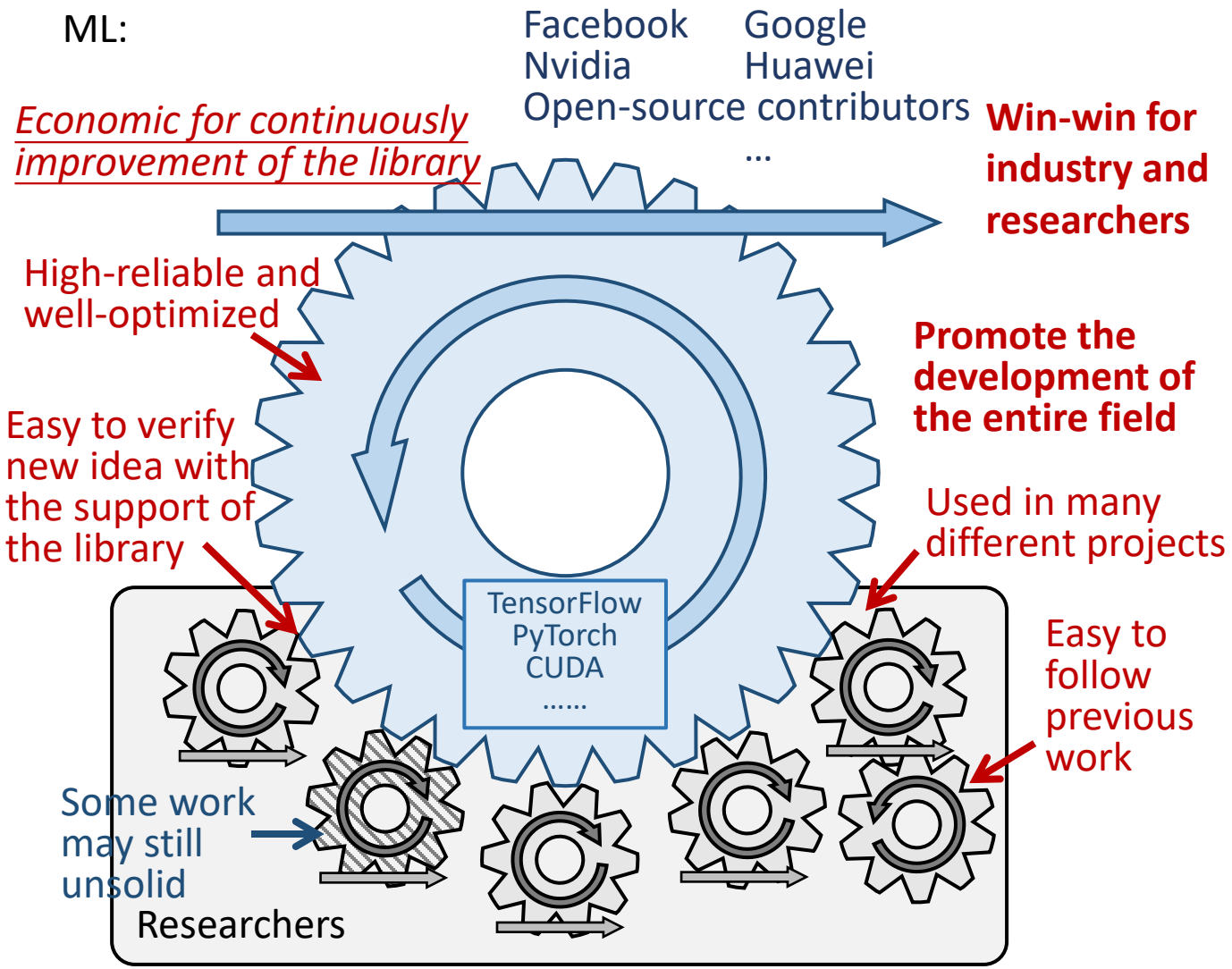
Uneconomic to improve code quality

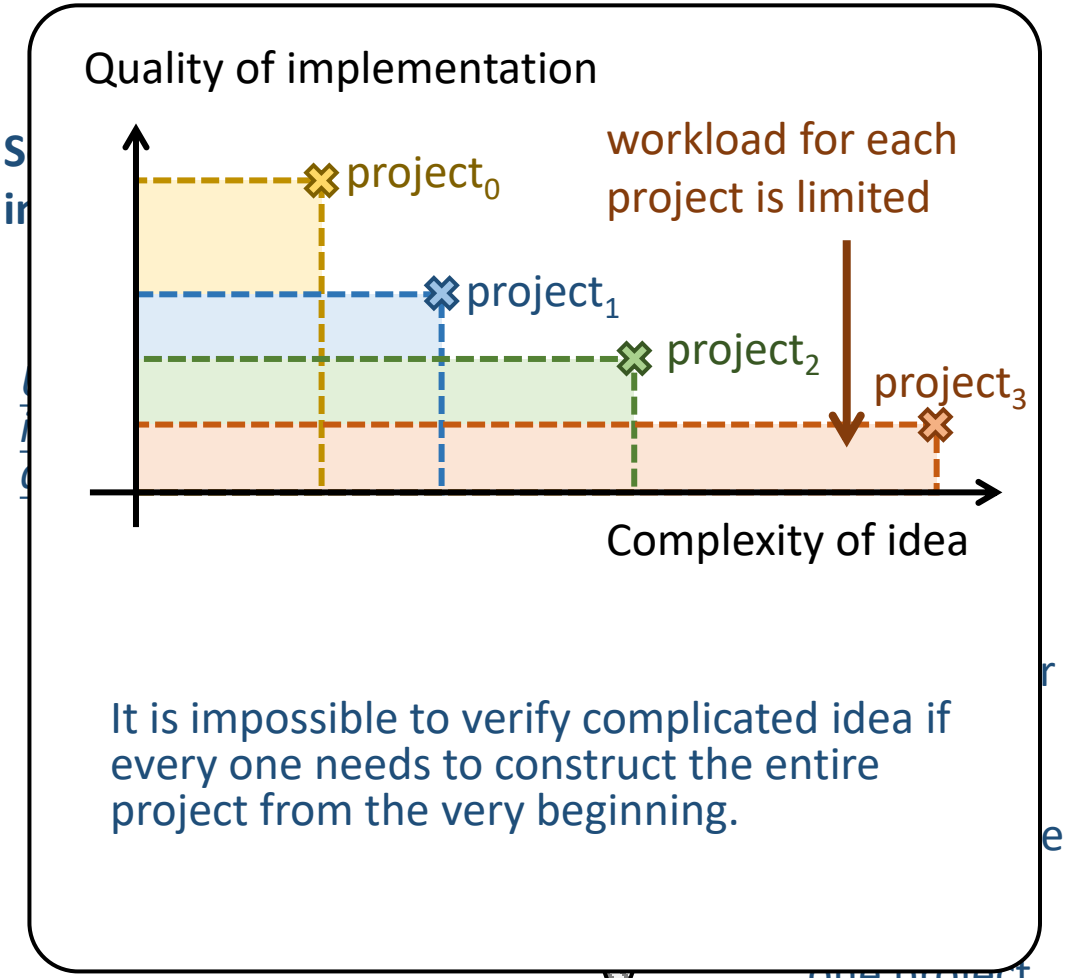
Low quality code due to the lack of opportunity for continuous optimization

Meaningless reconstruction for every beginner

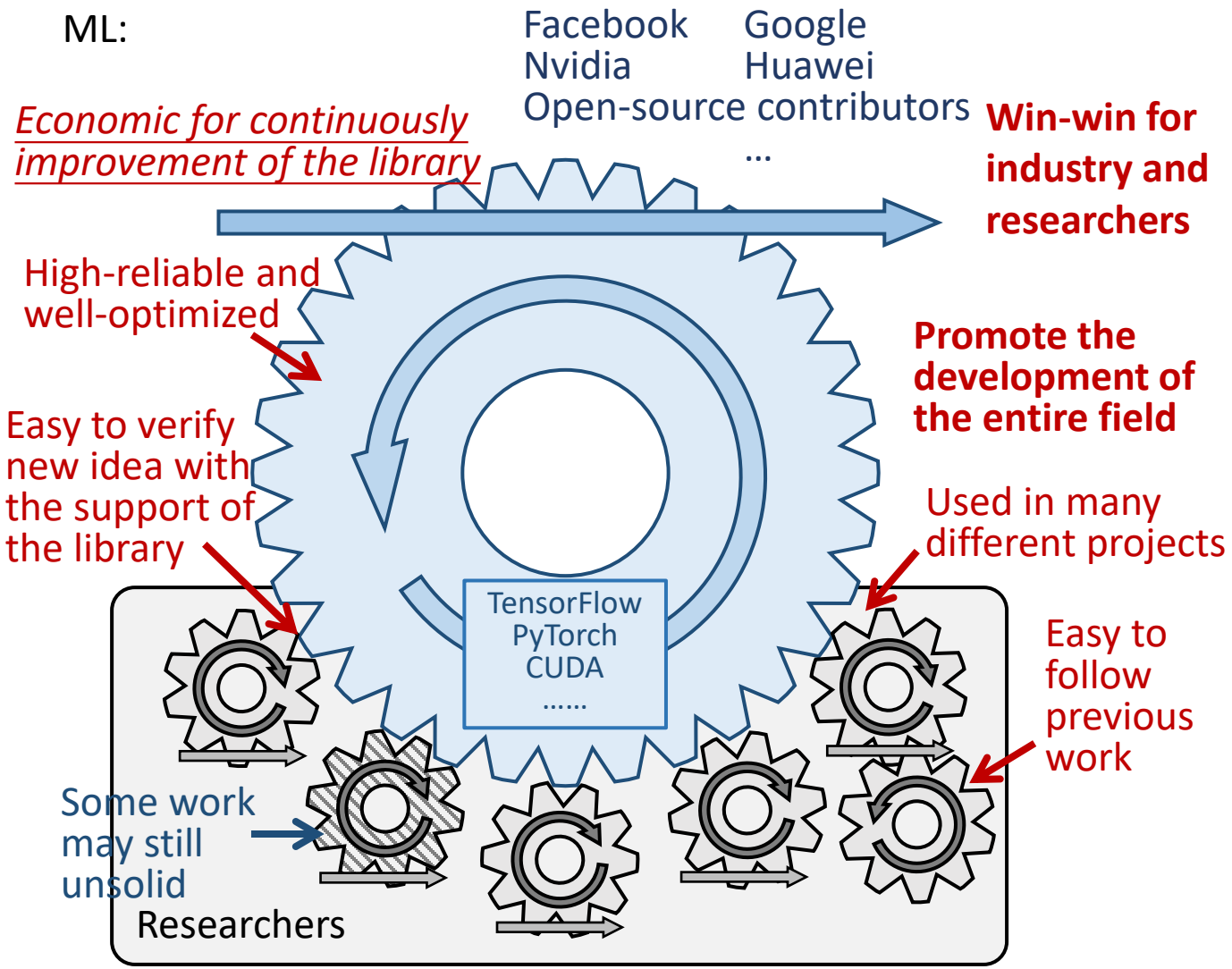
Most of code would be used in only one project

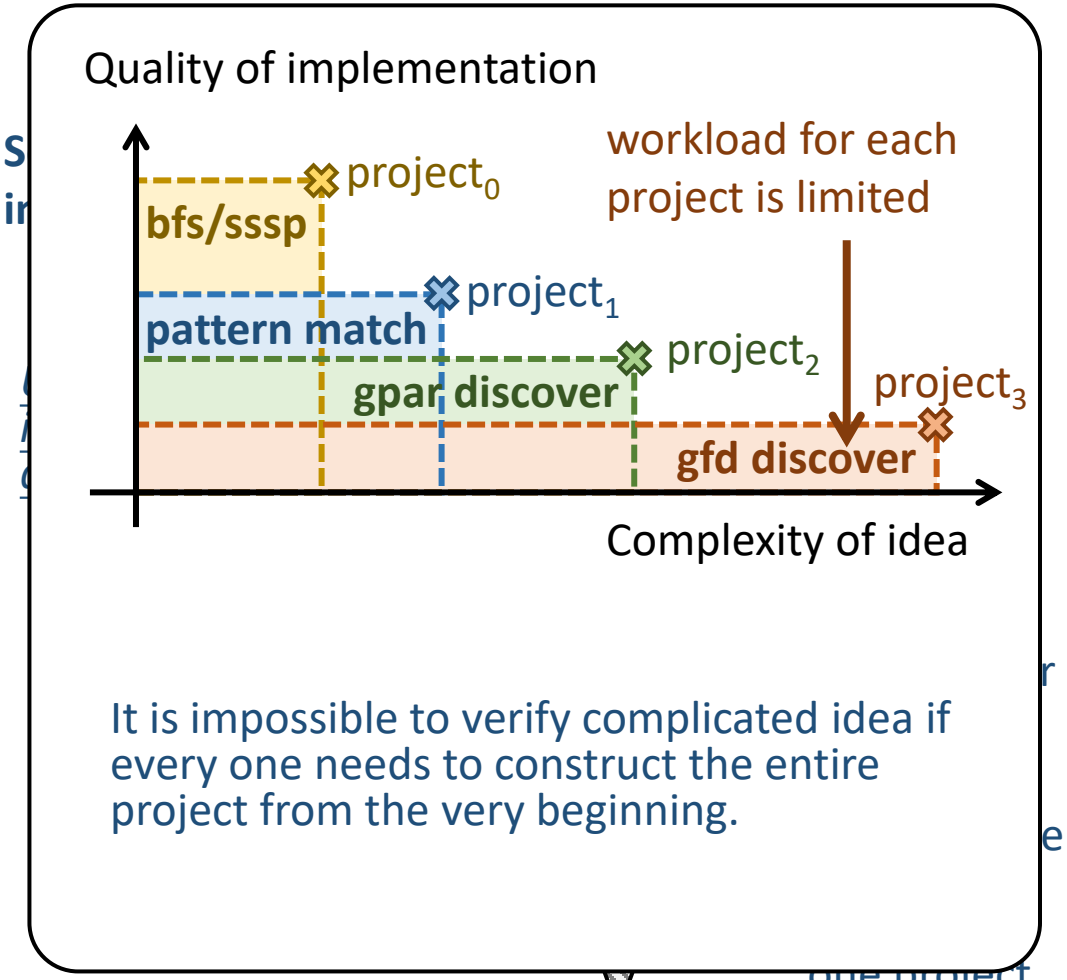
GUNDAM, provides an efficient, sustainable, industrialized development model of this field



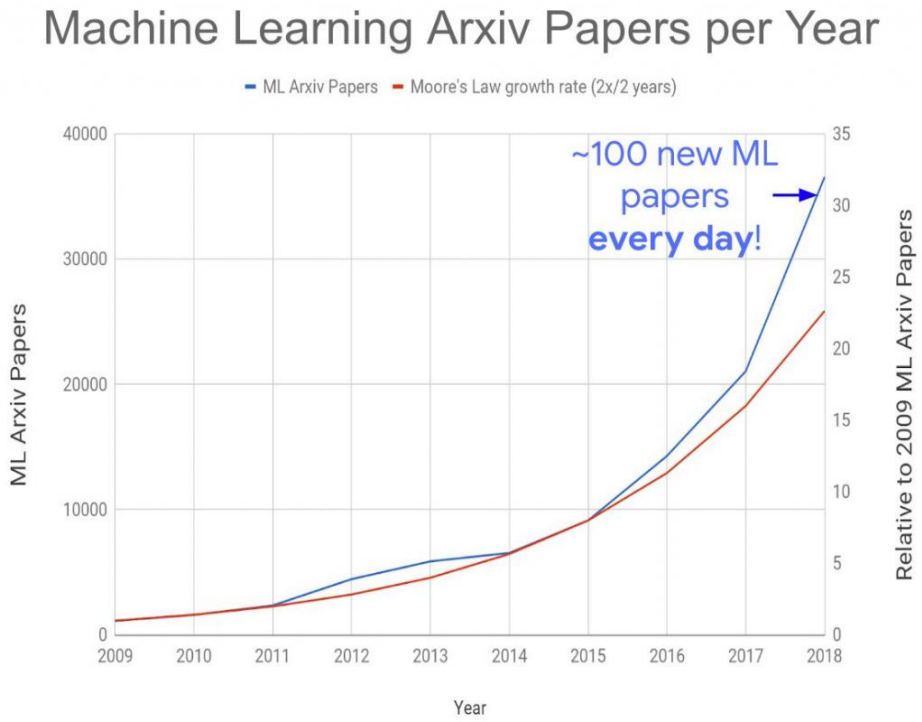


GUNDAM, provides an efficient, sustainable, industrialized development model of this field





GUNDAM, provides an efficient, sustainable, industrialized development model of this field



More than 90k paper are published in ML this year, even if 99.9% of them are rubbish, there still a lot of meaningful work.

Researchers

Tens of papers published each year

Lack of open-sourced code make most of the works hard for other to follow.

Low quality code even make it hard *for ourselves to follow previous works with the source code.*

Only works on toy examples, *scale matters.*

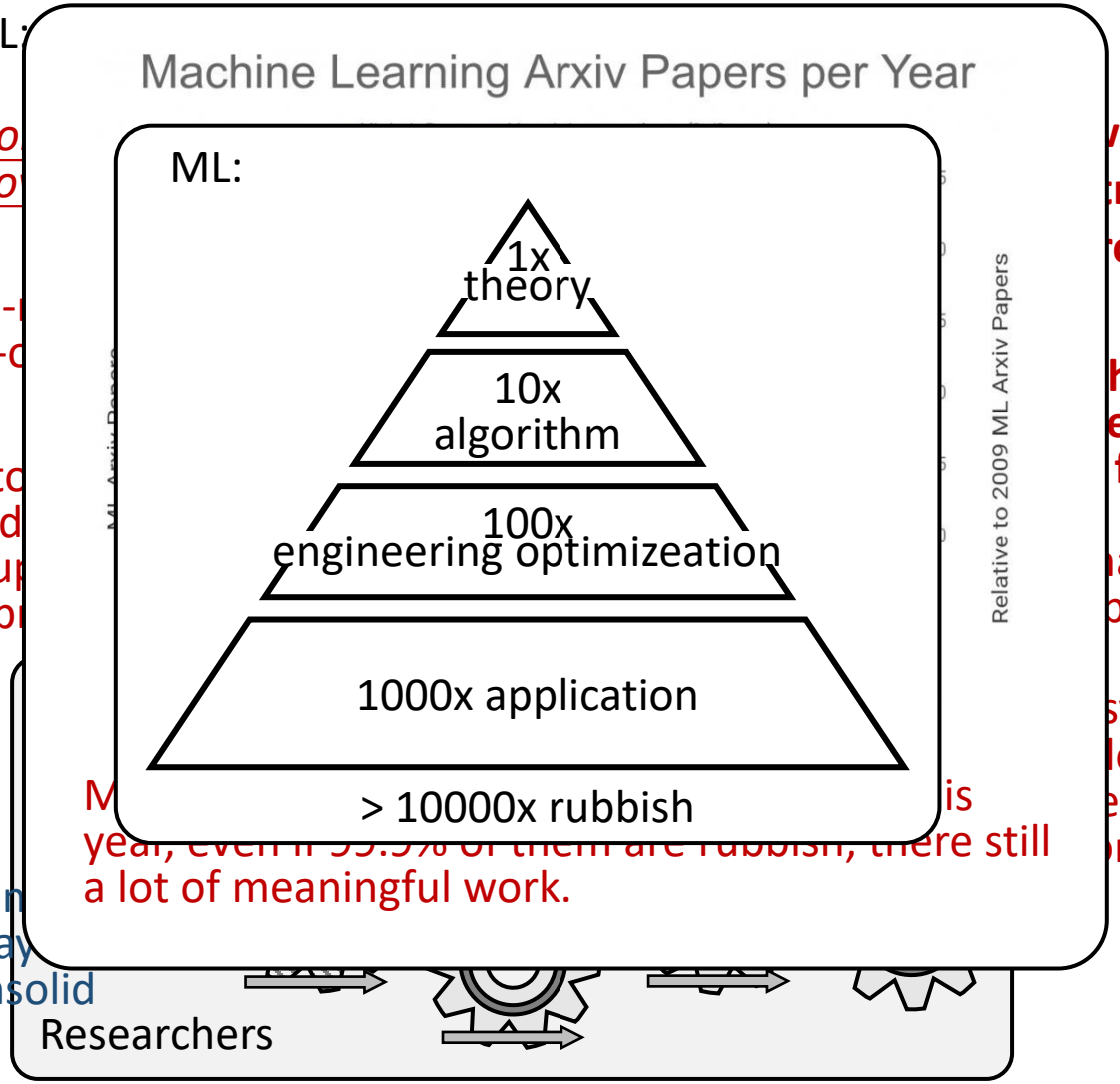
Some basic problems have not been discussed.

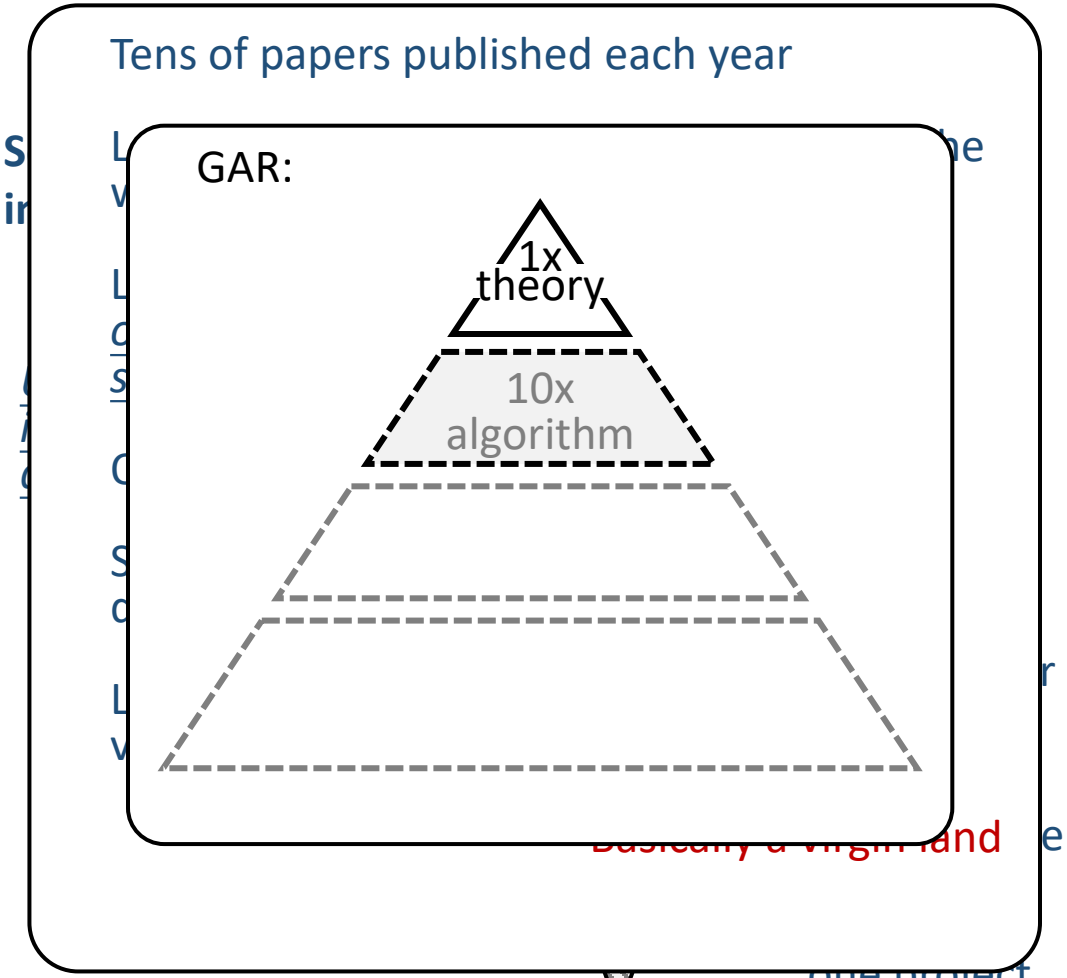
Lots of trivial ideas cannot find people to verify.

Basically a virgin land

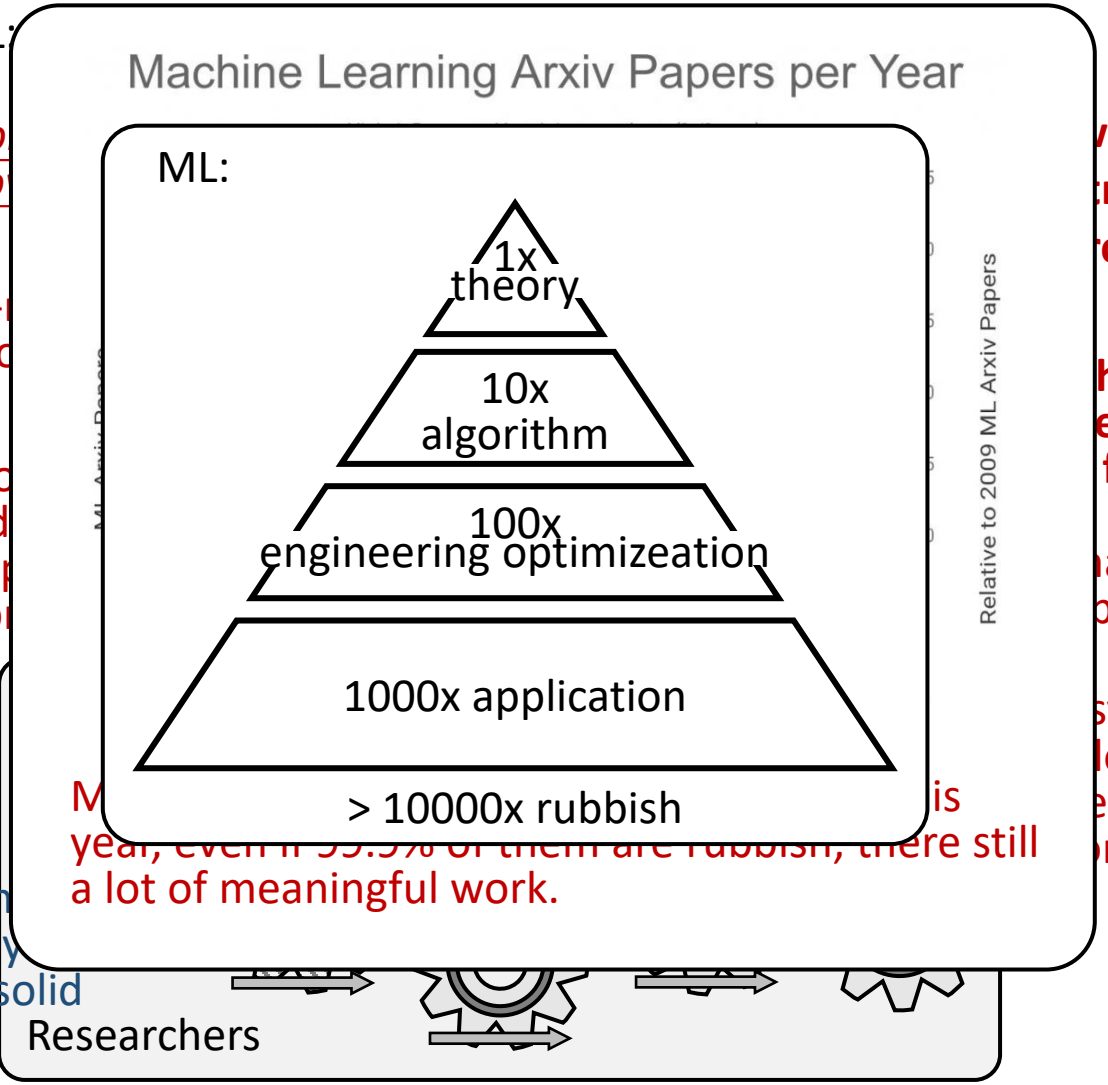
把蛋糕做大符合本组长期利益

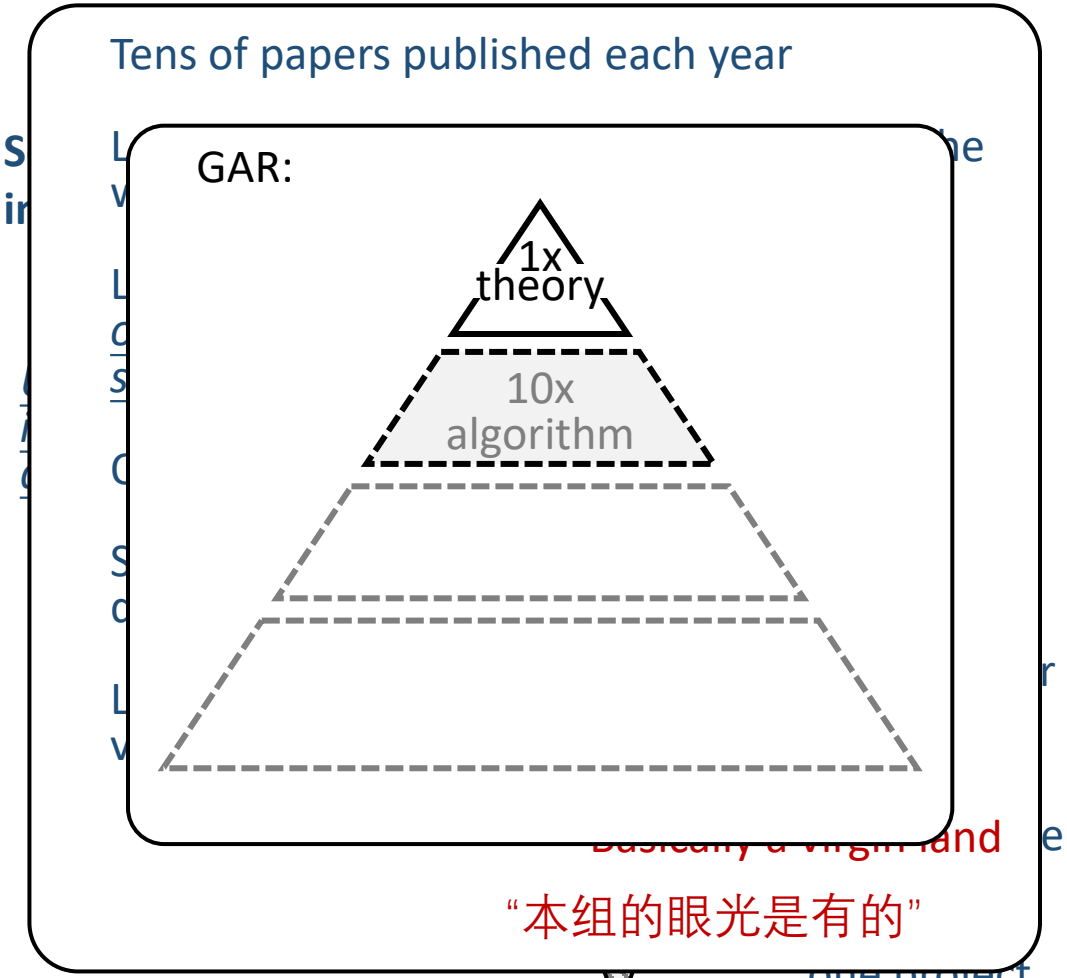
GUNDAM, provides an efficient, sustainable, industrialized development model of this field



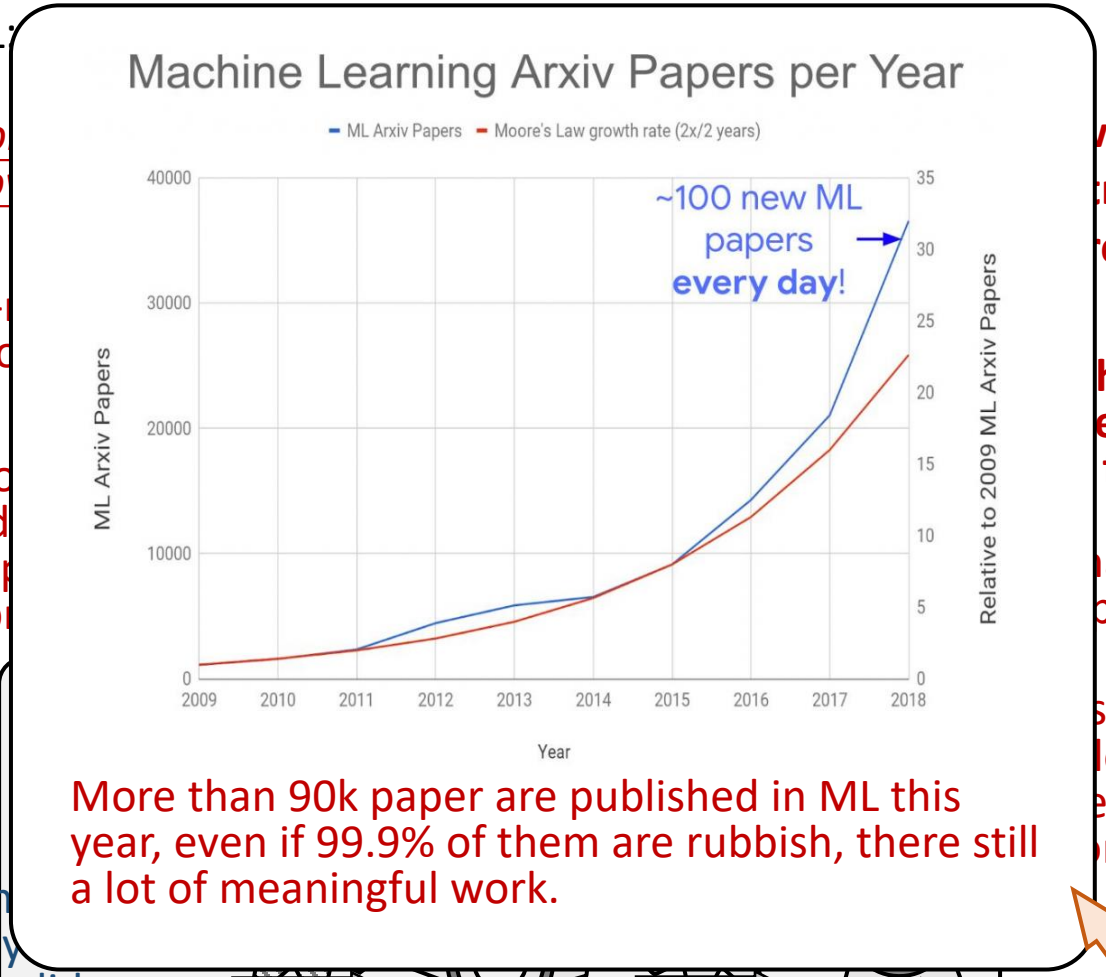


GUNDAM, provides an efficient, sustainable, industrialized development model of this field





GUNDAM, provides an efficient, sustainable, industrialized development model of this field



More than 90k paper are published in ML this year, even if 99.9% of them are rubbish, there still a lot of meaningful work.

Researchers

Tens of papers published each year

Lack of open-sourced code make most of the works hard for other to follow.

Low quality code even make it hard *for ourselves to follow previous works with the source code.*

Only works on toy examples, *scale matters.*

Some basic problems have not been discussed.

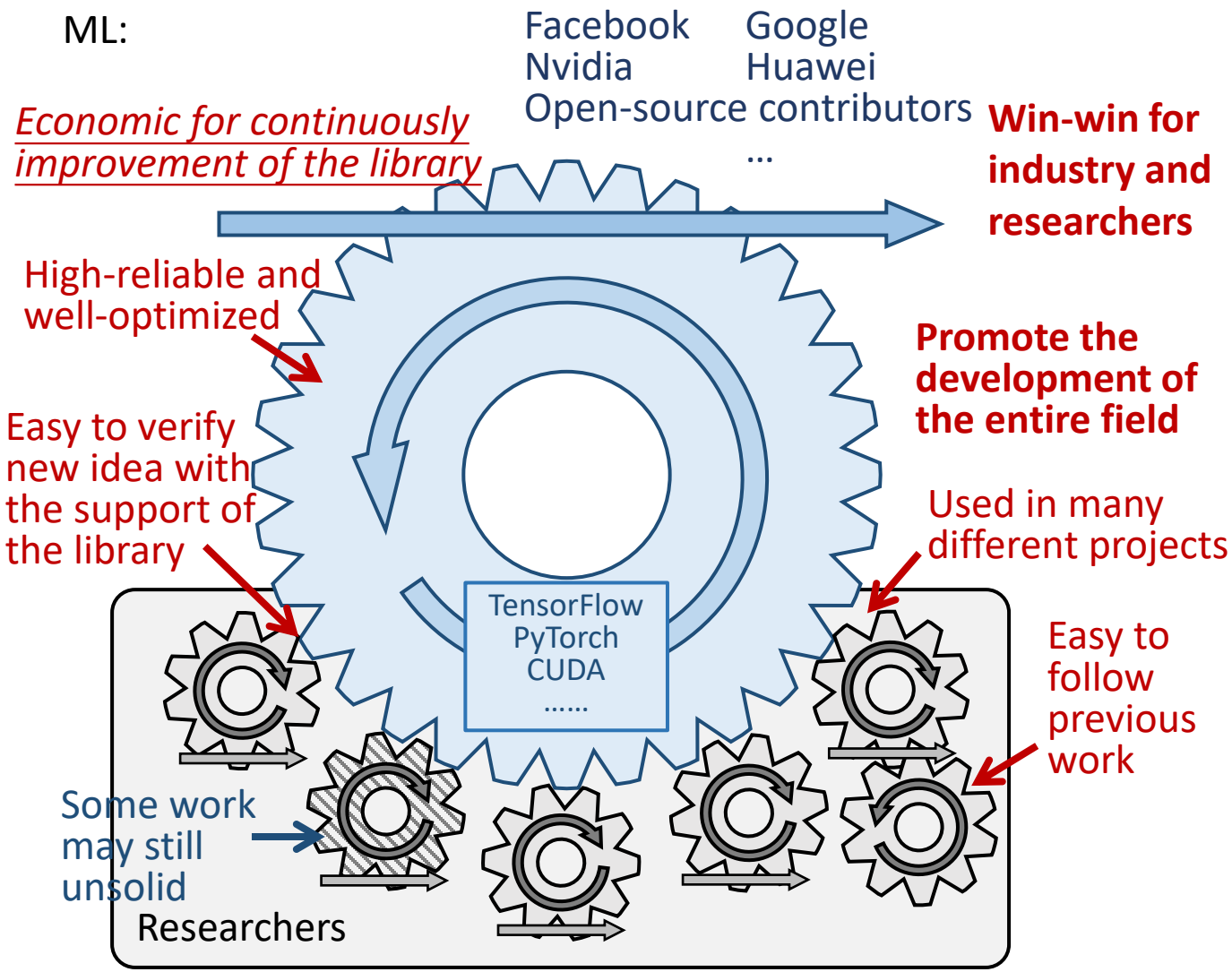
Lots of trivial ideas cannot find people to verify.

Basically a virgin land

New researchers & Investors

Which side would they chose?

GUNDAM, provides an efficient, sustainable, industrialized development model of this field



A vicious circle, the expected AlextNet-like success ***would never come!***

For deep learning, the development of CUDA and GPU provides a solid foundation for the success of AlexNet.

CUDA and GPU *are not brought by the inner logic of the development of deep learning.*

We cannot sit and wait for the such a "*deus ex machina*" happens in our field!

GUNDAM, provides an efficient, sustainable, industrialized development model

在无法取得业界支持的情况下，利用教学场景构建高效的、可持续的工业化发展模式，以推动本领域持续发展

ML:

Facebook Google
Nvidia Huawei
Open-source contributors
…

Economic for continuously improvement of the library

Win-win for industry and researchers

High-reliable and well-optimized

Promote the development of the entire field

Easy to verify new idea with the support of the library

Used in many different projects

TensorFlow
PyTorch
CUDA
……

Easy to follow previous work

Some work may still unsolid

Researchers

GAR:

参与迭代优化、实现细节功能，迅速提升编程能力

硕士

（找工作有更大的竞争力？）

GUNDAM

博士

设计开发核心算法&架构

博士

更快、更可靠地验证想法

解放生产力

- 新生可快速提升编程能力、提高生产力
- 一个人的工作能帮助其他所有人的工作

改善师生关系

- “给实验室搬砖”与“找工作”不再对立，师生具有更大的共同利益
- 新生可以快速上手、接手前人工作，降低学生由于太出活而毕不了业的可能性

多方共赢

GUNDAM, provides an efficient, sustainable, industrialized development model

在无法取得业界支持的情况下，利用教学场景构建高效的、可持续的工业化发展模式，以推动本领域持续发展

ML:

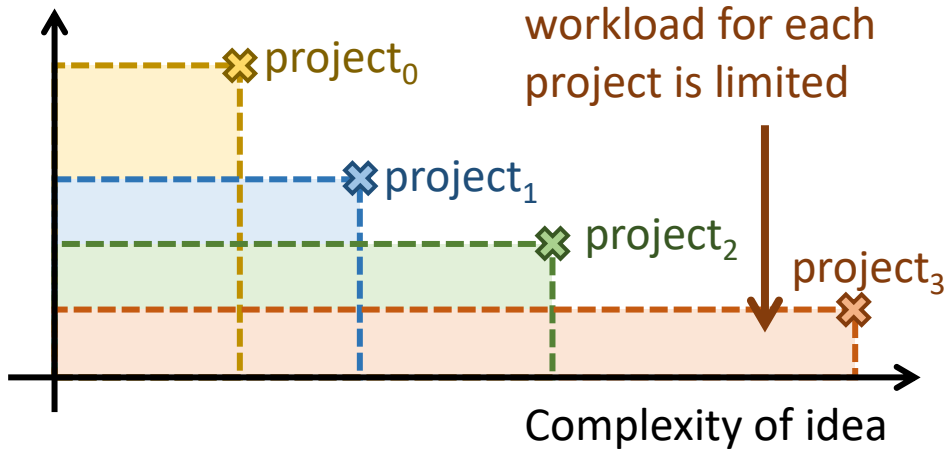
Economic for cont
improvement of t

High-reliable and well-optimized

Easy to verify new idea with the support of the library

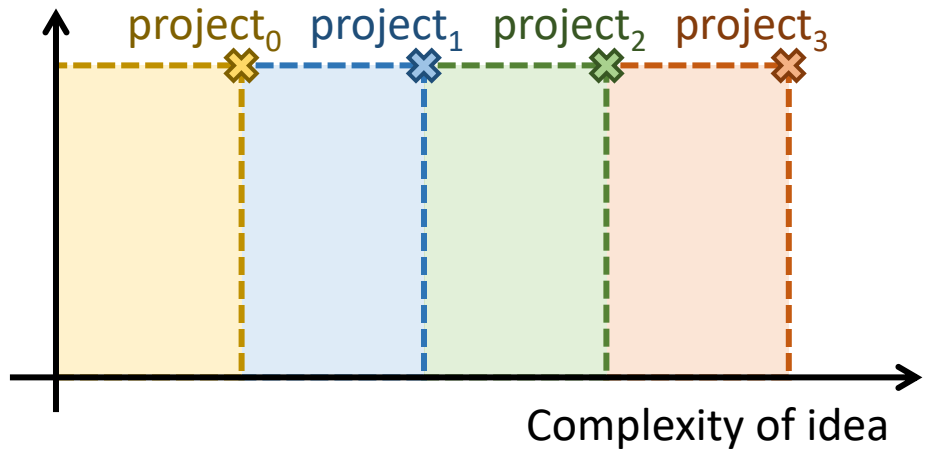
Some work may still unsolid

Researchers



It is impossible to verify complicated idea if every one needs to construct the entire project from the very beginning.

Without gundam



It is possible to verify complicated idea based on previous work!

With gundam

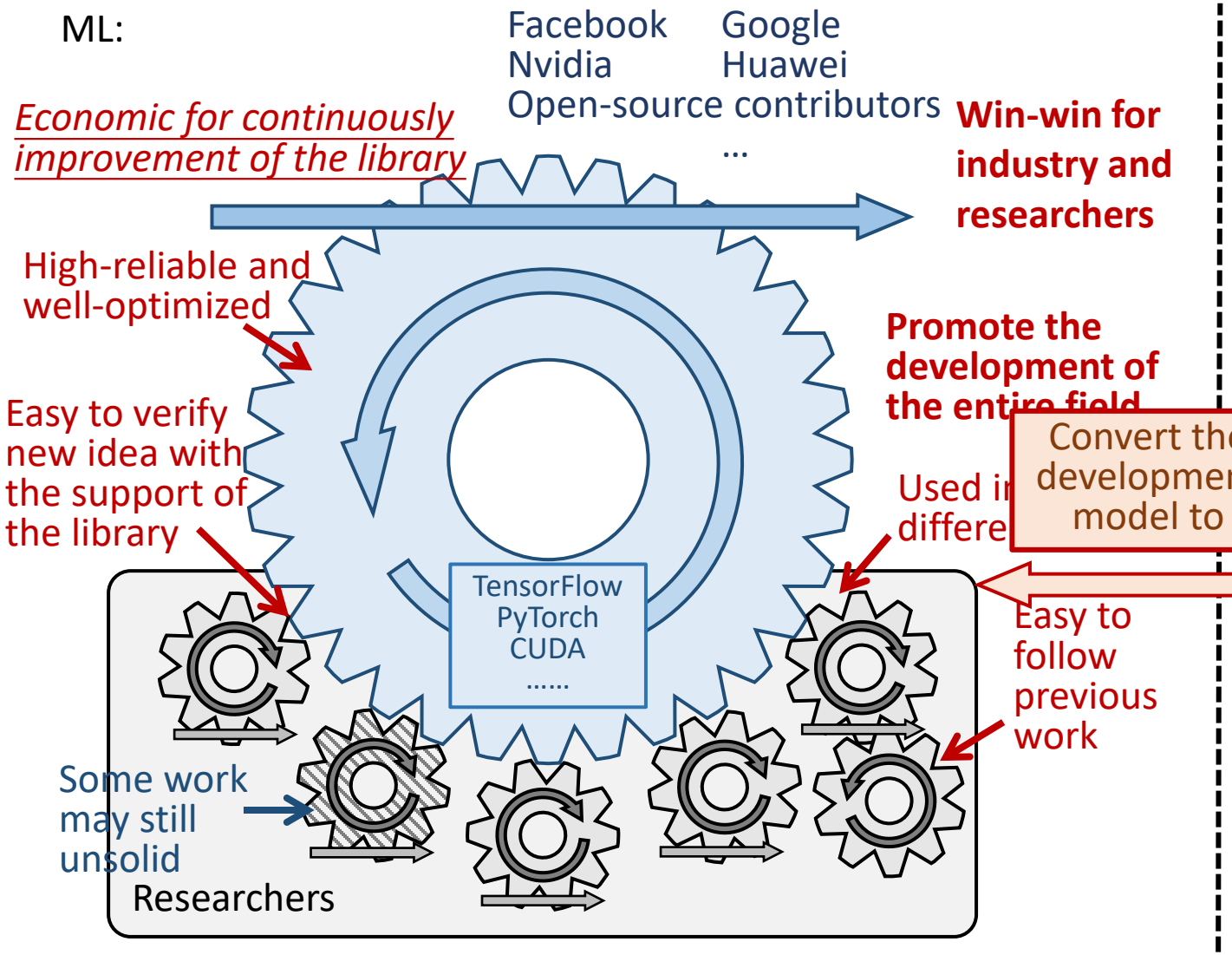
- 新生可以快速上手、接手前人工作，降低学生由于太出活而毕不了业的可能性

多方共赢

相较于Pytorch/Tensorflow，更应当贴近matlab的发展模式

GUNDAM, provides an efficient, sustainable, industrialized development model

在无法取得业界支持的情况下，利用教学场景构建高效的、可持续的工业化发展模式，以推动本领域持续发展



A virtuous circle, the expected AlextNet-like success *would come at some time!*

With such a success, can earn the support of the industry and more developer for the public library.

There must be such a library and it needs to be ours!

- 新生可以快速上手、接手前人工作，降低学生由于太出活而毕不了业的可能性

多方共赢

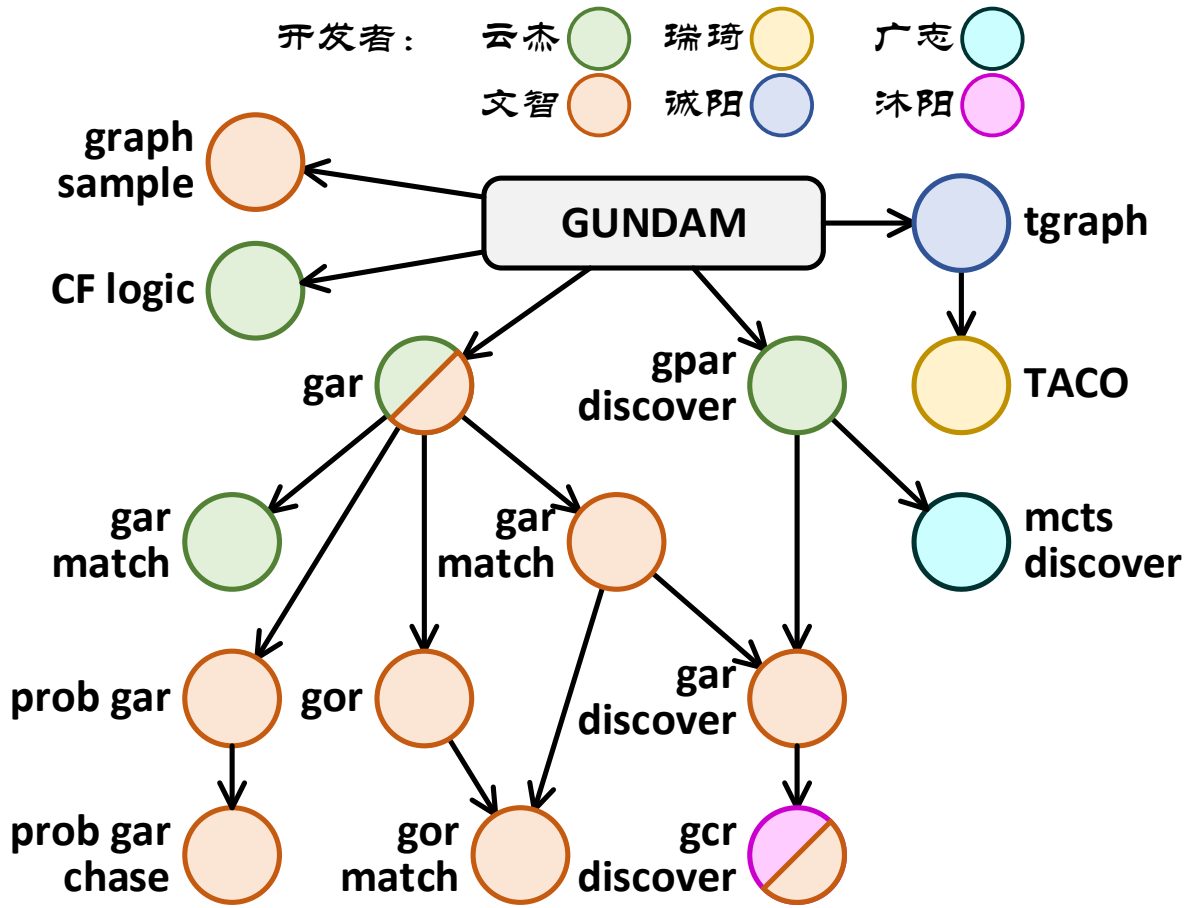
GUNDAM 现状

- 性能最好、通用性最强、设计理念最先进、最符合本组需求的图计算基础库
  - 各项均没有“之一”
- 已有一定积累
  - 可基于已有代码简单拓展新的工作
  - 已有工作均有可靠的baseline
- **产出 > 投入**

- 当前特性
  - 统一接口
  - 支持多种图
  - 静态类型 & 编译时优化

GUNDAM 现状

Works accumulated so far:



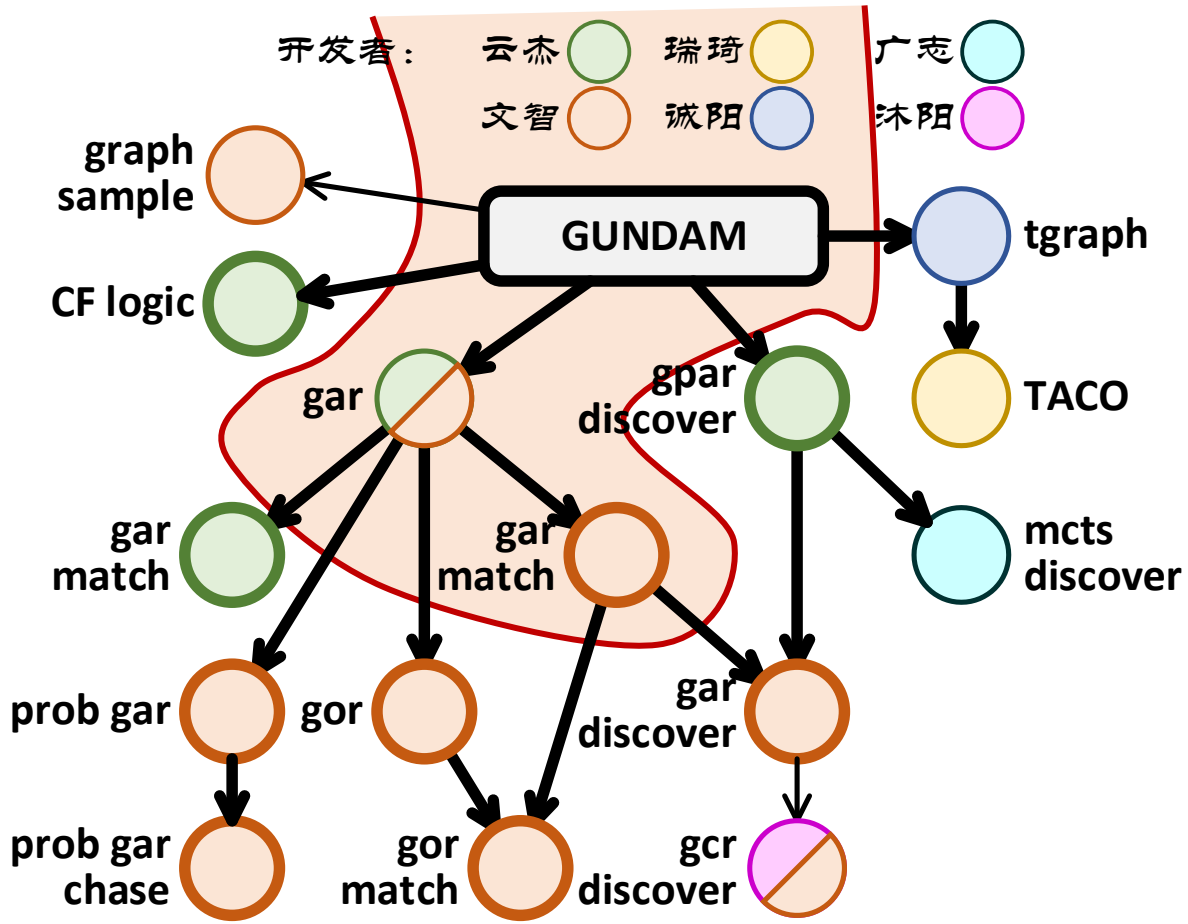
- Useful tools:
  - Gar tools:
    - Gar support
    - Gar imply
    - Gar analysis
    - Same gar
    - ……
  - Graph tools:
    - Same pattern
    - Connected
    - Diameter
    - ……

- Algorithm
  - Graph Iso
    - DagDp
    - VF2
    - ……
  - Bfs
  - Dfs
  - TopoSort
  - …

GUNDAM 现状

Works accumulated so far:

Bug free zone

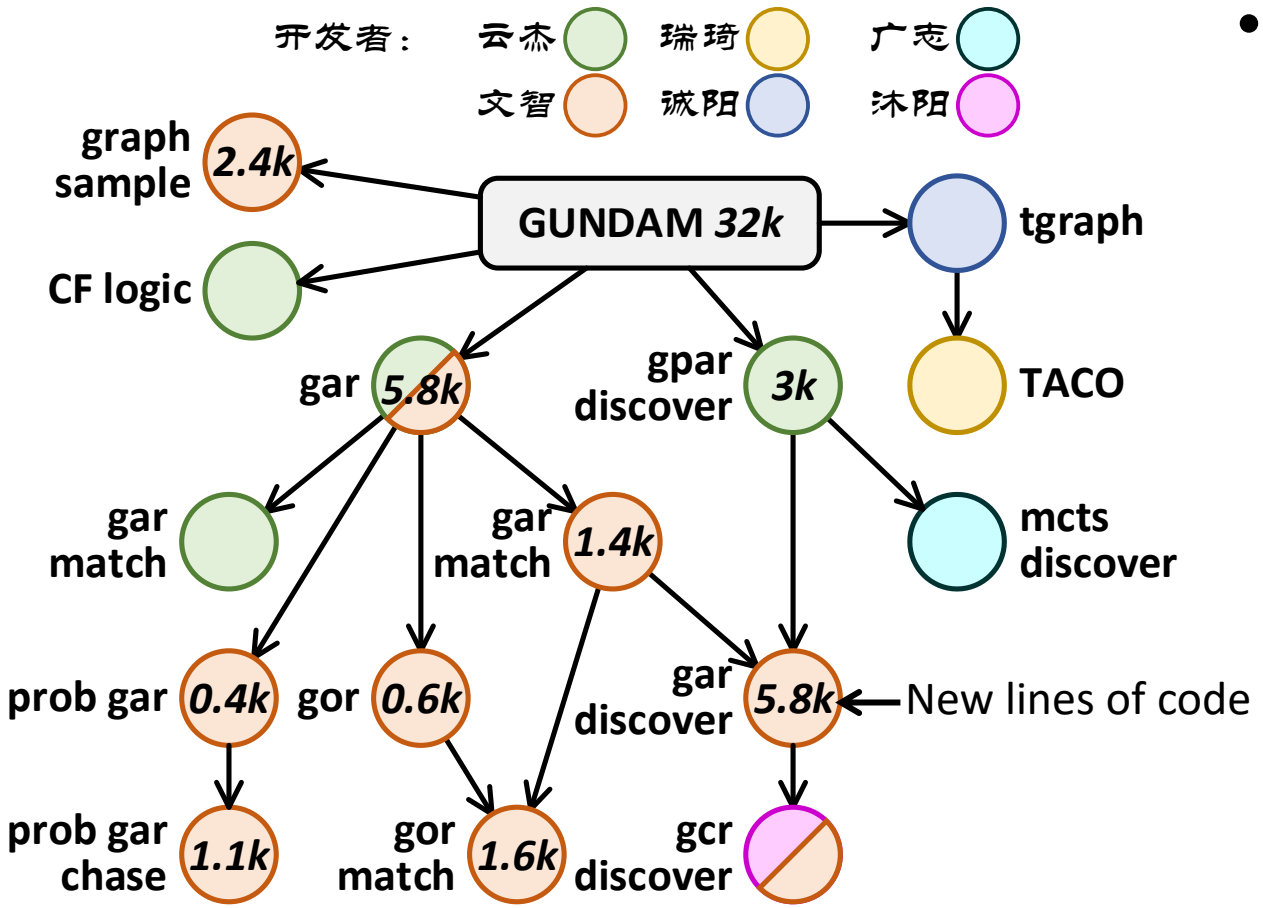


- Useful tools:
  - Gar tools:
    - Gar support
    - Gar imply
    - Gar analysis
    - Same gar
    - ……
  - Graph tools:
    - Same pattern
    - Connected
    - Diameter
    - ……

- Algorithm
  - Graph Iso
    - DagDp
    - VF2
    - ……
  - Bfs
  - Dfs
  - TopoSort
  - …

GUNDAM 现状

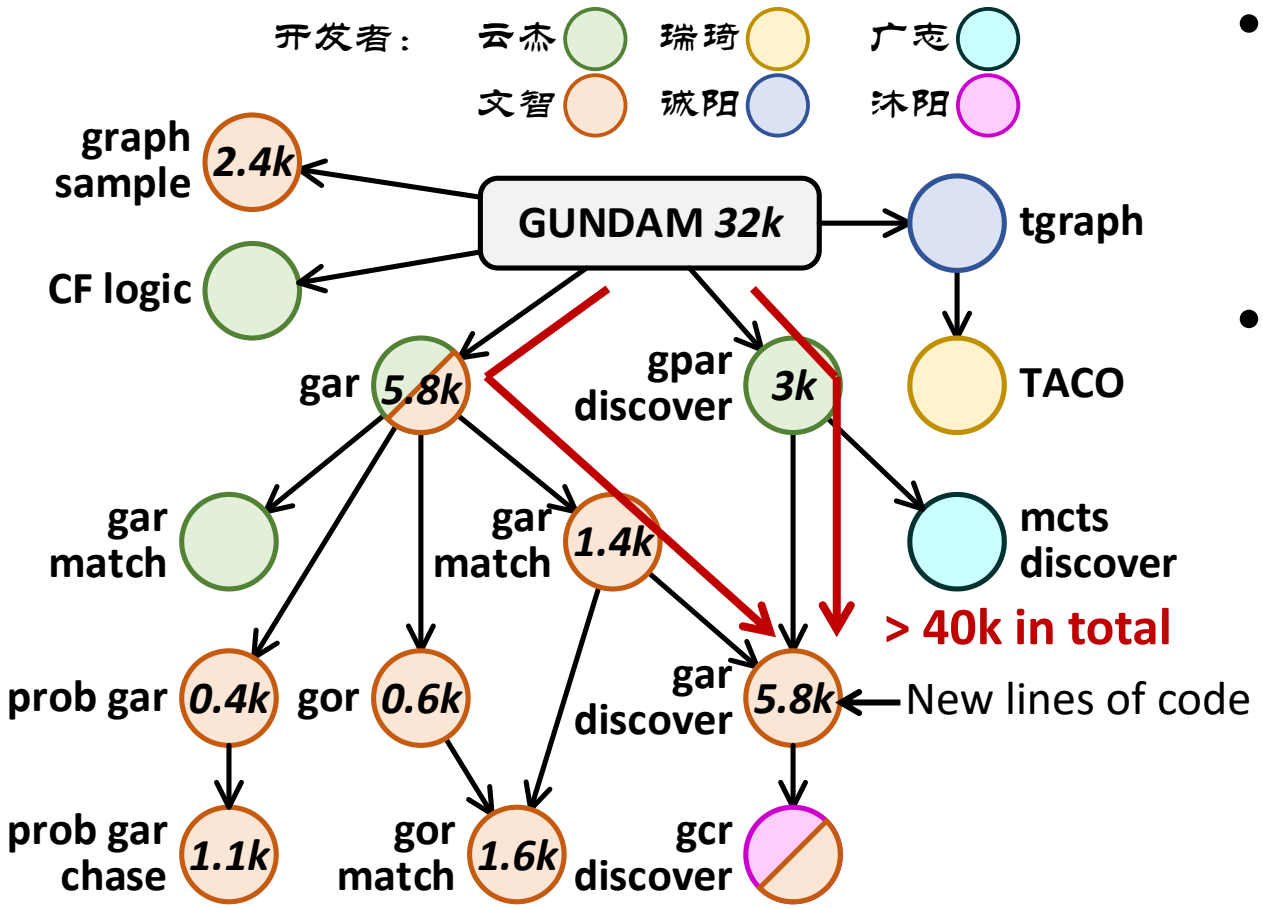
Works accumulated so far:



- 基于现有积累，可以使用相对较少的代码行数开发新的项目

GUNDAM 现状

Works accumulated so far:



- 基于现有积累，可以使用相对较少的代码行数开发新的项目

- “科学技术是第一生产力”
  - 绝对劳动时间的剥削是有极限的
    - 996 -> 007: **2.3**倍生产力提升
  - 只有技术进步能持续提升生产力
    - 40k -> 5.8k: **6.9**倍生产力提升
    - 随着库的各种功能不断完善，实现相同功能所需代码行数会持续降低

GUNDAM 现状

- Other useful resources relate to GUNDAM
  - GraphShow

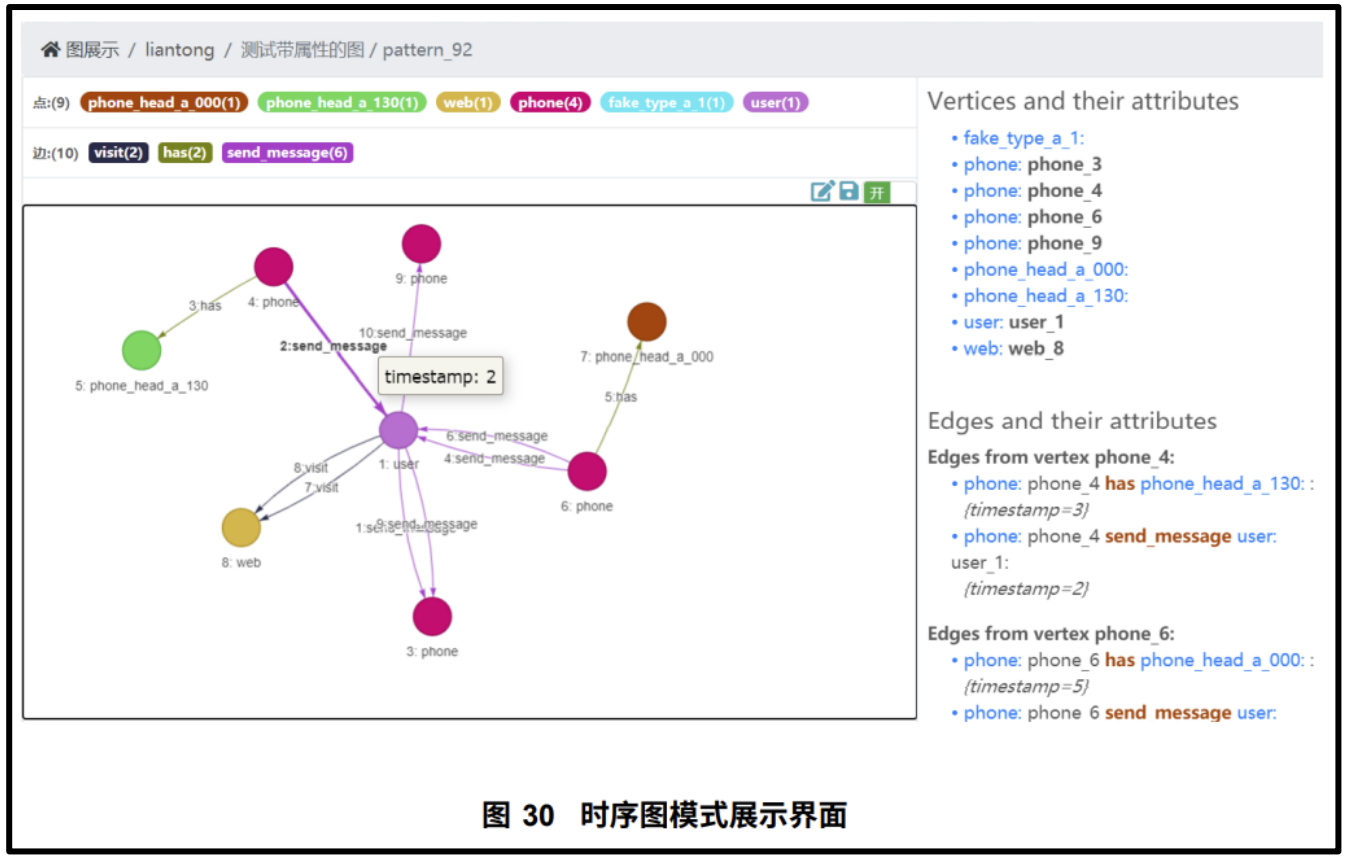


图 30 时序图模式展示界面

via 刘诚阳硕士论文

GUNDAM 现状

- Other useful resources relate to GUNDAM
  - GraphShow

图 31　时序图关联规则发现工具界面

GUNDAM 现状

- Other useful resources relate to GUNDAM
  - GraphShow
  - GraphBuilder
    - MySQL -> Neo4j -> CSV

```yaml
Relations:
#  - 边 type: 重命名边 type
  - CALL: Call
  - VISIT: Visit
  - SEND MESSAGE: Message

Nodes:
#  - 节点 label: [属性 1,属性 2…]
  - User: [UserID]
  - Web: [Web_name,Access_type]
  - Phone: [Phone_number,Phone_head,Phone_length]
  - Fake: [Fake_label]

#这里配置希望导出属性的节点,以及需要的属性 ,不写的默认不导出属性
# 填写的属性名需要对应Neo4j图内存在的属性名，同时它也会是全展开图生成的新节点label
#比如导出Phone的Phone_head属性
#Phone: [Phone_head]
```

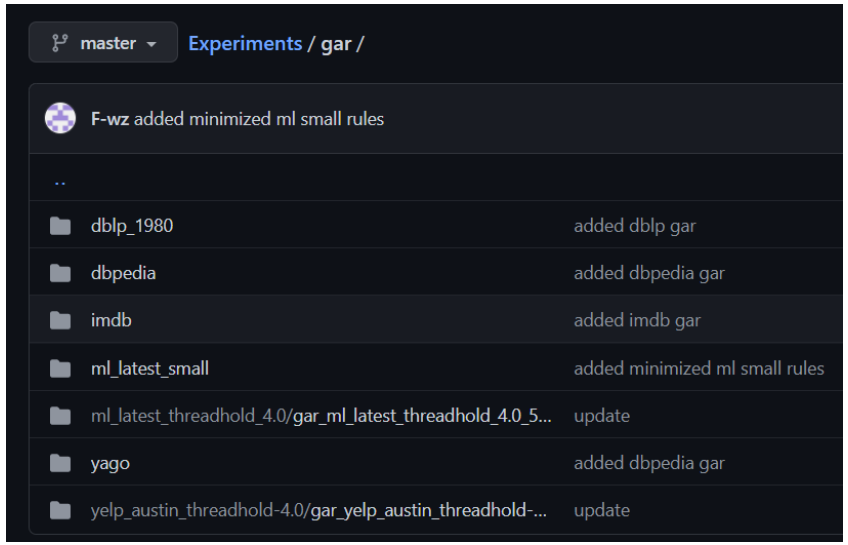
export.yaml

```yaml
#注意:
#1.  文件中冒号，逗号等均为英文字符，冒号后有空格
#2. 填写文件名时需要包含后缀
Nodes:
#  - MySql 表名: [
#    节点 Label 名: [
#      属性名 1:  重命名,
#      属性名 2:  重命名,
#      ...   ]
#  ]
  - user: [
    User: [
      uid: UserID]
  ]
  - voice_dura: [
    Phone:[
      opp_num: Phone_number,
      opp_head: Phone_head,
      opp_len: Phone_length ]
  ]
```

import.yaml

GUNDAM 现状

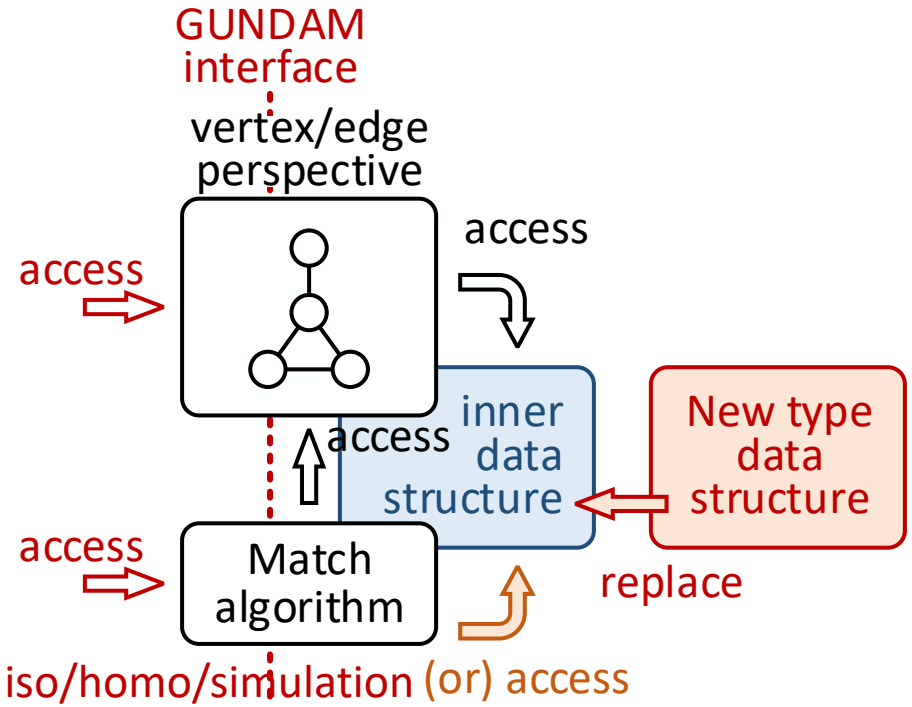
- Other useful resources relate to GUNDAM
  - GraphShow
  - GraphBuilder
    - MySQL -> Neo4j -> CSV
  - Experiments
    - Discovered and evaluated gars accumulated so far

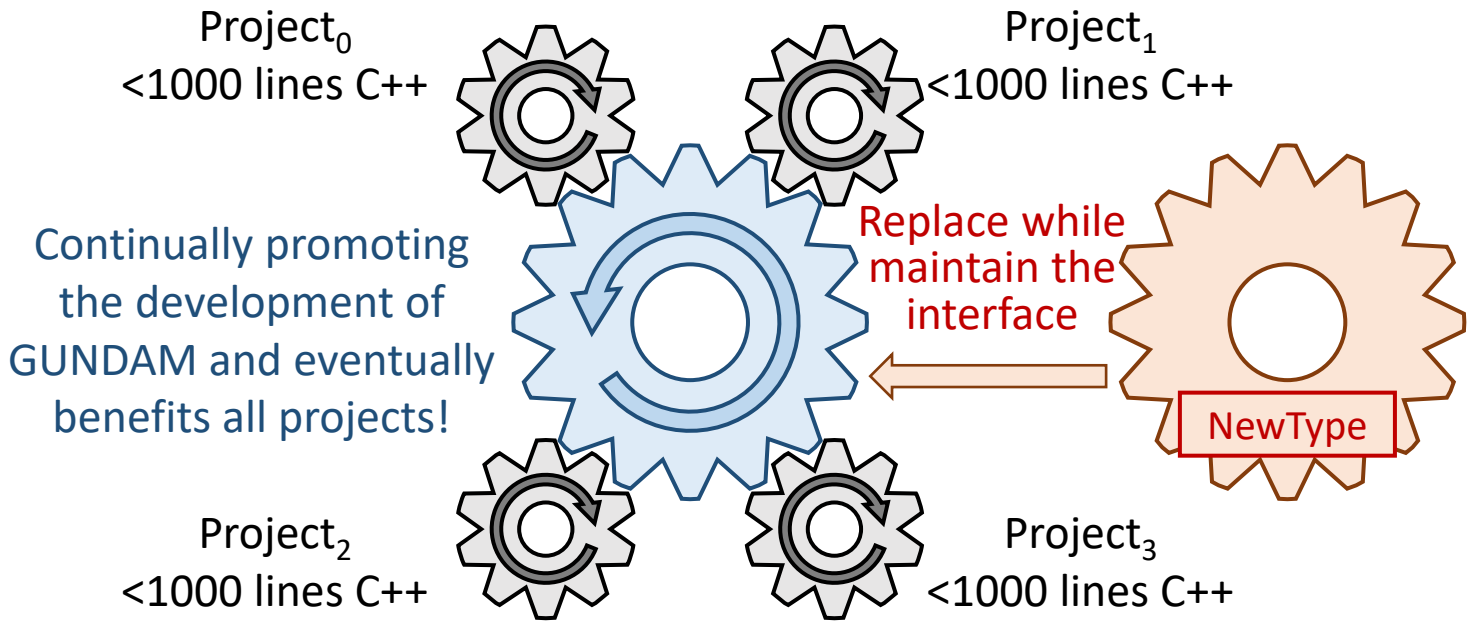


GUNDAM 现状

There must be such a library and it needs to be ours!

- *NewType* is under development, would inherit the interface in order to succeed all applications developed based on it so far.
  - The only risk is that the *NewType* does not perform that well.





Thanks!