

DCS292: 编译器构造实验(Compiler Construction)

Lab0: 开启“智能编译”

Yat Compiler Construction with AI

yatcc-ai.com

YatCC团队

中山大学 计算机学院

国家超级计算广州中心

2026.3.2

www.nscg-gz.cn



OUTLINE

目 录

中山大学计算机学院

School of Computer Science & Engineering

一、YatCC-AI介绍

二、整体实验设计

三、总结与展望

YatCC-AI? YatCC?

□ **YatCC-AI: Yat Compiler Construction with AI** [AI赋能]

□ **Yat Compilation Course: From University** [来自中大]

□ **Yet Another Tiny C Compiler: For Practical Training** [面向实践]

□ **Your AI Time Cool Compiler: Follow Technical Trend** [紧跟前沿]

□ **Young Architect Training, Coaching, Cultivating: Further Goal** [不止编译]

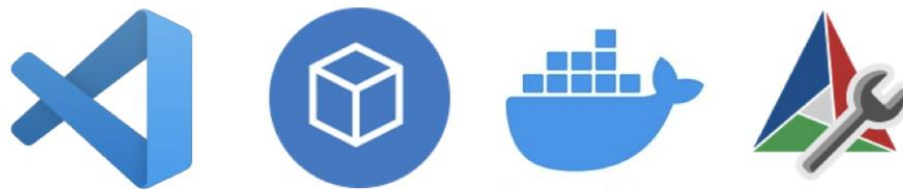
□ **Young, Active Team with Close Collaboration: Fantastic Team** [卓越团队]



演示 Demo



网页即开即用、AI全程赋能



YatCC-AI Hub



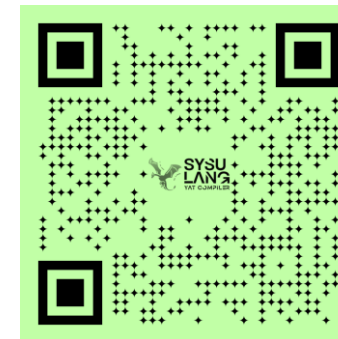
YAT COMPILER

Your AI Time Cool Compiler

Just open browser — no setup, no hassle!

Chat, Code, Create — rock your LLM agent!

- **课程网站:** yatcc-ai.com, [yatcc.github.io](https://github.com/yatcc/yatcc)
- **实验代码:** <https://github.com/arcsysu/YatCC>
- **实验文档:** <https://docs.yatcc-ai.com/#/>
- **视频教程:** <https://space.bilibili.com/3546650047941291>



Resource Hub

Course Archive

Catalog of past course information and teaching materials of 2025, 2024, 2023, 2022, 2021.

Reference Manual

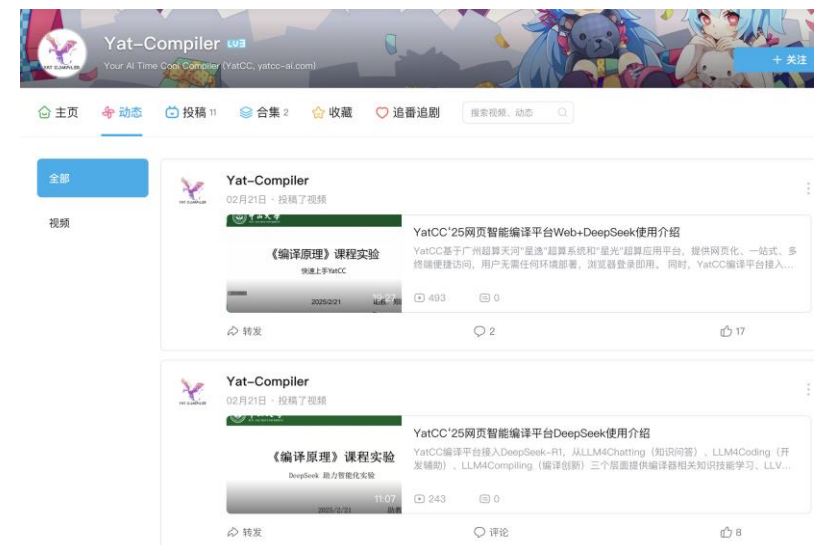
Documents that cover the complete process of experimentation and offer a wealth of practical examples.

Source Code

Detailed source code framework with 300+ commits, 200+ stars.

Video Tutorial

A pile of video clips providing step-by-step operational guidance.



OUTLINE

目 录

中山大学计算机学院

School of Computer Science & Engineering

一、YatCC-AI介绍

二、整体实验设计

三、总结与展望

□ 编译器是持续演进的核心基础软件

- 衔接上层应用程序和底层硬件架构
- 计算机生态一直在改变，带动编译器持续发展

PyTorch



编译器



□ LLVM是学习编译与开展编译工作的理想平台

- 发展逾20年，设计良好、易于上手、社区活跃
- 工业界和学术界需求广泛，但面临严峻人才缺口

人民日报 | 有品质的新闻

打开

推动国产基础软件加快发展（
创新谈）



人民日报 海生 2024年7月1日 00:00

方舟编译器

多端多语言，轻量低开销

拥有强大的基础软件，信息产业、数字经济
的“大厦”才能建得高。反之，如果缺少
强大的基础软件，其他软件和信息服务的发
展也将受到制约



□ 既有教学与业界存在一定偏差

- 强经典编译技术阶段，弱实用编译器全局实践
- 闭环小型教学项目 vs. 工业级编译基础设施
- AI应用对传统教学模式的冲击

基于 LLVM 的编译实践教学体系



<https://clang.llvm.org/>

https://clang.llvm.org/get_started.html

基于友好开发体验的LLVM编译实践教学

□ 打通课程教学和业界需求，促进产教融合

- 基于LLVM：平衡实验难度，理论结合实际
- 学习LLVM：了解复杂软件，衔接业界实践
- 对齐LLVM：参考标准实现，向业界前沿看齐

□ 定义一套现代化的标准实验环境

- 开箱即用，**全自动、零配置**的实验环境
- 提供完整的开发、测试、打包等**全流程**支持
- 降低上手难度，提高实验完成率

□ 向领域专家成长，or像领域专家一样探索

- 前沿研究或工程实践

S Y s Y 类C 编程语言



分阶段构建小型编译器

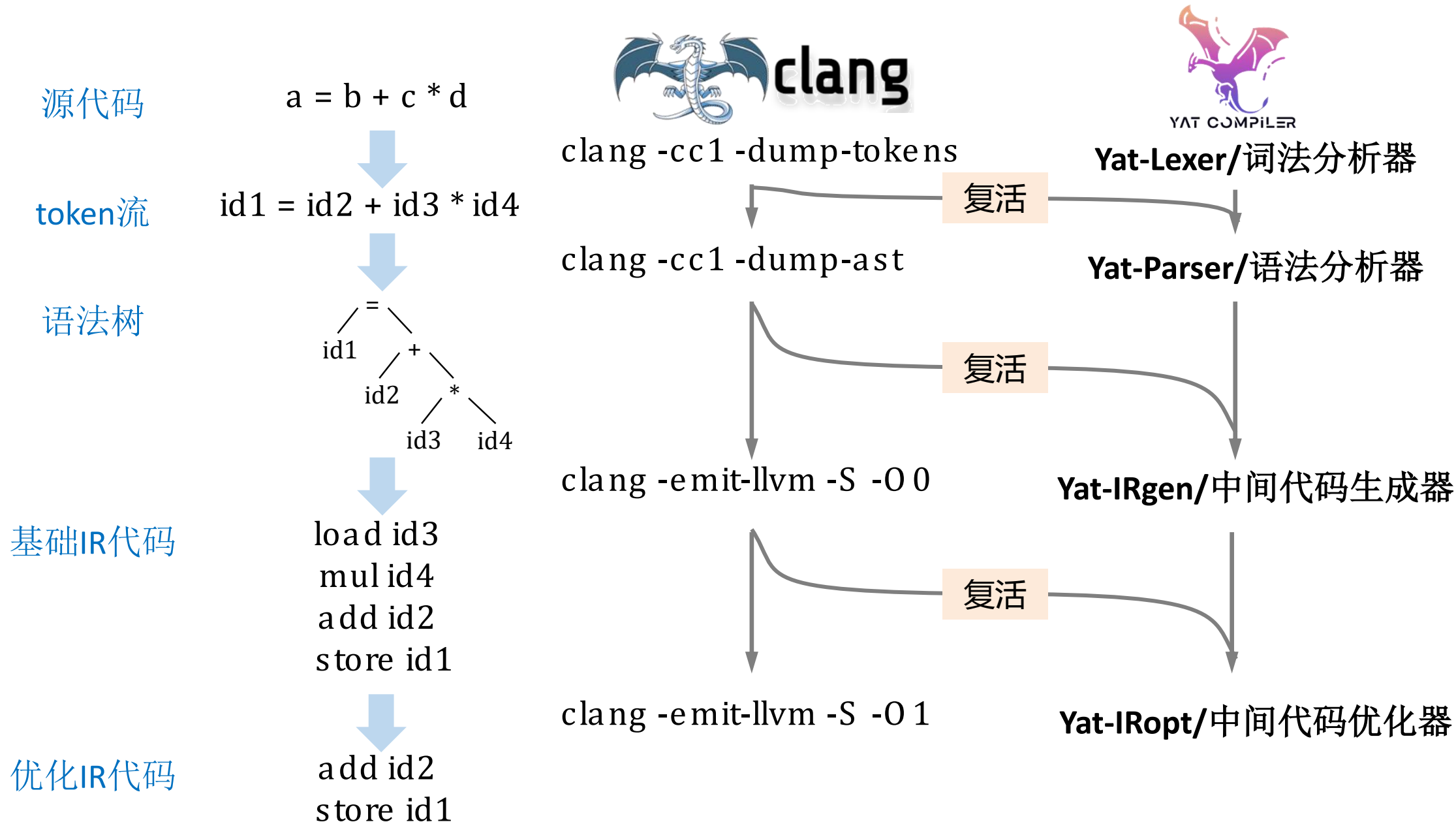


LLVM 编译基础设施



<https://compiler.educg.net/#/>

面向实践：基于LLVM开展实验



□ 一站式、零配置的云端开发环境

- 容器部署自动化，提供**开箱即用的云端实验环境**
- 完整设计项目框架，**全流程自动化，零命令行操作**
- 预置**良架构的基础示例代码**，实验难度可调节



YatCC-AI Hub



□ 分阶段推进，实验模块之间解耦

- 对应四个编译阶段，“**复活**”以隔绝前一阶段的影响



YAT COMPILER

Your A I T i m e C o o l C o m p i l e r

Just open browser — no setup, no hassle!

□ 集成单元测试，本地自动评测系统

- 以LLVM为参考答案，**自动运行测例、生成成绩单**：
本地实时反馈

Chat, Code, Create — rock your LLM agent!

LLM4Chatting

面向所有课

- 知识问答推理

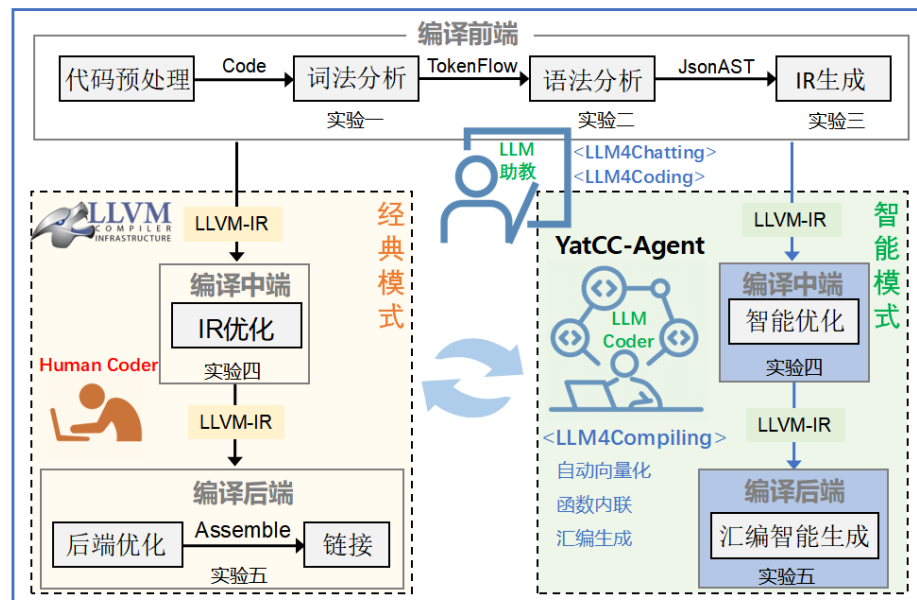
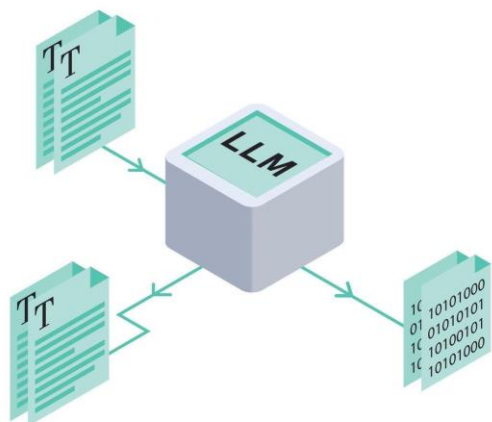
LLM4Coding

面向有编程实践的课

- 代码开发调试

LLM4Compiling

- 编译优化、汇编生成



YatCC-AI Hub



LLM Compiler

Meta LLM Compiler is a state-of-the-art LLM that builds upon Code Llama with improved performance for code optimization and compiler reasoning.

 huggingface.co

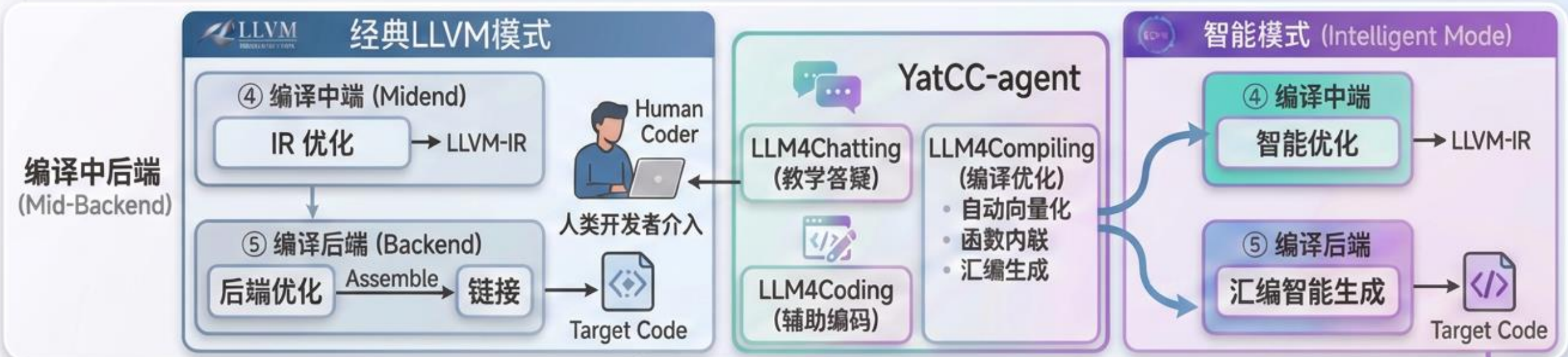


YatCC: AI-驱动的编译实践

Your AI Time, Cool Compiler



arcSYSu
国家超算广州中心
中山大学计算机学院



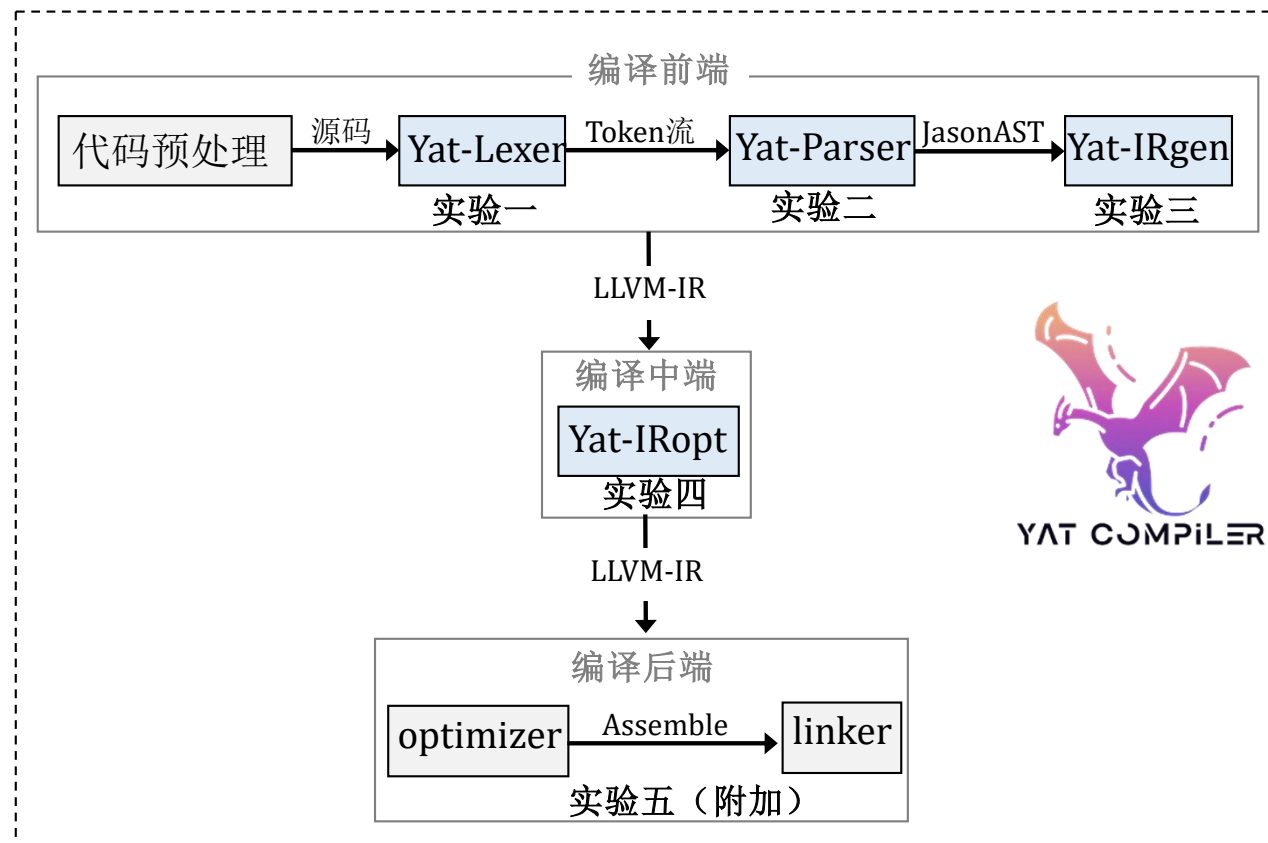
整体任务安排

实验分数构成:

- 最终测评得分
- 报告分+改进建议分

DDL slip day:

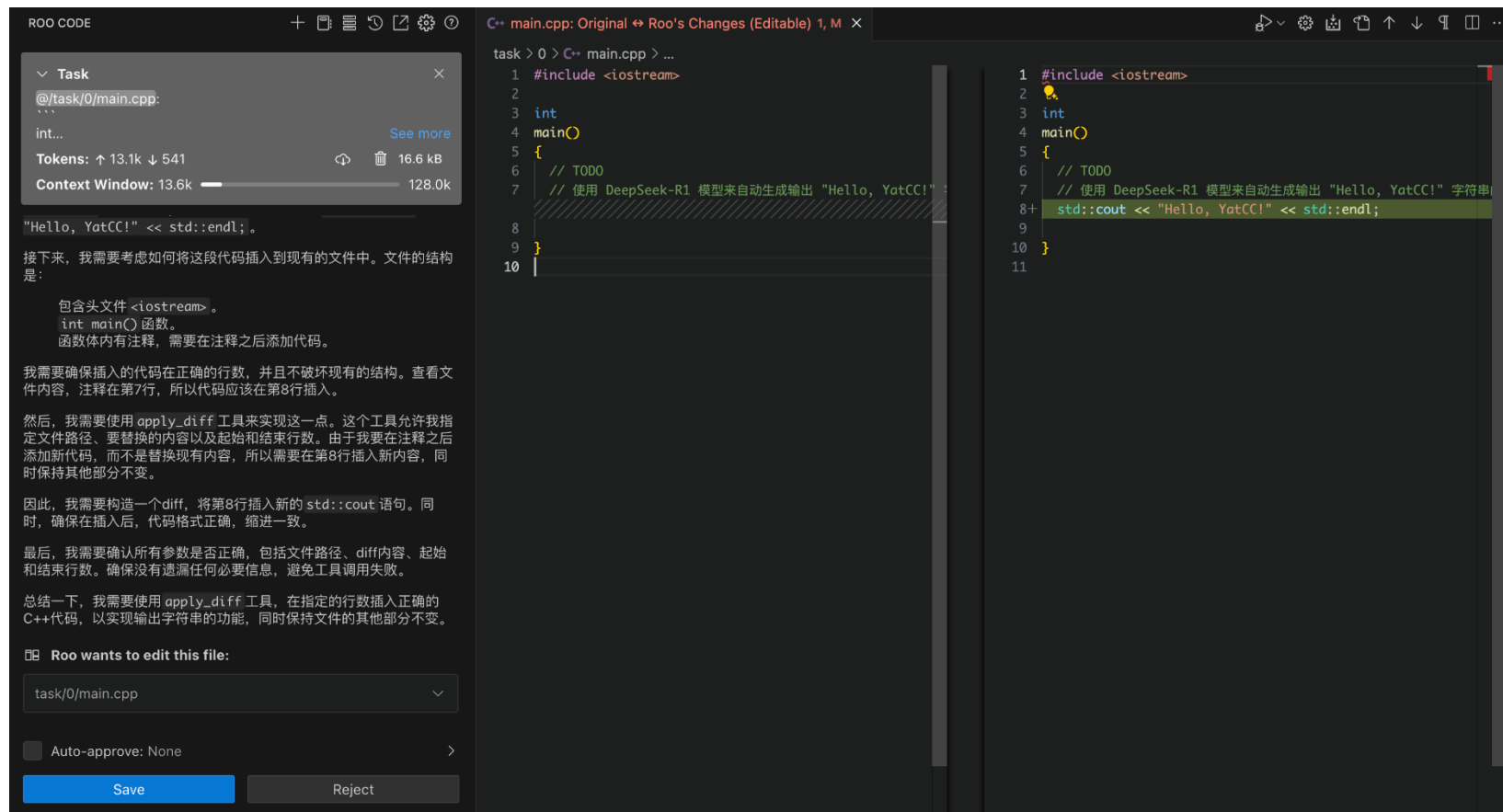
- 10 slip days 学生自行决定分配
- 每个实验逾期每天扣除10%得分
- 每个实验最多扣除40%得分



实验	0	一	二	三	四	五
内容	Env	Lexer	Parser	IRgen	IRopt	ASMgen
代码量/LOC	0	250	1000	1500	~500	<500
分数/百分比	0	15	15	25	30	5
用时/小时	1~2	2~6	24~72	36~108	24~108	<72
预留时间/天	7	14	30	30	30	14

目标要求：熟悉环境，开启“智能编译”之旅

- 上手云端开发环境
- 熟悉开发工具
- 熟悉任务构建、评分流程
- 熟悉AI大模型工具
- . . .

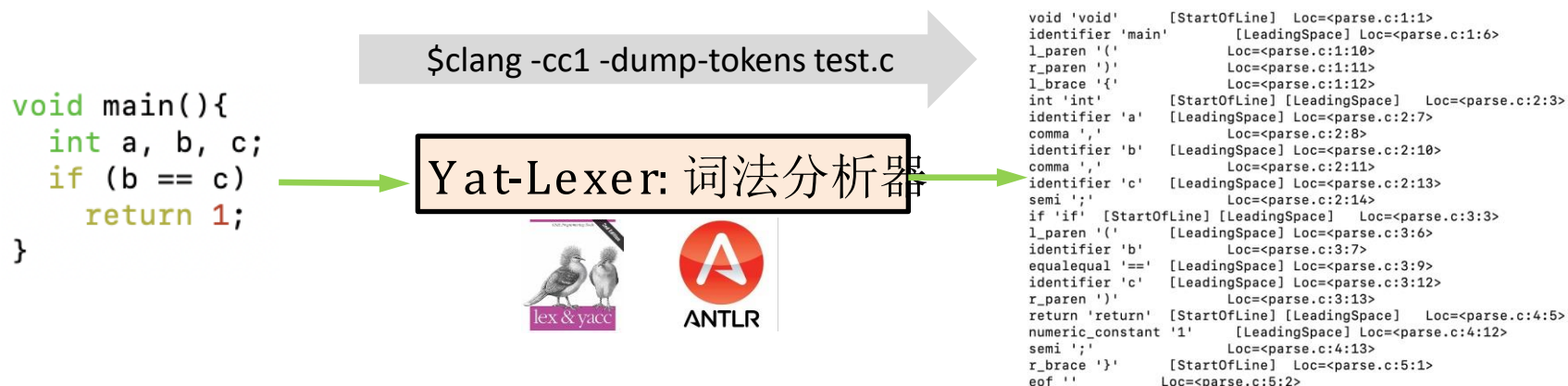


□ 目标要求：实现词法分析器，识别解析源代码token

- 采用Flex或ANTLR作为开发工具，输出结果与Clang输出匹配
- 需识别记录词号，对应文本内容、文件路径和词的行列号

□ 输入输出

- 输入：clang -E预处理过的源代码
- 输出：与Clang词法分析器一致的token流



目标要求：实现语法分析器，识别源程序语法结构

- 基于ANTLR或Flex+Bison自动解析工具，写出相应语法规则
- 识别语言的所有可能语法结构，构建解析树 (parse tree)

输入输出

- 输入：task1或Clang词法分析器生成的token流
- 输出：json格式抽象语法树，与标准Clang AST一致

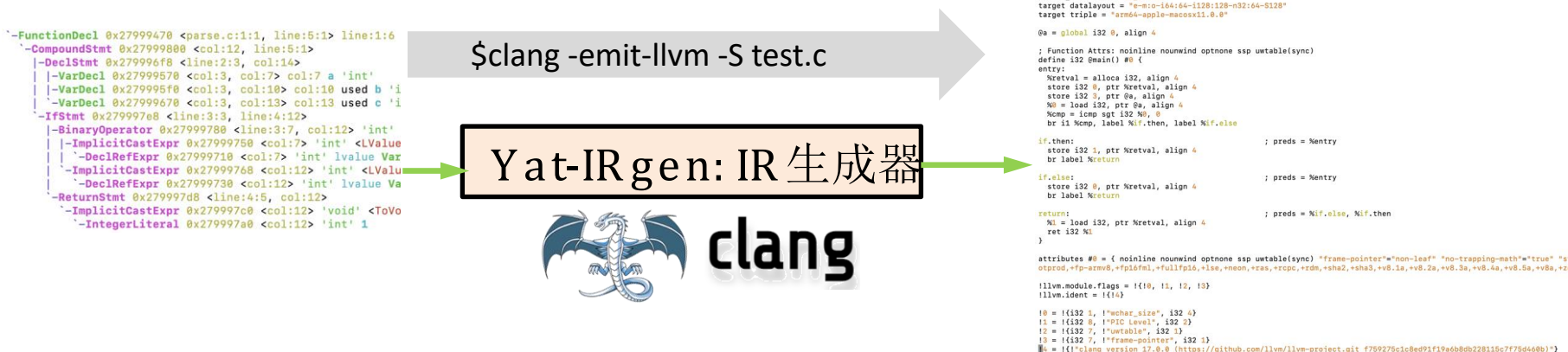


目标要求：实现中间代码生成器，生成源语言IR表示

- 基于LLVM设施和所提供实验框架实现，调用相关API和库
- 翻译出源程序对应的LLVM格式可读IR

输入输出

- 输入：task2或Clang语法分析器输出的json格式AST
- 输出：符合LLVM IR规范的中间表示文件

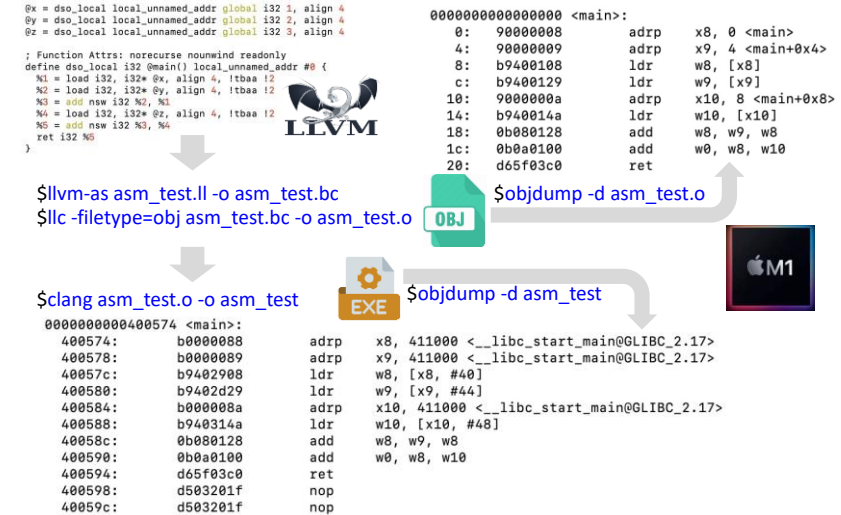


目标要求：将IR转换为目标机器汇编指令

- 使用LLVM库，将LLVM IR转换为汇编指令
- 生成的汇编指令能正确编译运行

输入输出

- 输入：task4或Clang提供的LLVM IR
- 输出：汇编指令



```

; Function Attrs: nounwind
; @run@%debug
define dso_local @main() local_unnamed_addr #1 {
entry:
  %call = tail call @_sys_getint() #3
  store i32 %call, ptr @loopCount, align 4, !tbaa !2
  tail call void @_sys_printf(i32 @roundof 2) #3
  %M = load i32, ptr @loopCount, align 4, !tbaa !2
  %cmp17 = icmp sgt i32 %M, 0
  br i1 %cmp17, label %while.cond1.preheader, label %while.end7

while.cond1.preheader:
; preds = %entry, %while.cond1.preheader
  %Mdivs.iv = phi i32 [ %Mdivs.iv.next, %while.cond1.preheader ], [ 0, %entry ]
  %M1819 = phi i32 [ %Madd6, %while.cond1.preheader ], [ 0, %entry ]
  %sum.018 = phi i32 [ %sum, %while.cond1.preheader ], [ 0, %entry ]
  %Mdiv = add i32 %Mdivs.iv, 6
  %Madd5 = add nsw i32 %div, %sum.018
  %sum = srem i32 %add5, 134209537
  %Madd6 = add nsw nsw i32 %M1819, 1
  %Mdivs.iv.next = add i32 %Mdivs.iv, 6
  %nextcond.not = icmp eq i32 %Madd6, %M
  br i1 %nextcond.not, label %while.cond.while.end7.crit_edge, label %while.cond1.preheader, !llvm.loop !6

while.cond.while.end7.crit_edge:
; preds = %while.cond1.preheader
  %M1 = add nsw i32 %M, -1
  store i32 %M1, ptr @global, align 4, !tbaa !2
  br label %while.end7

while.end7:
; preds = %while.cond.while.end7.crit_edge, %entry
  %sum.018ssa = phi i32 [ %sum, %while.cond.while.end7.crit_edge ], [ 0, %entry ]
  tail call void @_sys_stoptime(i32 @roundof 3) #3
  tail call void @_sys_printf(i32 @roundof %sum.018ssa) #3
  tail call void @_sys_printf(i32 @roundof 10) #3
  ret i32 0
}
    
```

\$llvm-as test.ll -o test.bc

Yat-AS Mgen: ASM生成器



```

31 main:
32   pushq   %rbp
33   .seh_pushreg   %rbp
34   movq    %rsp, %rbp
35   .seh_setframe %rbp, 0
36   subq   $48, %rsp
37   .seh_stackalloc 48
38   .seh_endprologue
39   call   __main
40   movl   $0, -4(%rbp)
41   movl   $0, -8(%rbp)
42   call   _sys_getint
43   movl   %eax, loopCount(%rip)
44   movl   $15, %ecx
45   call   _sys_starttime
46   jmp    .L4
47 .L7:
48   movl   $0, -12(%rbp)
49   movl   $0, -16(%rbp)
50   jmp    .L5
51 .L6:
52   movl   -8(%rbp), %eax
53   movl   %eax, %ecx
54   call   func
55   movl   global(%rip), %eax
    
```

OUTLINE

目 录

中山大学计算机学院

School of Computer Science & Engineering

一、YatCC-AI介绍

二、整体实验设计

三、总结与展望

学生竞赛巅峰



2023年全国大学生
计算机系统设计大赛
(华为毕昇杯) 全国
一等奖 (历史性突破)

教学案例标杆



2024年中国高校计算机
教育大会教学案例大赛
一等奖
2025年中国高校计算机
教育大会教学案例大赛
特等奖 (显示持续进步)

校级最高认可



2025年中山大学本科
/研究生教育教学成果
特等奖 (双特等奖)
2023年中山大学本科
教育教学成果 一等奖

学术与教改认可



2022年中国高校计算
机教育大会一等奖论文
2025年中山大学校级
教改项目立项

国际影响力认可



- 2025年“中日韩前喻计划 (A3)”研讨会最佳海报奖
- 2025年亚太人工智能学会优秀案例奖
- APAN'59、ISC25等国际会议上持续亮相

教学成果与学生竞赛成绩双丰收，充分验证了 YatCC-AI 教学范式的先进性与有效性。

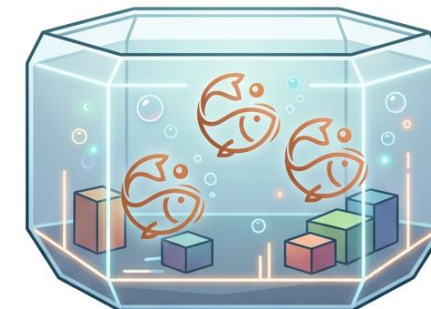
□ 定位核心基础专必课程本质，落实产教融合目标

- 动手构建编译器：基于理论教学，互补理论教学

□ 面向实践，强化过程，突出特色

- 丰富的教程资源，现代一站式开发体验，分阶段解耦推进
- 以LLVM为蓝本，大模型智能赋能，看齐业界与未来
- 平衡基础性、实践性和前沿性，构建编译系统全局观

A "MINI-COMPILER" MICRO-HABITAT FOR LEARNING



The "mini-compiler" simulation
(three colored-icons)



THE 'YAT COMPILER' PROGRAM
(as a mental/skill pattern)



YatCC-AI 智能 API 门户

依托中山大学与国家超算广州中心“天河星逸”，深度融合 HPC 与 DeepSeek 前沿大模型。
为您提供稳定、高性能的编译与 AI 算力服务接口。

立即开始探索

7x24h
稳定在线服务

DeepSeek
本地化推理加速

PF级
超算算力支撑



实验贡献者

实验文档

- 前言: 顾宇浩、郑腾扬、欧阳易芃、潘文轩
- Task0: 顾宇浩、单招文、潘文轩
- Task1: 单招文、何静仪、朱祉昕、**刘齐诚**
- Task2: 单招文、席梦悦、吴露、黄鑫、**刘齐诚**
- Task3: 郑中淳、孙高锦、黄鑫、潘胤臻、**刘齐诚**
- Task4: 郑中淳、黄瀚、潘文轩、**刘齐诚**

- [update\(task4\): 完成 task4 文档的调整](#)
#41 by Smallorange666 was merged on Jul 21, 2025
- [update\(task4\): 提升文档可读性](#)
#40 by Smallorange666 was merged on Jun 2, 2025
- [update: 增补内容, 增加可读性](#)
#39 by Smallorange666 was merged on May 22, 2025
- [update\(task2\): 重写antlr部分文档, 调整整体文档结构](#)
#38 by Smallorange666 was merged on Apr 24, 2025
- [update\(task1, task2\): 统一文档结构, 增加引导](#)
#37 by Smallorange666 was merged on Apr 24, 2025
- [更新了QA.md中关于task2的三个疑难问题](#)
#36 by hyy898573902 was merged on Apr 9, 2025



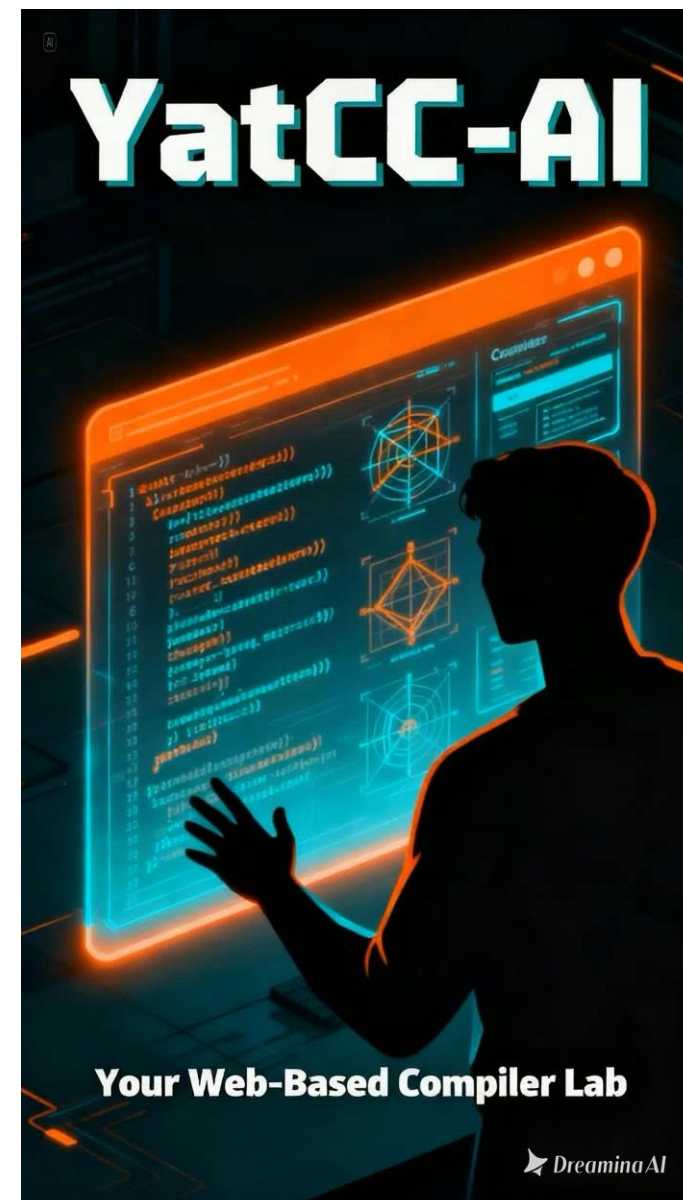
YatCC for Agent

□ 目标：构建一个覆盖 开发、训练、评测、部署、运营、治理的LLM智能体一体化平台

- Past: 面向教学的智能化代码开发、调试、运行的一站式一体化服务平台
- Now: 面向科研实践的智能化开发、部署、运行的一站式一体化服务平台
- Next: 面向LLM智能体全生命周期的一站式一体化服务平台

□ 小型项目：构建赋能教学与实践的智能体

- 部署 agent 同步 PPT、视频、文档、群聊、线下答疑等
- 构建 agent community, 专属 agent 的 StackOverflow
- 任你想象 ...



谢谢!

