

# 中国DevOps社区峰会 2024 · 上海

10.19 | 上海龙之梦大酒店



# 源于社区 服务社区

 中国DevOps社区峰会 2024 · 上海



## 生成式AI提升开发者效能

郑予彬 开发者布道师 亚马逊云科技





# 目录

- 1 当前生成式人工智能的趋势和挑战
- 2 技术革新与开发思维的变革
- 3 生成式人工智能赋能开发全流程
- 4 生成式人工智能对开发模式的影响





# 1

## PART 01

### 当前生成式人工智能的趋势和挑战



# 2023

## The Year of POCs



什么是生成式AI?

安全吗?

我需要成为prompt 的工程师吗?

如何选择模型?

我应该从哪里开始?



如何对我的业务产生影响?

怎么用在生成中?



这些模型都有哪些特性?

什么是基础模型?

什么是大语言模型?





# 2024 The Year of Production

(更多)



如何确定项目的优先次序?

如何降低成本?

如何落地?

哪种自定义手段适合我?



如何扩展?

哪个模型适合这个应用场景?

我应该训练自己的模型吗?

如何控制风险?



怎么在行业中创新?





# 生成式人工智能的未来 会怎样？

多种模型

多种模式

Agents

人工智能政策和标准





# 2

## PART 02

### 技术革新与开发思维变革







## 用于人工智能应用构建的工具和服务

全流程

自动化

智能化

## 用于调用大语言模型和基础模型构建应用的平台

模型选择

应用集成和调用能力

自定义能力

## 用于推理和预训练的计算架构



计算能力



预训练能力



推理能力



模型部署能力和推理架构兼容性



分布式系统和  
并行计算



接口与适配的SDK



弹性计算和有效成本



数据传输和同步





## 用于人工智能应用构建的工具和服务

全流程

自动化

智能化

## 用于调用大语言模型和基础模型构建应用的平台

模型选择

应用集成和调用能力

自定义能力

## 用于推理和预训练的计算架构



计算能力



预训练能力



推理能力



模型部署能力和推理架构兼容性



分布式系统和  
并行计算



接口与适配的SDK



弹性计算和有效成本



数据传输和同步





# 利用云环境进行软件开发

- 计算资源的可扩展性
- 数据存储和管理
- 协作与共享
- 成本优化
- 安全与合规
- 持续集成和部署
- 最新技术支持





## 用于人工智能应用构建的工具和服务

全流程

自动化

智能化

## 用于调用大语言模型和基础模型构建应用的平台

模型选择

应用集成和调用能力

自定义能力

## 用于推理和预训练的计算架构



计算能力



预训练能力



推理能力



模型部署能力和推理架构兼容性



分布式系统和  
并行计算



接口与适配的SDK



弹性计算和有效成本



数据传输和同步



# 模型评估



## Playground

直接在控制台的  
**Playground** 上试用不同的  
模型，比较成本、延时等



## 自动评估

使用通用数据集对于典型  
任务如分类、生成等自动  
进行评估，评价指标包括  
准确性、有害性和鲁棒性



## 利用自己的数据集 进行定制的评估

使用 统一的评价框架来组  
织自己的团队进行在自己  
的数据集上评估。



## 采用云供应商资源专 业评估

如果您内部没有人力或专  
业知识，可以利用云供应  
商专家团队进行评估

在定义使用场景并识别  
FM 基础效果时使  
playground

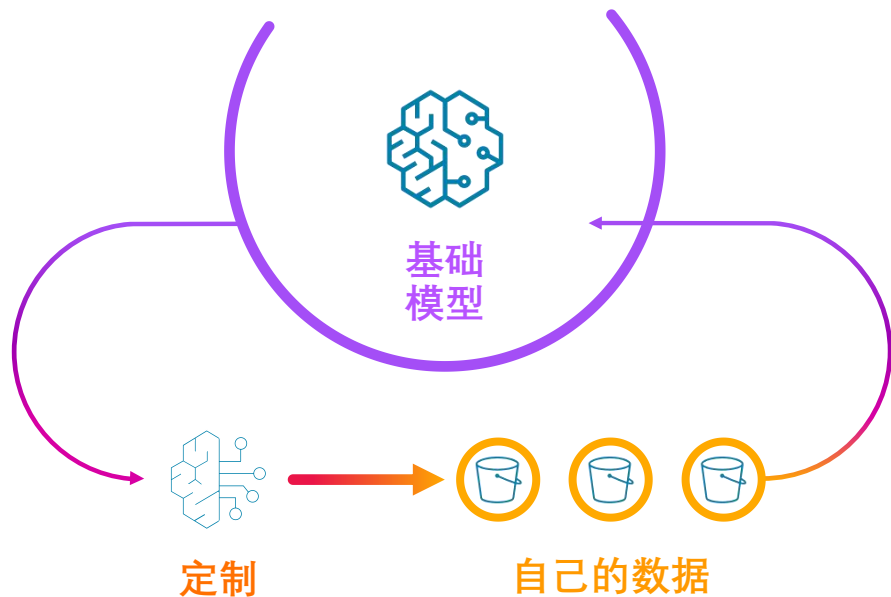
在迭代用例或模型时使用  
自动评估

当开始测试第一个原型应用或  
准备试运行，带上自己的团  
队进行评估

云供应商专业托管团队，为  
应用程序的生产发布提供专  
业的基础平台



# 利用自己的数据定制专属模型



检索增强生成

微调

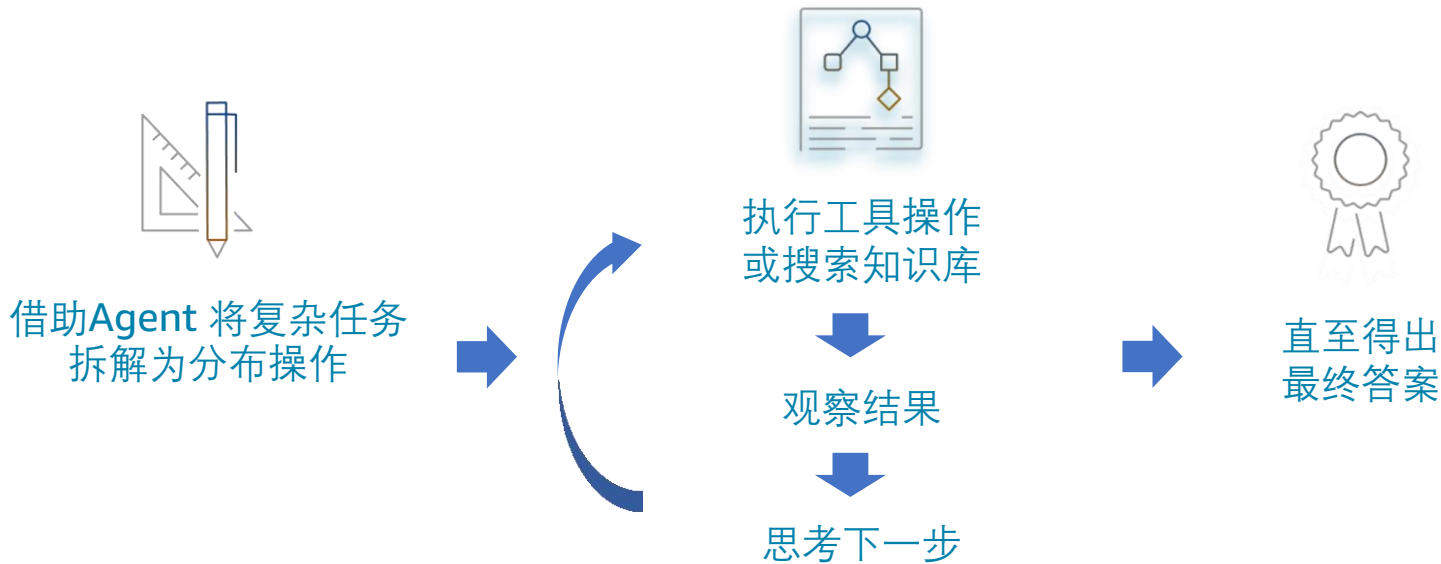
继续预训练





# 生成式AI的Agent

让生成式 AI 应用程序利用自己的系统和数据源执行多步骤任务





# 保护数据安全和隐私



不使用任何客户数据来训练底层模型

所有数据在传输和存储时均进行加密

用于训练自定义模型的数据保留在您的虚拟私有云中

支持多种标准，包括 **GDPR** 和 **HIPAA**





## 用于人工智能应用构建的工具和服务

全流程

自动化

智能化

## 用于调用大语言模型和基础模型构建应用的平台

模型选择

应用集成和调用能力

自定义能力

## 用于推理和预训练的计算架构



计算能力



预训练能力



推理能力



模型部署能力和推理架构兼容性



分布式系统和  
并行计算



接口与适配的SDK



弹性计算和有效成本



数据传输和同步





# 开发者是核心生产力

提升开发者满意度和幸福感

## 优化开发体验，提高开发效率

- 优化开发工具链
- 提升生产力
- 改善可维护性
- 简化工作流程
- 持续改进和创新



# 生成式人工智能在多个方面增强开发者工具并提升开发体验

**满意度：**提供智能代码建议、自动补全和代码生成等功能,减轻开发人员的工作负担,降低认知负荷,从而提高开发人员的**满意度和幸福感**

**开发速度：**实现的自动化和辅助功能,有助于更快地完成编码、调试和重构等任务,从而提高整体开发速度和效率。

**代码质量：**自动检测并发现代码中的错误和潜在问题,建议优化方案,并生成高质量的文档和注释,有助于提高代码质量和可维护性

**安全性：**可以被训练用于检测安全漏洞和不当编码实践,为开发人员提供安全建议,从而提高应用程序的安全性。

**入门速度：**对于新加入项目的开发人员,生成式AI可以生成代码解释和上下文文档,帮助他们更快地熟悉现有代码库,缩短入门时间

**技术债务：**识别代码库中的技术债务,提出重构建议,并协助自动化重构过程,帮助开发团队逐步消化





# 3

## PART 03

### 生成式人工智能的开发场景

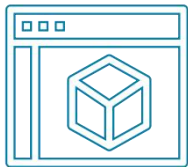




# 生成式人工智能参与开发全流程



# 人工智能工具已经就绪



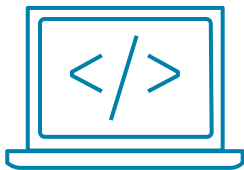
Amazon Management  
Console



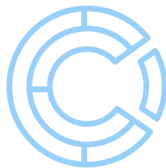
Amazon  
Documentation



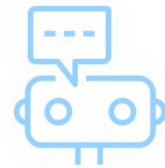
Amazon Console  
Mobile App



IDEs



Amazon  
CodeCatalyst



Slack and Teams  
(via AWS Chatbot)



- EC2 Dashboard
- EC2 Global View
- Events
- Instances
- Instances
- Instance Types
- Launch Templates
- Spot Requests
- Savings Plans
- Reserved Instances
- Dedicated Hosts
- Capacity Reservations
- Images
- AMIs
- AMI Catalog
- Elastic Block Store
- Volumes
- Snapshots
- Lifecycle Manager
- Network & Security
- Security Groups
- Elastic IPs
- Placement Groups
- Key Pairs
- Network Interfaces
- Load Balancing
- Load Balancers
- Target Groups
- Trust Stores
- Auto Scaling
- Auto Scaling Groups

Instances Info

Find Instance by attribute or tag (case-sensitive)

All states

Launch instances

Connect

Instance state

Actions

1

Name

Instance ID

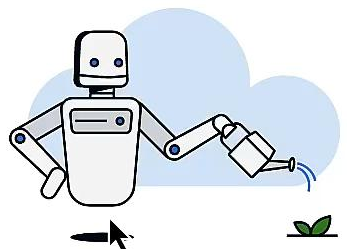
Instance state

Instance type

Status check

Alarm status

Availability Zone



No instances

You do not have any instances in this region

Launch instances

Select an instance



# Console Home

Reset to default layout

+ Add widgets

## Recently visited

- IAM
- CloudWatch
- Lambda
- S3
- DynamoDB
- API Gateway

View all services

## Welcome to AWS

### Getting started with AWS



Learn the fundamentals and find valuable information to get the most out of AWS.

### What's new with AWS?



Discover new AWS services, features, and Regions.

### Training and certification



Learn from AWS experts and advance your skills

## Explore AWS

### Build Web and Mobile Apps Fast

Learn how to build, scale, and deploy modern apps while leveraging AWS...

### AWS Support

Save time and move faster with

### Amazon Bedrock is Now General Availability

Amazon Bedrock is a fully managed service that offers leading foundation models...

### Calling All Java and Python Developers

Join the AWS BugBust challenge to

## Build a solution

Start building with simple wizards and automated workflows.



### Launch a virtual machine

With EC2 (2 mins)



### Start migrating to AWS

With AWS MGN (2 mins)



### Register a domain

With Route 53 (3 mins)

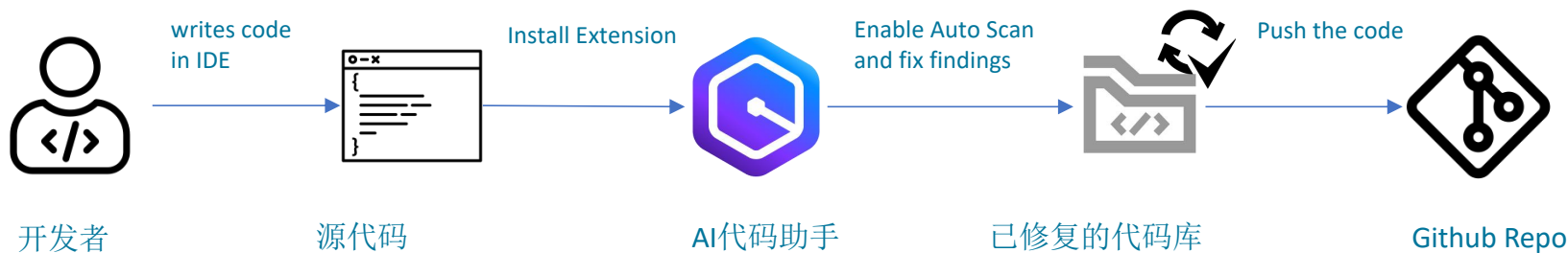


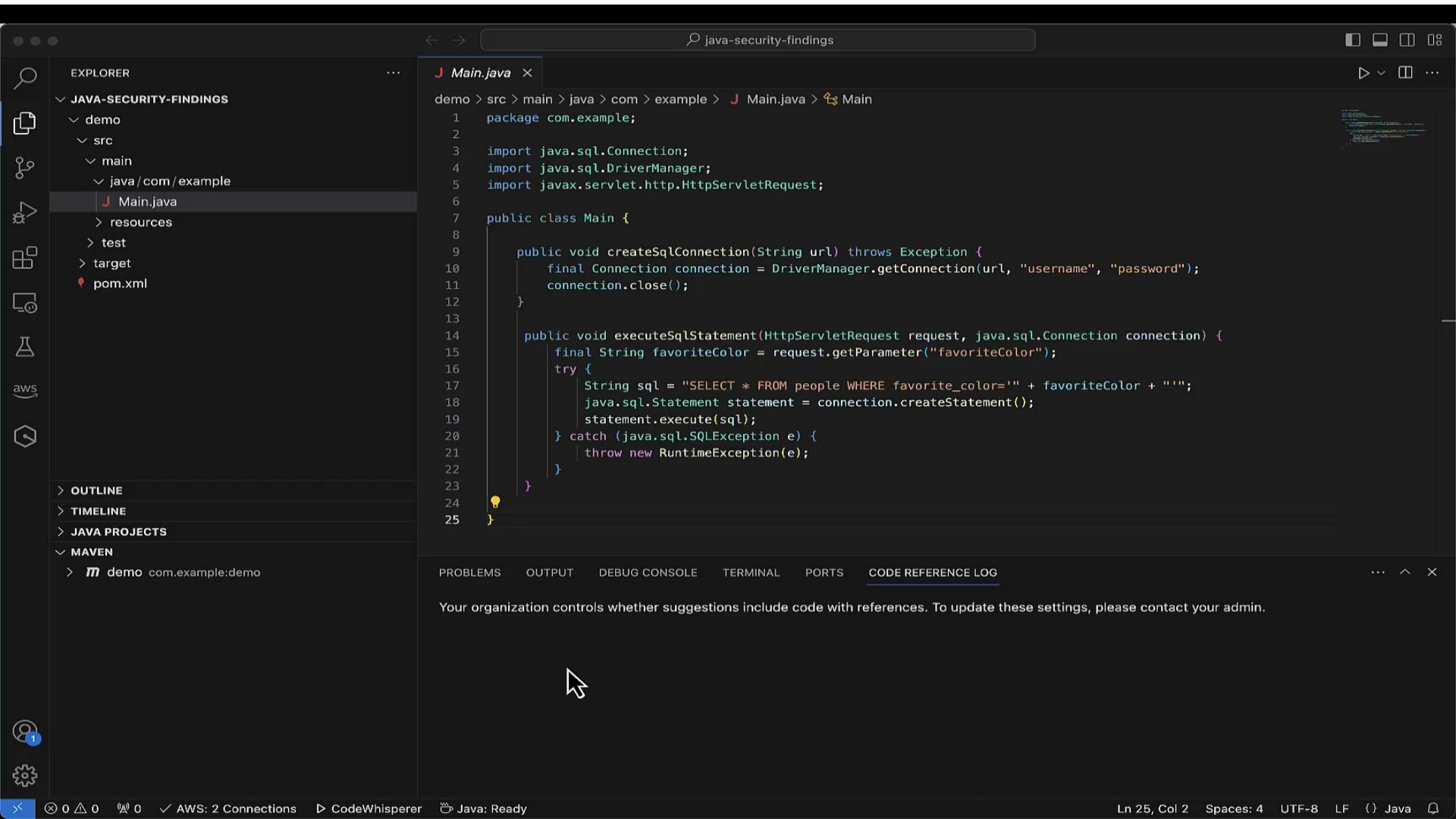
### Host a static web app

With AWS Amplify Console (2 mins)



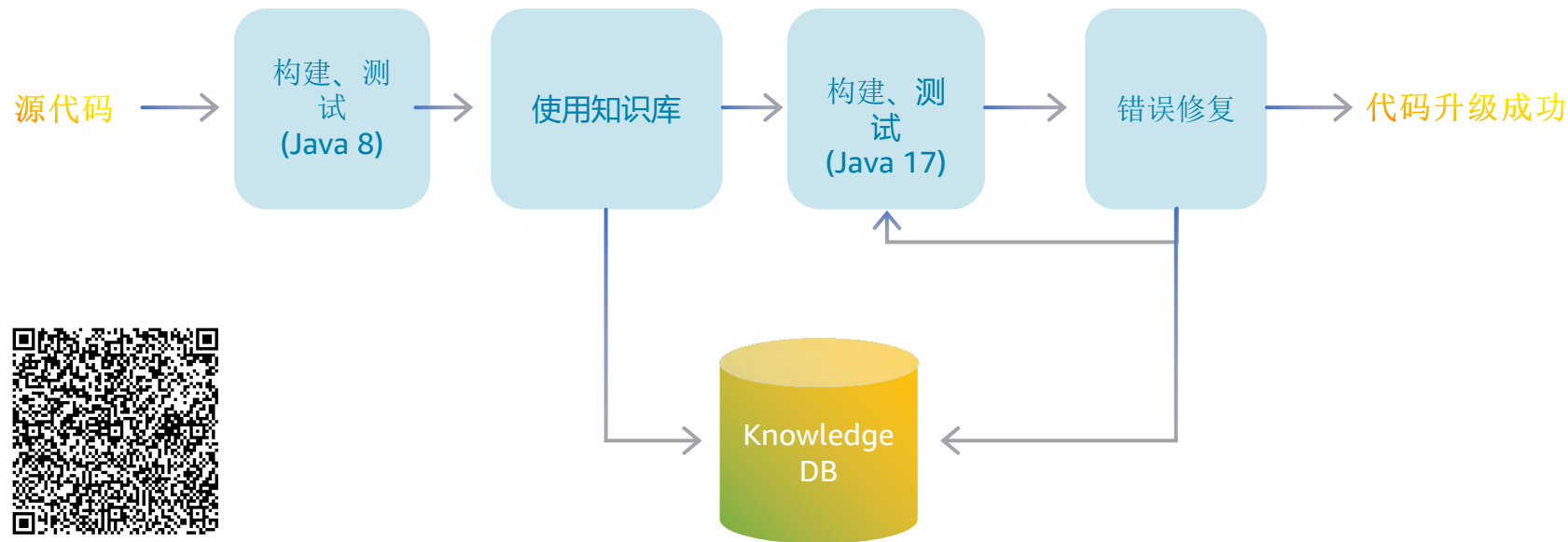
# 生成式AI轻松实现安全左移

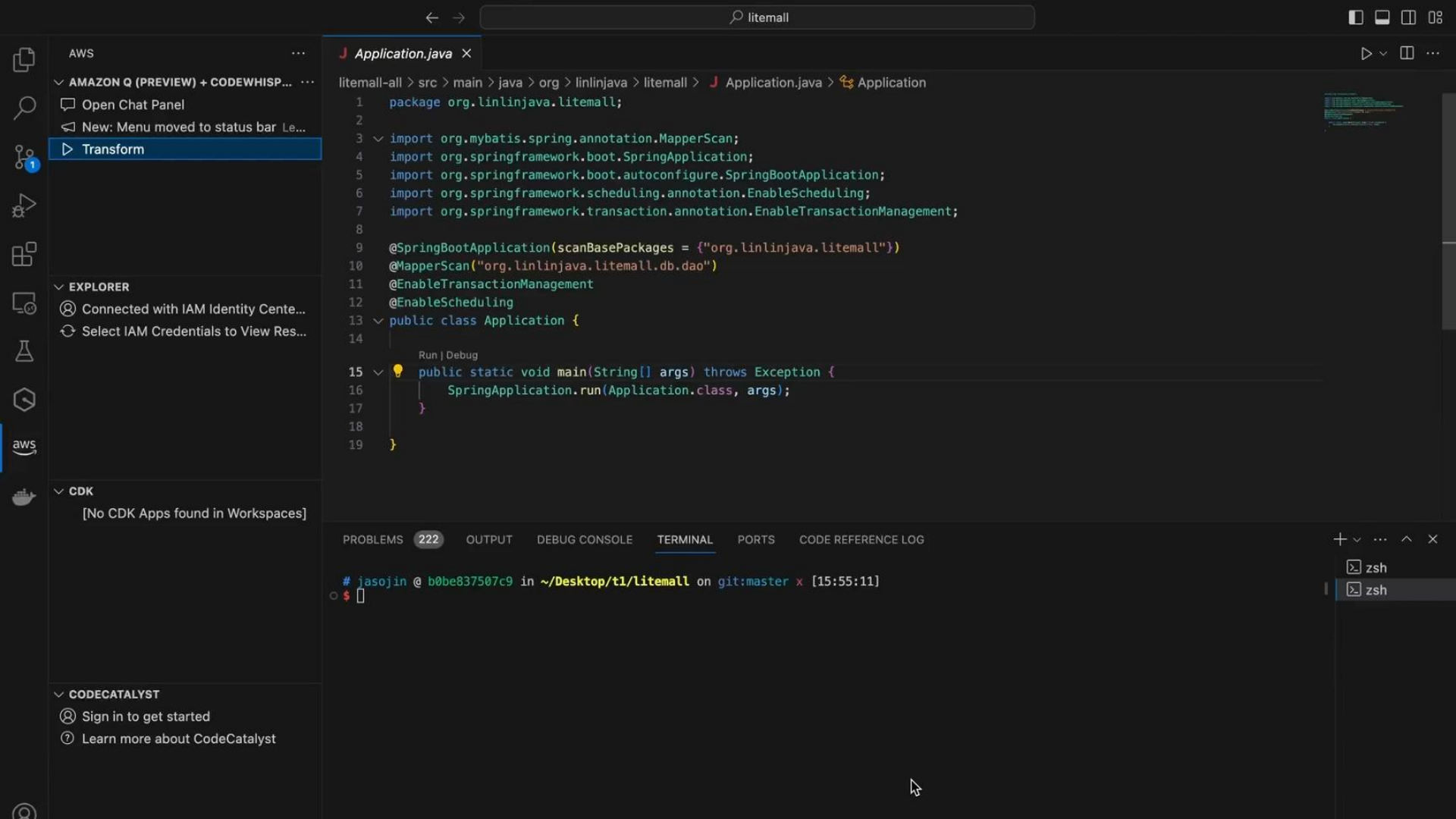




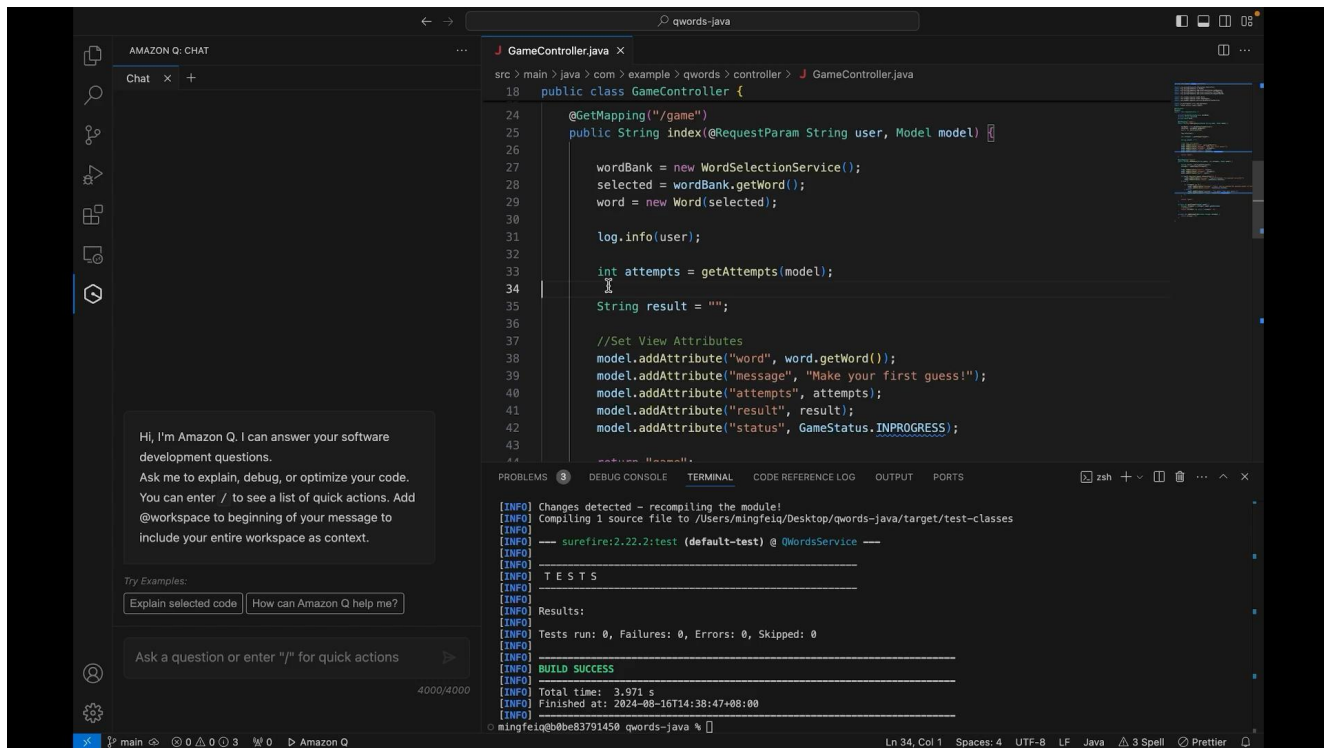


# 语言版本的快速升级

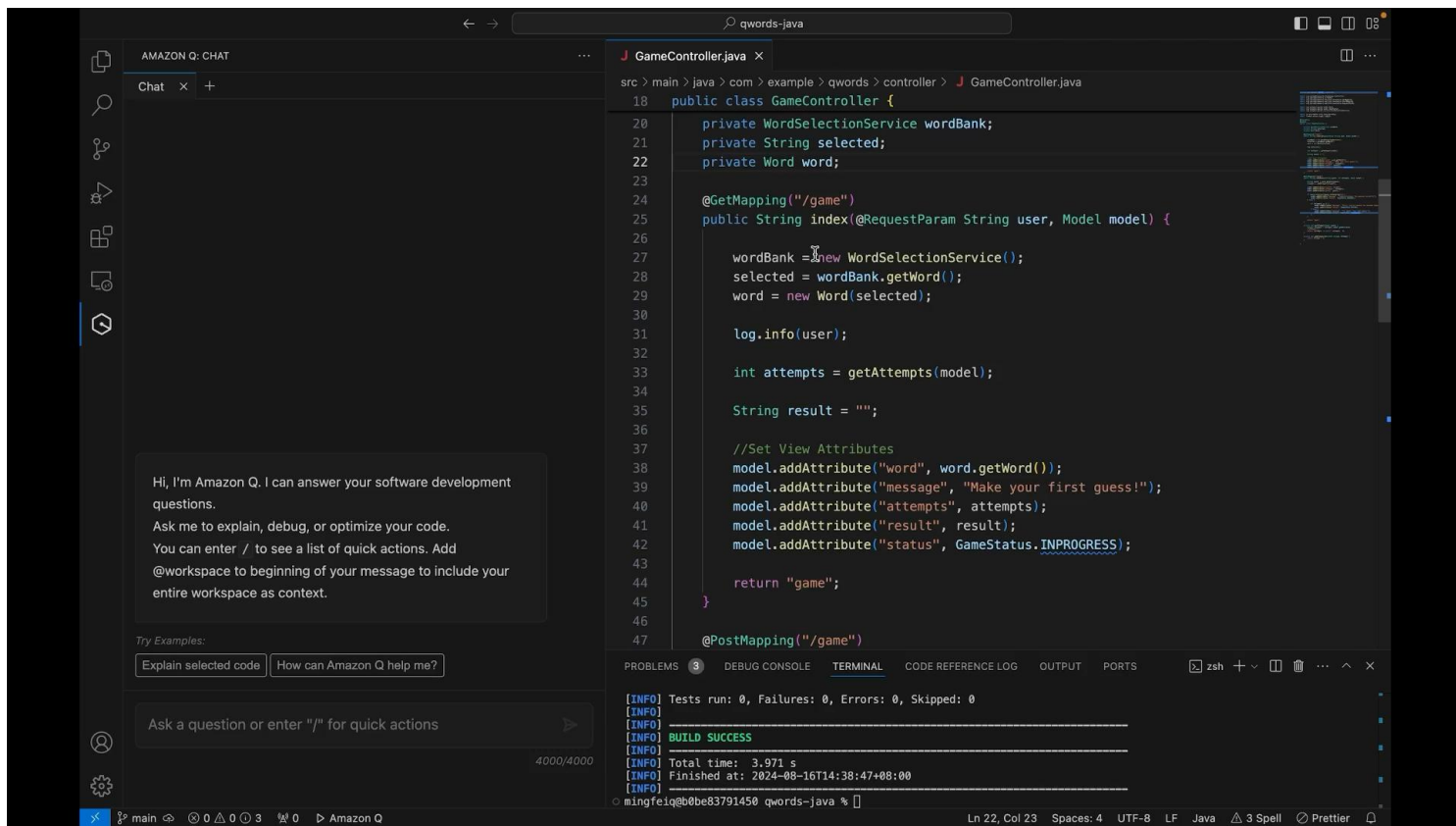




# 通过自然语言生成新功能



# 自动文档生成





# 4

## PART 04

### 生成式人工智能对未来开发模式的影响



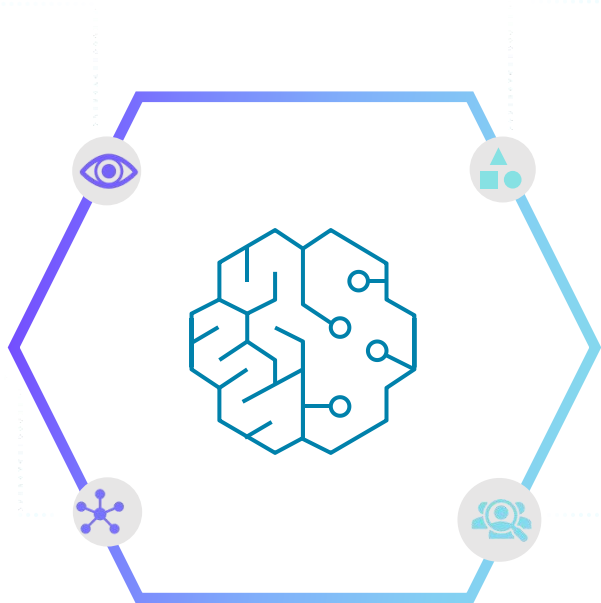
# 利用生成式AI继续提高开发效能

## 保持开放和谨慎的态度

- 探索尝试新的技术和方法
- 警惕过度依赖

## 灵活性是关键

- 避免技术和方法论的锁定,以保证随时调整创新的方向.
- 基础架构保持与第三方生成式AI集成的能力.
- 提供广泛的服务以确保长期灵活性和商业价值



## 因地制宜

- 生成式AI是生成的助力不是颠覆.
- 根据具体的业务场景也需求制定策略和方案, 以及选择合适的工具
- 评估和权衡自主开发和使用商业服务

## 倾听和赋能

- 获得并理解开发者意见和反馈.
- 为开发配置合适的工具, 为开发者赋能.



# 源于社区 服务社区

## THANKS!

