

JUNE 21 – 24, 2022 | LAS VEGAS, NV

The future of software development, powered by machine learning

Atul Deo

aws

Director, Product Management AWS Parminder Bhatia

Manager, Applied Science AWS

Typical journey of a developer



ML-assisted reviews and observability



Leveraging ML for writing code



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

CodeWhisperer: ML-powered coding companion

ACCELERATE APPLICATION DEVELOPMENT WITH CONTEXTUAL CODE RECOMMENDATIONS

aws

Generate ready-to-use code

Generate code from plain English comments

Generate complete functions

Available in all major IDEs as an extension

import csv import json # Function to convert a JSON file into a CSV file # Keys of the JSON are used as the headers of the CSV file def json_to_csv(json_file, csv_file): with open(json_file) as f:_______ data = json.load(f)

Open a csv writer
csv_writer = csv.writer(csv_file)

Write the headers
csv_writer.writerow(data[0].keys())

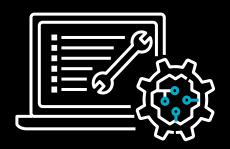
Write the data
csv_writer.writerows(data)

Suggestion by CodeWhisperer

What we're building

IMPROVE DEVELOPER PRODUCTIVITY WITH REAL-TIME CODE RECOMMENDATIONS







Languages: Python, Java, JavaScript, C#, TypeScript IDEs: JetBrains, VS Code, AWS Cloud9, AWS Lambda console, Visual Studio

First-class support for AWS APIs



The power of AI comes with a responsibility

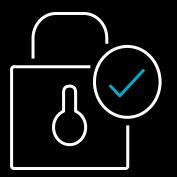
"CCS researchers find [TOOL] generates vulnerable code 40% of the time"

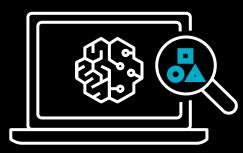
Center for Cybersecurity

"OpenAI warns AI behind [TOOL] may be susceptible to bias"

VentureBeat

We're building responsible AI







Security scan Reference tracker Bias avoidance

Customer & AWS Partner quotes

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.



"Distractions are a constant challenge while coding, especially when it's necessary to switch context to look up code samples and documentation on the web. Amazon CodeWhisperer keeps me focused on the code by automatically offering helpful suggestions right when I need them, so I never have to leave my editor."

Ryan Grove, Staff Software Engineer at SmugMug

FINRA

"Developers spend significant time looking up documentation for various APIs, cloud services, and libraries, like for looking up AWS documentation to describe security groups. This is where Amazon CodeWhisperer can be helpful for FINRA. For example, we can write a comment in the IDE to explain the task that may say something like, 'describe security groups based on a VPC ID in AWS.' Amazon CodeWhisperer would then automatically recommend the relevant code. In this way, CodeWhisperer can save developers time by providing real-time code recommendations directly in the IDE."

Brijesh Patel, Lead Software Engineer at FINRA



"We are excited to work with AWS on bringing Amazon CodeWhisperer to the IntelliJ Platform. At JetBrains, we aim to make software development a smooth and enjoyable experience. Availability of the plug-in for our tools will help developers stay focused in their IDE and reduce the need to search and customize code snippets from the web. As of today, users of IntelliJ IDEA, PyCharm, and WebStorm can start working with Amazon CodeWhisperer right in their IDE, with more IDEs to be supported in the near future."

Maxim Shafirov, JetBrains CEO

CodeWhisperer example

Project	AWS Toolkit: Developer Tools 🗸 🗸	© 🌣	🛵 json_exampl	ole.py 🛛 🛃 FileOps.py 🗶 🛃 main.py 🗶 📇 WelcomeToCodeWhisperer.md 🗡
Proj	 CodeWhisperer (Preview) Pause auto-suggestions Run Security Scan 			rt csv rt json
	Q Open Code Reference Panel			
			5 ⊟# Fun 6 ⊖# Key	nction to convert a JSON file to CSV ys of the JSON file are used as the header of the CSV file json_to_csv(json_file, csv_file):
				with open(json_file) as f:
				data = json.load(f)
			#	# Create a CSV writer
			c	csv_writer = csv.writer(csv_file)
			#	# Write the header
Structure			c	csv_writer.writerow(data[0].keys())
Stru			#	# Write the data
-			c	csv_writer.writerows(data)
🖈 Favorites				Insert Code Previous Next \neg \leftarrow \rightarrow Suggestion 3 of 5 from CodeWhisperer \vdots
×				
AWS Toolkit				
AWS			json_t	_to_csv()

CodeWhisperer example

Project	AWS Toolkit: Developer Tools 🛇 🗍 🌣	S3_example.py ×			
Proj	✓ CodeWhisperer (Preview) 1 II Pause auto-suggestions 2 ▶ Run Security Scan 3		import boto3	0 1 A	
	्ये Open Code Reference Panel		# Function to upload a file to S3 using server side encryption		
			<pre>def upload_file_to_s3_sse(file_name, bucket, object_name=None, sse_key=None):</pre>		
			"""Upload a file to an S3 bucket		
			:param file_name: File to upload		
			:param bucket: Bucket to upload to		
			:param object_name: S3 object name. If not specified then file_name is used		
			:param sse_key: Server side encryption key		
			return: True if file was uploaded, else False:		
			# If S3 object_name was not specified, use file_name		
			if object_name is None:		
			object_name = file_name		
			# Upload the file		
			s3_client = boto3.client('s3')		
re			try:		
Structure			response = s3_client.upload_file(file_name, bucket, object_name, ExtraArgs={'ServerSideEncryption': 'aws:kms', 'SSEKMSKeyId': sse	koul	
			except ClientError as e:	_кеууу	
-=			logging.error(e)		
es			return False		
Favorites			return True		
► Fav					
★			Insert Code Previous Next \rightarrow		
AWS Toolkit					
			Suggestion 1 of 2 from CodeWhisperer		
AWS			upload_file_to_s3_sse()		

CodeWhisperer security scan

ect	AWS Toolkit:Developer Tools 🛇 😂	🛃 json_example.py 🛛	ᡖ FileOps.py 🛛	樻 main.py 🛛	WelcomeToCodeWhisperer.md ×			
Structure	 CodeWhisperer (Preview) II Pause auto-suggestions Run Security Scan Q Open Code Reference Panel 	1 from flask 2 3 # This is 3 # This is 4 Gapp.route 5 def flask 6 from f 7 import 8 9 10 filena 11 12 12 # Secu 13 # with	<pre>from flask import app # This is a code sample to illustrate security scanning Gapp.route('/upload') def flask_SecurityExample(): from flask import request import re username = request.args.get('username') filename = request.files.get('attachment').filename # Security Issue: constructs regular expression from user input # without sanitization. re.search(username, filename)</pre>					
••								
ss		flask_Secur	ityExample()					
vorite	Problems: Current File 3 Project Errors	CodeWhisperer Securi	ty Issues 1 $ imes$			* -		
🖈 Favorites	Security Scan complete, 1 issues found in pythonProject Last Run 2022-06-21T07:30:51.173057Z							
🔇 AWS Toolkit 🛛	 FileOps.py /Users/ankurdes/PycharmProjects/pythonProject/FileOps.py 1 issue Resource leak: Allocated resources are not released properly. This can slow down or crash your system. They must be closed along all paths to prevent a resource leak. [14] 							

CodeWhisperer reference tracker

 CodeWhisperer (Preview) II Pause auto-suggestions Run Security Scan 							
Q Open Code Reference Panel							
						l	
	get_all_file	s()					
CodeWhisperer Reference Log						\$ -	
Don't want suggestions that include code from other sources? Edit in <mark>CodeWhisperer Settings</mark> .							
<pre>[21:37:58] ACCEPTED recommendation with the following code provided with reference under MIT from repository drfs. Added to main.py at line 5 to 9 return [os.path.join(root, name) for root, dirs, files in os.walk(directory) for name in files]</pre>							
≔ TODO ❶ Problems	Reference Log 🛛 Terminal	📚 Python Packages	Network Console			1 Event Log	
Amazon CodeWhisperer: Access code successfully verified! (1 hour ago)			9:6 LF UTF-8 4s	spaces Python 3.10 (pythonProject)	AWS: default@us-east-1	CodeWhisperer	

How it works

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Programming language processing (PLP)

Recent surge of applying deep learning on program analysis

Code search

Clone detection

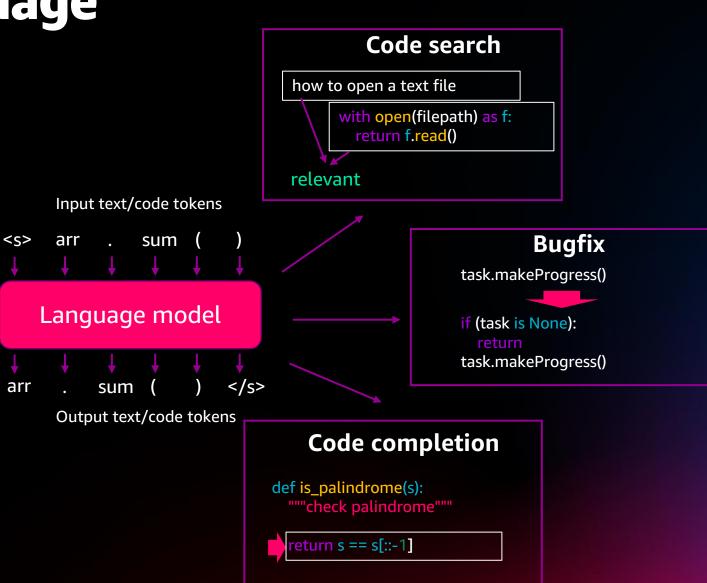
Code repair

. . .

aws

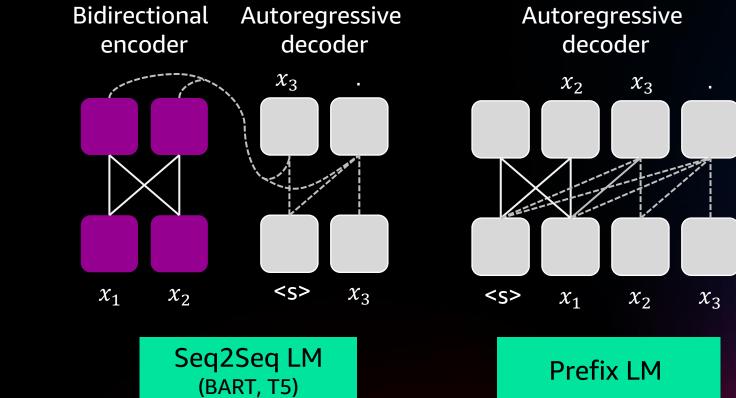
Code generation

Code summarization



Large language models

Autoregressive decoder x_1 x_2 x_3 <s> x_1 x_2 x_3 Language model (GPT)



Code-aware training



def max (MASK, b): return MASK if MASK > b else b

Recovered code

def max (a, b): return a if a > b else b

aws

(A) Masked token recovery

Corrupted code

def max (b): return if > b else b

Recovered code

def max (a, b):
 return a if a > b else b

(B) Deleted token recovery

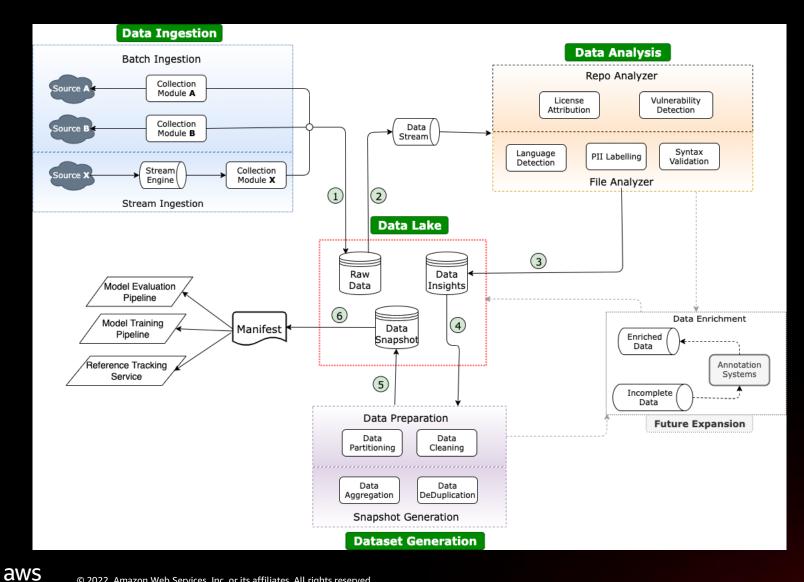
Corrupted code def MASK0 (MASK1, MASK2): return MASK1 if MASK1 > MASK2 else MASK2

Predicted tokens

MASKO max MASK1 a MASK2 b

(C) Masked identifier prediction

Data pipeline

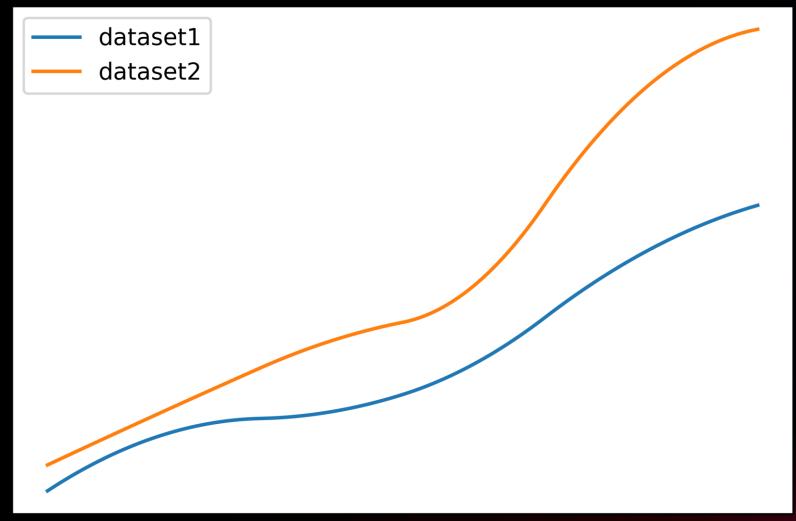


Major components

- Ingestion
- Aggregation
- Filtering
- Transformation
- License attribution
- Cleaning
- Deduplication
- Tokenization
- Formatting
- Chunking
- Sampling

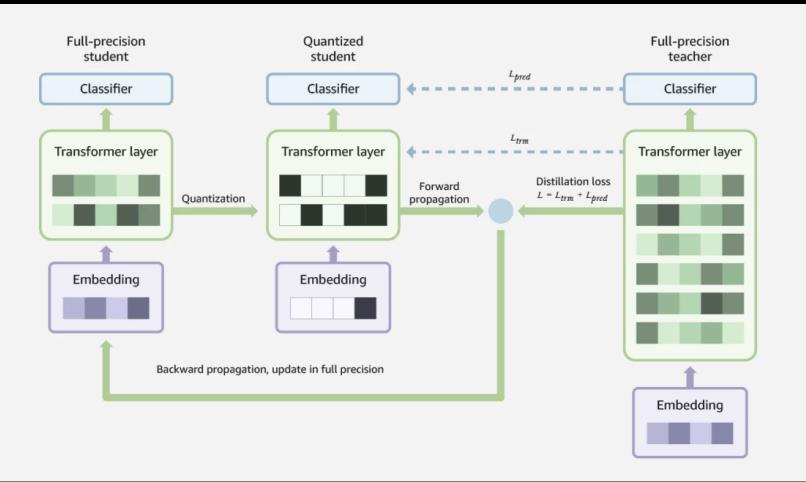
. . .

Evaluations



Model accuracy increases when model size increases

Efficient model compression techniques



Tensor parallelism

Distillation

Quantization

Compilation

"DQ-BART: Efficient sequence-to-sequence model via joint distillation and quantization" Figure based on: <u>TernaryBERT</u>: <u>Distillation-aware Ultra-low Bit BERT</u>

Sign up for the preview

Please visit <u>https://aws.amazon.com/codewhisperer</u> to request access for the preview

Users will receive an email with instructions for getting started

Thank you!

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.