



AWS Academy Natural Language Processing
Module 01 Student Guide
Version 0.1.1

200-ACMNLP-01-EN-SG

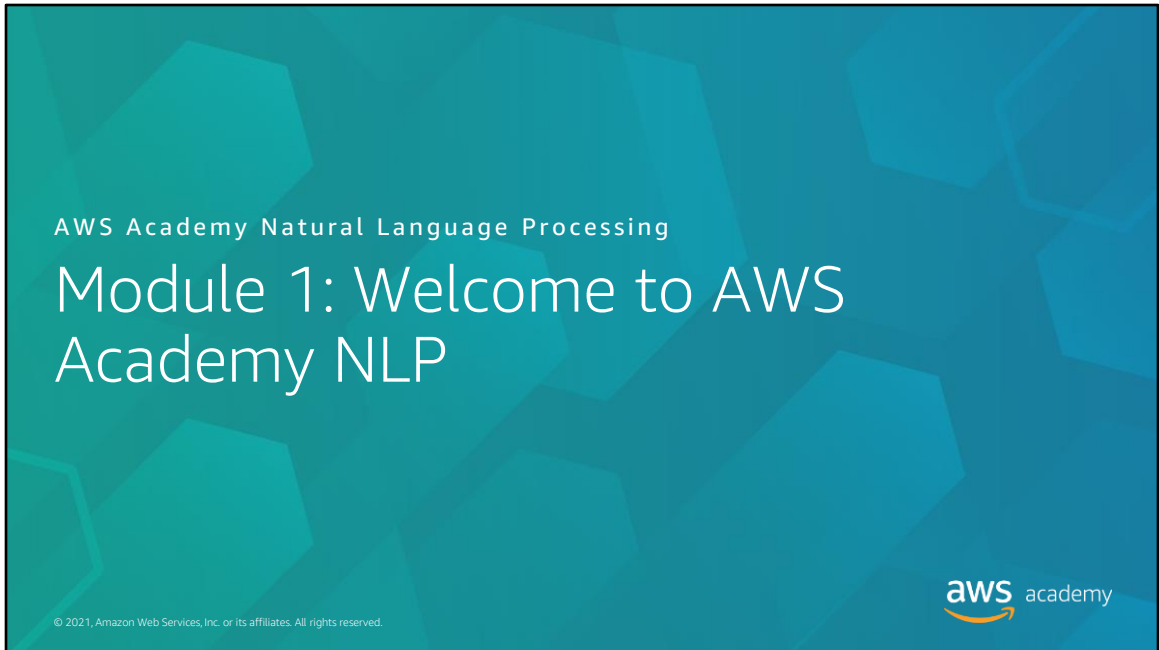
© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

This work may not be reproduced or redistributed, in whole or in part,
without prior written permission from Amazon Web Services, Inc.
Commercial copying, lending, or selling is prohibited.

All trademarks are the property of their owners.

Contents

Module 1: Welcome to AWS Academy NLP	4
--	---



Welcome to AWS Academy NLP

Module overview



Sections

1. Course overview
2. What is NLP?
3. Business problems solved by using NLP
4. NLP roles
5. What's new with AWS

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

2

This module covers the following sections:

1. Course overview
2. What is NLP?
3. Business problems solved by using NLP
4. NLP roles
5. What's new with AWS

Module objectives



At the end of this module, you should be able to:

- Identify course prerequisites and objectives
- Describe natural language processing (NLP)
- Describe business problems that are solved by using NLP
- Describe various roles that require NLP
- Identify NLP tasks for each role
- Explore new technologies at AWS

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

3

At the end of this module, you should be able to:

- Identify course prerequisites and objectives
- Describe NLP
- Describe the business problems that are solved by using NLP
- Describe the various roles that require NLP
- Identify NLP tasks for each role
- Explore new technologies at AWS

Module 1: Welcome to AWS Academy Natural Language Processing


Section 1: Overview

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.



This is section 1: Course prerequisites and objectives

Course prerequisites



- IT technical knowledge
- Completion of AWS Machine Learning Foundations
- Intermediate skills with Python programming
- General knowledge of applied statistics

General required knowledge

- Familiarity with machine learning (ML) concepts
- Working knowledge of distributed systems
- Familiarity with general networking concepts
- Working knowledge of multi-tier architectures

Preferred knowledge

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

5

Review the prerequisites of this course.

- First, you should already have some general **IT technical knowledge**. The foundational computer literacy skills students need to be successful include basic computer concepts, email, file management, and a good understanding of the internet.
- This course requires that you have completed the AWS Machine Learning Foundations course.
- You also should have intermediate skills with Python programming.
- Finally, you should have a general knowledge of applied statistics.

To ensure success in this course, it is preferred that you have:

- A general familiarity with **machine learning (ML)** concepts.
- A working knowledge of **distributed systems**.
- Familiarity with **general networking** concepts.
- And a working knowledge of **multi-tier architectures**.

Course objectives



- Describe NLP
- Understand how to apply the ML pipeline to NLP
- Implement text extraction to obtain data from webpages
- Create a solution for a sentiment analysis business problem
- Use Amazon Comprehend to implement a topic modeling solution
- Build a solution that uses AWS services to transcribe and translate text from multi-media
- Build a solution using a combination of algorithms and Amazon Machine Learning services

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.



By the end of the course you will be able to:

- Describe NLP
- Understand how to apply the ML pipeline to NLP
- Implement text extraction to obtain data from webpages
- Create a solution for a sentiment analysis business problem
- Use Amazon Comprehend to implement a topic modeling solution
- Build a solution that uses AWS services to transcribe and translate text from multi-media
- And, build a solution using a combination of algorithms and Amazon Machine Learning services

Course outline



- Module 1: Welcome to AWS Academy NLP
- Module 2: Introduction to NLP
- Module 3: Processing text for NLP
- Module 4: Implementing Sentiment Analysis
- Module 5: Introduction to Topic Modeling
- Module 6: Introduction to Information Extraction
- Module 7: Handling Language
- Module 8: Class Project
- Module 9: Course Review and Bridging to Certification



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

7

To achieve the course objectives, the course is divided into the following modules:

- This module is an overview of the course.
- Module 2 provides an Introduction to NLP
- In module 3 you will learn about processing text with NLP
- Module 4 describes how to implement Sentiment Analysis
- Module 5 Introduces topic modeling
- In module 6, you will learn how to extract information.
- Module 7 describes how to work with different human languages.
- In modules 8, you will work on a class project.
- And finally, module 9 will review the course. You will also learn more about the steps you must take to prepare for certification.

The next 10 slides provide more detail on what subtopics are covered in each module.

Module 2: Introduction to NLP



Module sections:

- NLP and machine learning
- Common NLP tasks
- Walkthrough of an NLP problem
- Evolution of NLP



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

8

The purpose of module 2 is to introduce you to the NLP domain.

- Section 1 describes the relationship of NLP and the general field of machine learning.
- Section 2 describes common NLP tasks
- Section 3 provides a walkthrough of an NLP problem
- Finally, section 4 describes how to apply machine learning to an NLP problem

Module 3: Processing text with NLP



Module sections:

- Text processing overview
- Extracting text
- Tokenizing text
- Common processing steps
- Advanced processing
- Storing and visualizing unstructured data



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

9

Module 3 provides an overview of how to process text with NLP.

- In Section 1, you are introduced to the overall process for extracting text.
- In Section 2, you learn how to extract text from an HTML page with open source tools and Amazon Textract.
- Section 3 reviews how to tokenize text.
- Section 4 describes common processing steps
- Section 5 covers some more advanced forms of text processing.
- And finally in Section 6 you learn how to store and visualized unstructured data

Module 4: Implementing sentiment analysis



Module sections:

- Scenario introduction
- Identifying text processing steps
- Examining algorithms for sentiment analysis
- Challenge lab: discussion and walkthrough



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

10

Module 4 provides an introduction to sentiment analysis

- In Section 1, you are introduced to a scenario that you will use in the module.
- Section 2 identifies the steps for processing text to analyze sentiment.
- Section 3 reviews common algorithms used for sentiment analysis
- Finally, section 4 reviews the challenge lab for sentiment analysis.

Module 5: Introducing information extraction



Module sections:

- Information extraction overview
- Types of information extraction



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

11

The purpose of module 5 is to introduce you to information extraction.

- In section 1, you will get an overview of information extraction and learn about common information extraction applications
- Section 2 covers several types of information extraction

Module 6: Introducing topic modeling



Module sections:

- Introduction to topic modeling
- Identifying the approach
- Introduction to Amazon Comprehend
- Algorithms
- Making recommendations based on text



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

12

Module 6 introduces how to create a model of topics from a group of documents.

- Section 1 provides a high-level overview of topic modeling concepts
- Section 2 describes the overall approach to building a topic model
- In Section 3, you will learn about Amazon Comprehend.
- Section 4 reviews different algorithms you can use for topic modeling
- And finally, Section 5 describes how to make recommendations based on text

.

Module 7: Handling language



Module sections:

- NLP and languages
- Identifying languages
- Translating text with Amazon Translate



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

13

In Module 7, you will learn how to work with different languages for NLP problems.

- Section 1 provides a high-level overview of the challenges of working with languages
- Section 2. describes the overall process for translating text.
- In section 3, you learn how to use Amazon Translate to translate text between languages.

Module 8: Course recap and bridging to certification



Module sections:

- Capstone project
- Course recap
- Bridging to certification



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

14

The final module provides a summary of the course and describes high-level steps for working towards achieving the AWS Certified Machine Learning – Specialty.

- Section 1 introduces the capstone project
- Section 2 provides a high-level review of the topics that the course covered.
- Section 3 reviews the requirements for attaining the AWS Certified Machine Learning – Specialty. It suggests some next steps for moving forward on your path to certification.

Module 1: Welcome to AWS Academy Natural Language Processing

Section 2: What is NLP?

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

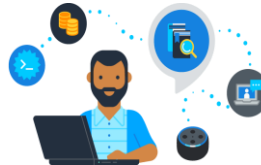


Welcome to Section 2: What is NLP?

NLP example



"Alexa, what's it like outside?"



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

16

Before you learn about natural language processing (NLP), review the following example of NLP: Amazon Alexa.

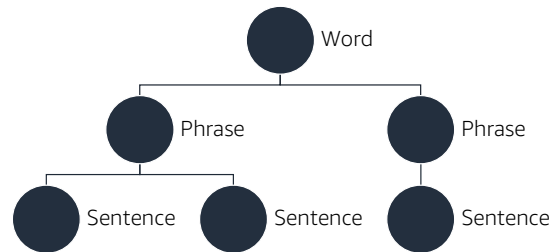
Here is a diagram of how Amazon Alexa uses NLP to create a conversation with users. First, an Amazon device, such as an Echo, records your words. The recording of your speech is sent to Amazon servers to be analyzed more efficiently. Amazon breaks down your phrase into individual sounds. Then, it connects to a database that contains the pronunciations of various words to find which words most closely correspond to the combination of individual sounds. Then, Alexa identifies important words to make sense of the tasks and to carry out corresponding functions. For instance, if Alexa notices words like *outside* or *temperature*, it opens the weather app. Finally, Amazon servers send the information back to your device, and Alexa speaks.

What is NLP?



NLP develops computational algorithms to analyze and represent human language automatically.

By evaluating the structure of language, ML systems can process large sets of words, phrases, and sentences.



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

17

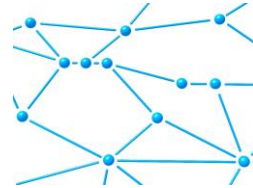
NLP is a broad term for a general set of business or computational problems that you can solve with machine learning (ML). NLP systems predate ML. Speech-to-text on your old cell phone and screen readers are both examples of NLP solutions. Many NLP systems now use some form of machine learning. NLP works with the hierarchical structure of language. Words are at the lowest layer of the hierarchy. A group of words makes a phrase. The next level up in the hierarchy is a group of phrases, which make a sentence, and ultimately, sentences convey ideas.

NLP systems face several significant challenges, which you will learn about next.

NLP challenges



Lack of precision



Many complex dependencies



Meaning that is based on context



Lack of structure

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

18

Language is not precise. Often, the same words or phrases can have multiple meanings. For example, consider the term *weather*. You might be *under the weather*, which means that you are sick. However, *there is wonderful weather outside* means that the weather conditions outside are good. Words can have different meanings, which are based on the other words that surround them. The words that surround the words are the context.

Words can also have different meaning based on inflection. For example, the phrase *Oh, really?* might convey surprise, disagreement, or many other meanings, depending on a combination of context and inflection.

Some of the main challenges for NLP include:

- Discovering the structure of the text – One of the first tasks of any NLP application is to break the text into meaningful units, such as words, phrases, and sentences.
- Labeling data – After the system converts the text to data, the next challenge is to apply labels that represent the various parts of speech. Every language requires a different labeling scheme to match the language's grammar.
- Representing context – Addressing this challenge is where machine learning can have a significant impact. Because word meaning depends on context, any NLP system needs a way to represent context. It is a large challenge because of the large number of contexts.

Converting context into a form that computers can understand is difficult.

- Applying grammar – Although grammar defines a structure for language, the application of grammar is nearly infinite. Dealing with the variation in how humans use language is a major challenge for NLP systems

Module 1: Welcome to AWS Academy Natural Language Processing

Section 3: Business problems solved by using NLP

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.



Section 3 reviews some of the business problems you can solve with NLP.

Natural language processing use cases



Search applications



Translations



Market and social research



Human machine interfaces



Content management



Chatbots

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

20

You can apply NLP to a wide range of problems. Some of the more common applications include:

- Search applications (such as Google and Bing)
- Translation services
- Sentiment analysis for marketing or political campaigns
- Social research that is based on media analysis
- Human machine interfaces (such as Alexa)
- Content management applications
- Chatbots to mimic human speech in applications

Module 1: Welcome to AWS Academy Natural Language Processing

Section 4: NLP roles

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

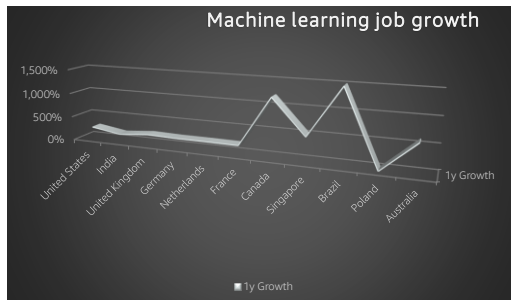


As you learned in the AWS Academy Machine Learning Foundations course, many different roles work with NLP. Review some of the different roles that work with NLP applications in the following slides.

Careers in ML



- Demand for machine learning (ML) skills is very high and growing
- Variety of ML roles work on NLP problems
- Technical skills with the fastest growth in demand for ML roles
 - Python programming
 - SQL
 - Data science
 - Data mining
 - Data analysis



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

22

The demand for machine learning (ML) skills is growing rapidly. The ML roles that you learned about in AWS Academy Machine Learning Foundations also apply to NLP. The technical skills with the highest demand for working in ML are:

- Python programming
- SQL
- Data science
- Data mining
- Data analysis

Data scientist or data analyst



- Applying knowledge of statistics and analytical skills to interpret data
- Data scientists often have degrees in statistics, computer science, or economics
- Some programming skills are required
- Typical tasks: Acquire and prepare data, exploratory data analysis, present data analysis



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

23

As a data scientist or data analyst, you use analytic and statistical skills to collect, analyze, and interpret large datasets. When you work on NLP applications, you might be asked to evaluate ML models to optimize outputs. For example, you could be asked to determine whether a topic modeling application correctly classifies topics, and whether classification errors are statistically significant. Data scientists and analysts often work with what are known as *big data*. Big data datasets are characterized by what is known as the *four V's* (volume, variety, velocity, and veracity).

Data scientists and analysts acquire and prepare data. They also explore and analyze datasets to help select appropriate algorithms and models. When data scientists and analysts work with big data datasets, they develop strategies for handling the four V's. Data scientists and analysts often validate findings and present recommendations to stakeholders.

ML engineer



- Emphasis on programming and system design skills
- Often have background as a developer or software architect
- Some knowledge of statistics required
- Typical tasks: Evaluating and selecting algorithms, training and evaluating models, designing NLP applications



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

24

NLP engineers need some skills that are similar to a data scientist's skills. However, as an engineer, you focus more on programming skills and software architecture. As an ML engineer, you apply those skills to designing and developing ML systems. An ML engineer who works on an NLP application would be more involved in designing and implementing the end-to-end solution for the NLP problem.

NLP engineers evaluate and select algorithms and models that are appropriate to the specific NLP business problem. They also design end-to-end solutions and select appropriate tools to create those solutions.

Applied science researcher



- Applies machine learning technology to a specific domain
- Requires knowledge of both the domain and machine learning
- Typical tasks: Using ML services to conduct research, applying NLP tools to a domain-specific application



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

25

You might also decide to work toward a career in science where you can apply ML technology to an NLP application. ML has had an impact on everything from astronomy to zoology, so many different paths are open to you. As an applied science researcher, your primary focus is on the type of science that you decide to concentrate on. For example, an applied science engineer might use ML to develop an NLP application that processes a large volume of medical diagnoses to look for patterns of misdiagnosis.

Applied science researchers can be from many different disciplines. The tasks they undertake are specific to their particular domain of expertise. For example, a biology researcher might use NLP tools to analyze findings from a collection of lab summaries.

Developer



- Integrates ML with software applications
- Requires strong application development skills and machine learning knowledge
- Typical tasks: Incorporating NLP models or services into an application



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

26

Many software developers are now adding NLP functionality to their applications. For example, voice interfaces and text analysis that are based on ML are becoming commonplace. As an ML developer who focuses on NLP, your primary focus is software development skills, but you also need some of the skills of a data scientist and ML engineer. As a software developer who works with NLP solutions, you are more likely to develop code that interfaces directly with an ML service (such as Amazon Comprehend). NLP developers often rely on either a data scientist or an ML engineer for their expertise with identifying and optimizing appropriate data sources and models.

Activity: NLP jobs scavenger hunt



- Find a job description for one of the four NLP roles outlined in this module
- Good search locations include:
 - [Indeed](#)
 - [Glassdoor](#)
 - [Monster](#)
- Be prepared to present your findings to the class



27

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

In this educator-led activity, you will be challenged to find a job description for an NLP role that you would like to work as in the future.



The “What’s new at AWS” section introduces new and emerging technologies at AWS.

What is Amazon CodeWhisperer?



CodeWhisperer



AI-powered code generator for IDEs and code editors

- **AI coding companion:**
 - Generates code suggestions based on comments and existing code
 - Offers real-time support for code authoring directly within your integrated development environment (IDE)
- **AI security scanner:**
 - Helps identify hard-to-find vulnerabilities
 - References multiple standards and best practices

Content processed by CodeWhisperer Professional is not stored or used for service improvement.

©2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

29

Amazon CodeWhisperer analyzes your comments and code as you write them in your integrated development environment (IDE). It goes beyond code completion by using natural language processing to comprehend the comments in your code. By understanding English comments, CodeWhisperer generates complete functions and code blocks that align with your descriptions. CodeWhisperer also analyzes the surrounding code, ensuring the generated code matches your style and naming conventions and seamlessly integrates into the existing context.

When scanning for security vulnerabilities, CodeWhisperer assesses your code against multiple sets of standards and best practices. This includes the following:

- Open Worldwide Application Security Project (OWASP) standards
- Crypto library best practices
- AWS security standards

The security scan feature is continuously updated to help keep applications free from new security vulnerabilities.

Compatibility: CodeWhisperer integrates with popular tools such as Visual Studio Code, JetBrains IDEs (IntelliJ IDEA, PyCharm, etc.), Amazon SageMaker Studio, JupyterLab, AWS Cloud9, and AWS Lambda console.

Support: CodeWhisperer supports a wide range of programming languages and development environments, including Python, Java, JavaScript, TypeScript, C#, Go, Rust, PHP, Ruby, Kotlin,

C, C++, shell scripting, structured query language (SQL), and Scala.

Installation: You can access CodeWhisperer by downloading and installing the AWS Toolkit IDE extension or plugin. You can also activate CodeWhisperer from directly within the AWS Lambda and AWS Cloud9 console code editors.

Installation instructions vary depending on the environment. For more information, see “Getting started” in the *CodeWhisperer User Guide* at <https://docs.aws.amazon.com/codewhisperer/latest/userguide/getting-started.html>.

Code generation

- Code suggestions
- Code completion
- Code generation from comments
- Alternate code suggestions
- Option to accept or reject
- Reference tracking for code that resembles open-source training data

```
1 import boto3
```

```
1 import boto3
2 # create an s3 bucket named cw95323
```

< 1/3 > Accept tab Accept Word ⌘ → ...

```
2 # create an s3 bucket named cw95323
  s3 = boto3.resource('s3')
  s3.create_bucket(Bucket='cw59323')
  # upload a file to the bucket
```



©2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

30

The code generation feature of CodeWhisperer offers code suggestions in real time in your development environment. It automatically offers code completion and code generation suggestions. It uses natural language processing of English comments in your code and an understanding of surrounding code to suggest whole lines of code, complete functions, and logical blocks of code. The generated code is aligned with your coding style and naming conventions. CodeWhisperer prioritizes secure coding and responsible artificial intelligence (responsible AI) practices. It's optimized for Amazon APIs and trained extensively on Amazon and open-source code. You have the option to accept the first suggestion, explore more suggestions, or continue writing your own code. It's important to review each code suggestion before accepting it because you might need to make edits to ensure that the suggestion aligns with your intended functionality.

User actions

- Previous and next suggestion: Use the left arrow and right arrow.
- Accept a suggestion: Press Tab.
- Reject a suggestion: Press Esc.
- Manually start code generation when typing a comment: On MacOS, press Option+C, and on Windows, press Alt+C.

Open Code Reference Log

CodeWhisperer learns from open-source projects and the code it suggests might occasionally resemble code samples from the training data. With the reference log, you can view references to code suggestions that are similar to the training data. When such occurrences happen, CodeWhisperer notifies you and provides repository and licensing information. Use this information to make decisions about whether to use the code in your project and properly attribute the source code as desired.

Security scan

```
1 import boto3
2
3 def upload_file(bucket_name, file_path):
4     s3 = boto3.client('s3')
5     with open(file_path, 'rb') as file:
6         s3.upload_fileobj(file, bucket_name, file_path)
7         print("File uploaded.")
8
9 bucket_name = input("Enter bucket name: ")
10 file_path = input("Enter file path to upload: ")
11 upload_file(bucket_name, file_path)
```

PROBLEMS



CWE-22 – Path traversal: Constructing path names with unsanitized user input can lead to...



©2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

31

The security scanning feature of CodeWhisperer detects security vulnerabilities in both CodeWhisperer-generated code and developer-written code. It scans the code to identify potential vulnerabilities and provides suggestions for remediation. This includes scanning for hard-to-find vulnerabilities that might be overlooked. The security scan is compatible with popular IDEs such as VS Code and JetBrains. It supports Python, Java, and JavaScript.

Benefits of Amazon CodeWhisperer



Value to developers

- Increase velocity.
- Spend less time writing code.
- Receive help directly within your IDE.
- Find security vulnerabilities in your code.



Value to organizations

- Use at all experience levels.
- Support open-source attribution.
- Reduce the risk of security vulnerabilities.
- Increase code quality and developer productivity.



©2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

32

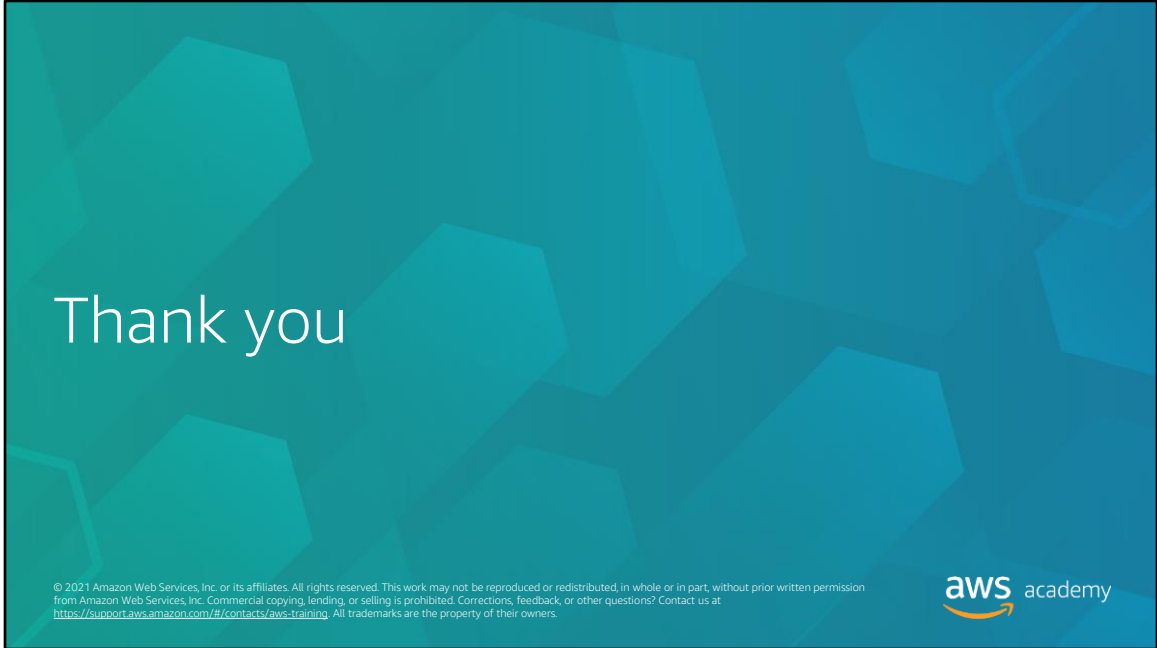
Automated code generation automates repetitive tasks and saves you time. It eliminates the need for you to invest excessive hours in exploring and learning new technologies. Instead, you can rely on high-quality code suggestions that match your coding style. This approach enhances your productivity so you can focus on critical tasks, which encourages innovation and progress in software development. With automated code generation, you can streamline your workflows and achieve significant time savings while ensuring the delivery of code that meets your standards.

CodeWhisperer code generation offers many benefits for software development organizations. It accelerates application development for faster delivery of software solutions. By automating repetitive tasks, it optimizes the use of developer time, so developers can focus on more critical aspects of the project. Additionally, code generation helps mitigate security vulnerabilities, safeguarding the integrity of the codebase.

CodeWhisperer also protects open source intellectual property by providing the open source reference tracker. CodeWhisperer enhances code quality and reliability, leading to robust and efficient applications. And it supports an efficient response to evolving software threats, keeping the codebase up to date with the latest security practices. CodeWhisperer has the potential to increase development speed, security, and the quality of software.

Resources:

- Visit “Getting started” at <https://aws.amazon.com/codewhisperer/resources/>.
- Dive deep with the “Amazon CodeWhisperer – Getting Started” course on AWS Skill Builder at <https://explore.skillbuilder.aws/learn/course/external/view/elearning/16405/amazon-codewhisperer-getting-started>.
- Learn how to build an event-driven serverless app at <https://catalog.us-east-1.prod.workshops.aws/workshops/a33a5d69-1417-4d5f-acc9-ae5c7fba665b/en-US/>.



Thank you for completing this module. In the next module, you will learn about a common approach for working on NLP applications.