Microsoft

Microsoft Official Course

AI-900T00

Microsoft Azure AI Fundamentals

AI-900T00
Microsoft Azure AI Fundamentals

1 http://www.microsoft.com/trademarks

MICROSOFT LICENSE TERMS

MICROSOFT INSTRUCTOR-LED COURSEWARE

These license terms are an agreement between Microsoft Corporation (or based on where you live, one of its affiliates) and you. Please read them. They apply to your use of the content accompanying this agreement which includes the media on which you received it, if any. These license terms also apply to Trainer Content and any updates and supplements for the Licensed Content unless other terms accompany those items. If so, those terms apply.

BY ACCESSING, DOWNLOADING OR USING THE LICENSED CONTENT, YOU ACCEPT THESE TERMS. IF YOU DO NOT ACCEPT THEM, DO NOT ACCESS, DOWNLOAD OR USE THE LICENSED CONTENT.

If you comply with these license terms, you have the rights below for each license you acquire.

1. **DEFINITIONS.**

    1. "Authorized Learning Center" means a Microsoft Imagine Academy (MSIA) Program Member, Microsoft Learning Competency Member, or such other entity as Microsoft may designate from time to time.
    2. "Authorized Training Session" means the instructor-led training class using Microsoft Instructor-Led Courseware conducted by a Trainer at or through an Authorized Learning Center.
    3. "Classroom Device" means one (1) dedicated, secure computer that an Authorized Learning Center owns or controls that is located at an Authorized Learning Center's training facilities that meets or exceeds the hardware level specified for the particular Microsoft Instructor-Led Courseware.
    4. "End User" means an individual who is (i) duly enrolled in and attending an Authorized Training Session or Private Training Session, (ii) an employee of an MPN Member (defined below), or (iii) a Microsoft full-time employee, a Microsoft Imagine Academy (MSIA) Program Member, or a Microsoft Learn for Educators – Validated Educator.
    5. "Licensed Content" means the content accompanying this agreement which may include the Microsoft Instructor-Led Courseware or Trainer Content.
    6. "Microsoft Certified Trainer" or "MCT" means an individual who is (i) engaged to teach a training session to End Users on behalf of an Authorized Learning Center or MPN Member, and (ii) currently certified as a Microsoft Certified Trainer under the Microsoft Certification Program.
    7. "Microsoft Instructor-Led Courseware" means the Microsoft-branded instructor-led training course that educates IT professionals, developers, students at an academic institution, and other learners on Microsoft technologies. A Microsoft Instructor-Led Courseware title may be branded as MOC, Microsoft Dynamics, or Microsoft Business Group courseware.
    8. "Microsoft Imagine Academy (MSIA) Program Member" means an active member of the Microsoft Imagine Academy Program.
    9. "Microsoft Learn for Educators – Validated Educator" means an educator who has been validated through the Microsoft Learn for Educators program as an active educator at a college, university, community college, polytechnic or K-12 institution.
    10. "Microsoft Learning Competency Member" means an active member of the Microsoft Partner Network program in good standing that currently holds the Learning Competency status.
    11. "MOC" means the "Official Microsoft Learning Product" instructor-led courseware known as Microsoft Official Course that educates IT professionals, developers, students at an academic institution, and other learners on Microsoft technologies.
    12. "MPN Member" means an active Microsoft Partner Network program member in good standing.

13. "Personal Device" means one (1) personal computer, device, workstation or other digital electronic device that you personally own or control that meets or exceeds the hardware level specified for the particular Microsoft Instructor-Led Courseware.
14. "Private Training Session" means the instructor-led training classes provided by MPN Members for corporate customers to teach a predefined learning objective using Microsoft Instructor-Led Courseware. These classes are not advertised or promoted to the general public and class attendance is restricted to individuals employed by or contracted by the corporate customer.
15. "Trainer" means (i) an academically accredited educator engaged by a Microsoft Imagine Academy Program Member to teach an Authorized Training Session, (ii) an academically accredited educator validated as a Microsoft Learn for Educators – Validated Educator, and/or (iii) a MCT.
16. "Trainer Content" means the trainer version of the Microsoft Instructor-Led Courseware and additional supplemental content designated solely for Trainers' use to teach a training session using the Microsoft Instructor-Led Courseware. Trainer Content may include Microsoft PowerPoint presentations, trainer preparation guide, train the trainer materials, Microsoft One Note packs, classroom setup guide and Pre-release course feedback form. To clarify, Trainer Content does not include any software, virtual hard disks or virtual machines.

2. **USE RIGHTS.** The Licensed Content is licensed, not sold. The Licensed Content is licensed on a **one copy per user basis**, such that you must acquire a license for each individual that accesses or uses the Licensed Content.

- 2.1 Below are five separate sets of use rights. Only one set of rights apply to you.

    1. **If you are a Microsoft Imagine Academy (MSIA) Program Member:**
        1. Each license acquired on behalf of yourself may only be used to review one (1) copy of the Microsoft Instructor-Led Courseware in the form provided to you. If the Microsoft Instructor-Led Courseware is in digital format, you may install one (1) copy on up to three (3) Personal Devices. You may not install the Microsoft Instructor-Led Courseware on a device you do not own or control.
        2. For each license you acquire on behalf of an End User or Trainer, you may either:
            1. distribute one (1) hard copy version of the Microsoft Instructor-Led Courseware to one (1) End User who is enrolled in the Authorized Training Session, and only immediately prior to the commencement of the Authorized Training Session that is the subject matter of the Microsoft Instructor-Led Courseware being provided, **or**
            2. provide one (1) End User with the unique redemption code and instructions on how they can access one (1) digital version of the Microsoft Instructor-Led Courseware, **or**
            3. provide one (1) Trainer with the unique redemption code and instructions on how they can access one (1) Trainer Content.
        3. For each license you acquire, you must comply with the following:
            1. you will only provide access to the Licensed Content to those individuals who have acquired a valid license to the Licensed Content,
            2. you will ensure each End User attending an Authorized Training Session has their own valid licensed copy of the Microsoft Instructor-Led Courseware that is the subject of the Authorized Training Session,
            3. you will ensure that each End User provided with the hard-copy version of the Microsoft Instructor-Led Courseware will be presented with a copy of this agreement and each End

User will agree that their use of the Microsoft Instructor-Led Courseware will be subject to the terms in this agreement prior to providing them with the Microsoft Instructor-Led Courseware. Each individual will be required to denote their acceptance of this agreement in a manner that is enforceable under local law prior to their accessing the Microsoft Instructor-Led Courseware,

4. you will ensure that each Trainer teaching an Authorized Training Session has their own valid licensed copy of the Trainer Content that is the subject of the Authorized Training Session,
5. you will only use qualified Trainers who have in-depth knowledge of and experience with the Microsoft technology that is the subject of the Microsoft Instructor-Led Courseware being taught for all your Authorized Training Sessions,
6. you will only deliver a maximum of 15 hours of training per week for each Authorized Training Session that uses a MOC title, and
7. you acknowledge that Trainers that are not MCTs will not have access to all of the trainer resources for the Microsoft Instructor-Led Courseware.

2. **If you are a Microsoft Learning Competency Member:**
   1. Each license acquire may only be used to review one (1) copy of the Microsoft Instructor-Led Courseware in the form provided to you. If the Microsoft Instructor-Led Courseware is in digital format, you may install one (1) copy on up to three (3) Personal Devices. You may not install the Microsoft Instructor-Led Courseware on a device you do not own or control.
   2. For each license you acquire on behalf of an End User or MCT, you may either:
      1. distribute one (1) hard copy version of the Microsoft Instructor-Led Courseware to one (1) End User attending the Authorized Training Session and only immediately prior to the commencement of the Authorized Training Session that is the subject matter of the Microsoft Instructor-Led Courseware provided, **or**
      2. provide one (1) End User attending the Authorized Training Session with the unique redemption code and instructions on how they can access one (1) digital version of the Microsoft Instructor-Led Courseware, **or**
      3. you will provide one (1) MCT with the unique redemption code and instructions on how they can access one (1) Trainer Content.
   3. For each license you acquire, you must comply with the following:
      1. you will only provide access to the Licensed Content to those individuals who have acquired a valid license to the Licensed Content,
      2. you will ensure that each End User attending an Authorized Training Session has their own valid licensed copy of the Microsoft Instructor-Led Courseware that is the subject of the Authorized Training Session,
      3. you will ensure that each End User provided with a hard-copy version of the Microsoft Instructor-Led Courseware will be presented with a copy of this agreement and each End User will agree that their use of the Microsoft Instructor-Led Courseware will be subject to the terms in this agreement prior to providing them with the Microsoft Instructor-Led Courseware. Each individual will be required to denote their acceptance of this agreement in a manner that is enforceable under local law prior to their accessing the Microsoft Instructor-Led Courseware,

4. you will ensure that each MCT teaching an Authorized Training Session has their own valid licensed copy of the Trainer Content that is the subject of the Authorized Training Session,
5. you will only use qualified MCTs who also hold the applicable Microsoft Certification credential that is the subject of the MOC title being taught for all your Authorized Training Sessions using MOC,
6. you will only provide access to the Microsoft Instructor-Led Courseware to End Users, and
7. you will only provide access to the Trainer Content to MCTs.

3. **If you are a MPN Member:**
   1. Each license acquired on behalf of yourself may only be used to review one (1) copy of the Microsoft Instructor-Led Courseware in the form provided to you. If the Microsoft Instructor-Led Courseware is in digital format, you may install one (1) copy on up to three (3) Personal Devices. You may not install the Microsoft Instructor-Led Courseware on a device you do not own or control.
   2. For each license you acquire on behalf of an End User or Trainer, you may either:
      1. distribute one (1) hard copy version of the Microsoft Instructor-Led Courseware to one (1) End User attending the Private Training Session, and only immediately prior to the commencement of the Private Training Session that is the subject matter of the Microsoft Instructor-Led Courseware being provided, **or**
      2. provide one (1) End User who is attending the Private Training Session with the unique redemption code and instructions on how they can access one (1) digital version of the Microsoft Instructor-Led Courseware, **or**
      3. you will provide one (1) Trainer who is teaching the Private Training Session with the unique redemption code and instructions on how they can access one (1) Trainer Content.
   3. For each license you acquire, you must comply with the following:
      1. you will only provide access to the Licensed Content to those individuals who have acquired a valid license to the Licensed Content,
      2. you will ensure that each End User attending an Private Training Session has their own valid licensed copy of the Microsoft Instructor-Led Courseware that is the subject of the Private Training Session,
      3. you will ensure that each End User provided with a hard copy version of the Microsoft Instructor-Led Courseware will be presented with a copy of this agreement and each End User will agree that their use of the Microsoft Instructor-Led Courseware will be subject to the terms in this agreement prior to providing them with the Microsoft Instructor-Led Courseware. Each individual will be required to denote their acceptance of this agreement in a manner that is enforceable under local law prior to their accessing the Microsoft Instructor-Led Courseware,
      4. you will ensure that each Trainer teaching an Private Training Session has their own valid licensed copy of the Trainer Content that is the subject of the Private Training Session,

5. you will only use qualified Trainers who hold the applicable Microsoft Certification credential that is the subject of the Microsoft Instructor-Led Courseware being taught for all your Private Training Sessions,

6. you will only use qualified MCTs who hold the applicable Microsoft Certification credential that is the subject of the MOC title being taught for all your Private Training Sessions using MOC,

7. you will only provide access to the Microsoft Instructor-Led Courseware to End Users, and

8. you will only provide access to the Trainer Content to Trainers.

4. **If you are an End User:**
For each license you acquire, you may use the Microsoft Instructor-Led Courseware solely for your personal training use. If the Microsoft Instructor-Led Courseware is in digital format, you may access the Microsoft Instructor-Led Courseware online using the unique redemption code provided to you by the training provider and install and use one (1) copy of the Microsoft Instructor-Led Courseware on up to three (3) Personal Devices. You may also print one (1) copy of the Microsoft Instructor-Led Courseware. You may not install the Microsoft Instructor-Led Courseware on a device you do not own or control.

5. **If you are a Trainer.**

1. For each license you acquire, you may install and use one (1) copy of the Trainer Content in the form provided to you on one (1) Personal Device solely to prepare and deliver an Authorized Training Session or Private Training Session, and install one (1) additional copy on another Personal Device as a backup copy, which may be used only to reinstall the Trainer Content. You may not install or use a copy of the Trainer Content on a device you do not own or control. You may also print one (1) copy of the Trainer Content solely to prepare for and deliver an Authorized Training Session or Private Training Session.

2. If you are an MCT, you may customize the written portions of the Trainer Content that are logically associated with instruction of a training session in accordance with the most recent version of the MCT agreement.

3. If you elect to exercise the foregoing rights, you agree to comply with the following: (i) customizations may only be used for teaching Authorized Training Sessions and Private Training Sessions, and (ii) all customizations will comply with this agreement. For clarity, any use of "customize" refers only to changing the order of slides and content, and/or not using all the slides or content, it does not mean changing or modifying any slide or content.

- 2.2 **Separation of Components.** The Licensed Content is licensed as a single unit and you may not separate their components and install them on different devices.
- 2.3 **Redistribution of Licensed Content.** Except as expressly provided in the use rights above, you may not distribute any Licensed Content or any portion thereof (including any permitted modifications) to any third parties without the express written permission of Microsoft.
- 2.4 **Third Party Notices.** The Licensed Content may include third party code that Microsoft, not the third party, licenses to you under this agreement. Notices, if any, for the third party code are included for your information only.
- 2.5 **Additional Terms.** Some Licensed Content may contain components with additional terms, conditions, and licenses regarding its use. Any non-conflicting terms in those conditions and licenses also apply to your use of that respective component and supplements the terms described in this agreement.

3. **LICENSED CONTENT BASED ON PRE-RELEASE TECHNOLOGY.** If the Licensed Content's subject matter is based on a pre-release version of Microsoft technology ("**Pre-release**"), then in addition to the other provisions in this agreement, these terms also apply:

    1. **Pre-Release Licensed Content.** This Licensed Content subject matter is on the Pre-release version of the Microsoft technology. The technology may not work the way a final version of the technology will and we may change the technology for the final version. We also may not release a final version. Licensed Content based on the final version of the technology may not contain the same information as the Licensed Content based on the Pre-release version. Microsoft is under no obligation to provide you with any further content, including any Licensed Content based on the final version of the technology.

    2. **Feedback.** If you agree to give feedback about the Licensed Content to Microsoft, either directly or through its third party designee, you give to Microsoft without charge, the right to use, share and commercialize your feedback in any way and for any purpose. You also give to third parties, without charge, any patent rights needed for their products, technologies and services to use or interface with any specific parts of a Microsoft technology, Microsoft product, or service that includes the feedback. You will not give feedback that is subject to a license that requires Microsoft to license its technology, technologies, or products to third parties because we include your feedback in them. These rights survive this agreement.

    3. **Pre-release Term.** If you are an Microsoft Imagine Academy Program Member, Microsoft Learning Competency Member, MPN Member, Microsoft Learn for Educators – Validated Educator, or Trainer, you will cease using all copies of the Licensed Content on the Pre-release technology upon (i) the date which Microsoft informs you is the end date for using the Licensed Content on the Pre-release technology, or (ii) sixty (60) days after the commercial release of the technology that is the subject of the Licensed Content, whichever is earliest ("**Pre-release term**"). Upon expiration or termination of the Pre-release term, you will irretrievably delete and destroy all copies of the Licensed Content in your possession or under your control.

4. **SCOPE OF LICENSE.** The Licensed Content is licensed, not sold. This agreement only gives you some rights to use the Licensed Content. Microsoft reserves all other rights. Unless applicable law gives you more rights despite this limitation, you may use the Licensed Content only as expressly permitted in this agreement. In doing so, you must comply with any technical limitations in the Licensed Content that only allows you to use it in certain ways. Except as expressly permitted in this agreement, you may not:

    - access or allow any individual to access the Licensed Content if they have not acquired a valid license for the Licensed Content,
    - alter, remove or obscure any copyright or other protective notices (including watermarks), branding or identifications contained in the Licensed Content,
    - modify or create a derivative work of any Licensed Content,
    - publicly display, or make the Licensed Content available for others to access or use,
    - copy, print, install, sell, publish, transmit, lend, adapt, reuse, link to or post, make available or distribute the Licensed Content to any third party,
    - work around any technical limitations in the Licensed Content, or
    - reverse engineer, decompile, remove or otherwise thwart any protections or disassemble the Licensed Content except and only to the extent that applicable law expressly permits, despite this limitation.

5. **RESERVATION OF RIGHTS AND OWNERSHIP.** Microsoft reserves all rights not expressly granted to you in this agreement. The Licensed Content is protected by copyright and other intellectual property

laws and treaties. Microsoft or its suppliers own the title, copyright, and other intellectual property rights in the Licensed Content.

6. **EXPORT RESTRICTIONS.** The Licensed Content is subject to United States export laws and regulations. You must comply with all domestic and international export laws and regulations that apply to the Licensed Content. These laws include restrictions on destinations, end users and end use. For additional information, see www.microsoft.com/exporting.
7. **SUPPORT SERVICES.** Because the Licensed Content is provided "as is", we are not obligated to provide support services for it.
8. **TERMINATION.** Without prejudice to any other rights, Microsoft may terminate this agreement if you fail to comply with the terms and conditions of this agreement. Upon termination of this agreement for any reason, you will immediately stop all use of and delete and destroy all copies of the Licensed Content in your possession or under your control.
9. **LINKS TO THIRD PARTY SITES.** You may link to third party sites through the use of the Licensed Content. The third party sites are not under the control of Microsoft, and Microsoft is not responsible for the contents of any third party sites, any links contained in third party sites, or any changes or updates to third party sites. Microsoft is not responsible for webcasting or any other form of transmission received from any third party sites. Microsoft is providing these links to third party sites to you only as a convenience, and the inclusion of any link does not imply an endorsement by Microsoft of the third party site.
10. **ENTIRE AGREEMENT.** This agreement, and any additional terms for the Trainer Content, updates and supplements are the entire agreement for the Licensed Content, updates and supplements.
11. **APPLICABLE LAW.**
    1. United States. If you acquired the Licensed Content in the United States, Washington state law governs the interpretation of this agreement and applies to claims for breach of it, regardless of conflict of laws principles. The laws of the state where you live govern all other claims, including claims under state consumer protection laws, unfair competition laws, and in tort.
    2. Outside the United States. If you acquired the Licensed Content in any other country, the laws of that country apply.
12. **LEGAL EFFECT.** This agreement describes certain legal rights. You may have other rights under the laws of your country. You may also have rights with respect to the party from whom you acquired the Licensed Content. This agreement does not change your rights under the laws of your country if the laws of your country do not permit it to do so.
13. **DISCLAIMER OF WARRANTY. THE LICENSED CONTENT IS LICENSED "AS-IS" AND "AS AVAILABLE." YOU BEAR THE RISK OF USING IT. MICROSOFT AND ITS RESPECTIVE AFFILIATES GIVES NO EXPRESS WARRANTIES, GUARANTEES, OR CONDITIONS. YOU MAY HAVE ADDITIONAL CONSUMER RIGHTS UNDER YOUR LOCAL LAWS WHICH THIS AGREEMENT CANNOT CHANGE. TO THE EXTENT PERMITTED UNDER YOUR LOCAL LAWS, MICROSOFT AND ITS RESPECTIVE AFFILIATES EXCLUDES ANY IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NON-INFRINGEMENT.**
14. **LIMITATION ON AND EXCLUSION OF REMEDIES AND DAMAGES. YOU CAN RECOVER FROM MICROSOFT, ITS RESPECTIVE AFFILIATES AND ITS SUPPLIERS ONLY DIRECT DAMAGES UP TO US$5.00. YOU CANNOT RECOVER ANY OTHER DAMAGES, INCLUDING CONSEQUENTIAL, LOST PROFITS, SPECIAL, INDIRECT OR INCIDENTAL DAMAGES.**

This limitation applies to

- anything related to the Licensed Content, services, content (including code) on third party Internet sites or third-party programs; and
- claims for breach of contract, breach of warranty, guarantee or condition, strict liability, negligence, or other tort to the extent permitted by applicable law.

It also applies even if Microsoft knew or should have known about the possibility of the damages. The above limitation or exclusion may not apply to you because your country may not allow the exclusion or limitation of incidental, consequential, or other damages.

Please note: As this Licensed Content is distributed in Quebec, Canada, some of the clauses in this agreement are provided below in French.

Remarque : Ce le contenu sous licence étant distribué au Québec, Canada, certaines des clauses dans ce contrat sont fournies ci-dessous en français.

EXONÉRATION DE GARANTIE. Le contenu sous licence visé par une licence est offert « tel quel ». Toute utilisation de ce contenu sous licence est à votre seule risque et péril. Microsoft n'accorde aucune autre garantie expresse. Vous pouvez bénéficier de droits additionnels en vertu du droit local sur la protection dues consommateurs, que ce contrat ne peut modifier. La ou elles sont permises par le droit locale, les garanties implicites de qualité marchande, d'adéquation à un usage particulier et d'absence de contrefaçon sont exclues.

LIMITATION DES DOMMAGES-INTÉRÊTS ET EXCLUSION DE RESPONSABILITÉ POUR LES DOMMAGES. Vous pouvez obtenir de Microsoft et de ses fournisseurs une indemnisation en cas de dommages directs uniquement à hauteur de 5,00 $ US. Vous ne pouvez prétendre à aucune indemnisation pour les autres dommages, y compris les dommages spéciaux, indirects ou accessoires et pertes de bénéfices.

Cette limitation concerne:

- tout ce qui est relié au le contenu sous licence, aux services ou au contenu (y compris le code) figurant sur des sites Internet tiers ou dans des programmes tiers; et.
- les réclamations au titre de violation de contrat ou de garantie, ou au titre de responsabilité stricte, de négligence ou d'une autre faute dans la limite autorisée par la loi en vigueur.

Elle s'applique également, même si Microsoft connaissait ou devrait connaître l'éventualité d'un tel dommage. Si votre pays n'autorise pas l'exclusion ou la limitation de responsabilité pour les dommages indirects, accessoires ou de quelque nature que ce soit, il se peut que la limitation ou l'exclusion ci-dessus ne s'appliquera pas à votre égard.

EFFET JURIDIQUE. Le présent contrat décrit certains droits juridiques. Vous pourriez avoir d'autres droits prévus par les lois de votre pays. Le présent contrat ne modifie pas les droits que vous confèrent les lois de votre pays si celles-ci ne le permettent pas.

Revised April 2019

Contents

Module 0 Welcome

Welcome to the course

About this Course

Welcome to this course on *Azure AI Fundamentals*!

This course is designed for anyone who wants to learn about artificial intelligence (AI) and the services in Microsoft Azure that you can use to build AI solutions. The course provides a practical, hands-on approach in which you will get a chance to see AI in action and try Azure AI services for yourself.

The materials in this workbook are designed to be used alongside online modules in **Microsoft Learn**[1]. Throughout the course, you'll find references to specific Learn modules that you should use to supplement the information here.

Learning objectives

After completing this course, you will be able to:

- Describe Artificial Intelligence workloads and considerations.
- Describe fundamental principles of machine learning on Azure.
- Describe features of computer vision workloads on Azure.
- Describe features of Natural Language Processing (NLP) workloads on Azure.

Course Agenda

This course includes the following modules:

Module	Lessons
Explore Fundamentals of Artificial Intelligence	- Introduction to Artificial Intelligence - Artificial Intelligence in Microsoft Azure

1 https://docs.microsoft.com/learn/certifications/azure-ai-fundamentals

Module	Lessons
Explore Fundamentals of Machine Learning	- Introduction to Machine Learning - Azure Machine Learning
Explore Fundamentals of Computer Vision	- Computer Vision Concepts - Creating Computer Vision solutions in Azure
Explore Fundamentals of Natural Language Processing	- Introduction to Natural Language Processing - Building Natural Language Solutions in Azure

Lab environment

Labs in this course are based on exercises in Microsoft Learn. You will be provided with an Azure subscription for use in this class. Your instructor will provide details. ,

Module 1 Introduction to AI

Artificial Intelligence in Azure

What is Artificial Intelligence?

AI enables us to build amazing software that can improve health care, enable people to overcome physical disadvantages, empower smart infrastructure, create incredible entertainment experiences, and even save the planet!

Simply put, AI is the creation of software that imitates human behaviors and capabilities. Key elements include:

- Making decisions based on data and past experience
- Detecting anomalies
- Interpreting visual input
- Understanding written and spoken language
- Engaging in dialogs and conversations

Common Artificial Intelligence Workloads

Common AI-related workloads include:

- **Machine learning** - This is often the foundation for an AI system, and is the way we "teach" a computer model to make prediction and draw conclusions from data.
- **Anomaly detection** - The capability to automatically detect errors or unusual activity in a system.
- **Computer vision** - The capability of software to interpret the world visually through cameras, video, and images.
- **Natural language processing** - The capability for a computer to interpret written or spoken language, and respond in kind.
- **Knowledge mining** - The capability to extract information from data sources to create a searchable knowledge store.

Principles of Responsible AI

At Microsoft, AI software development is guided by a set of six principles, designed to ensure that AI applications provide amazing solutions to difficult problems without any unintended negative consequences.

Fairness

AI systems should treat all people fairly. For example, suppose you create a machine learning model to support a loan approval application for a bank. The model should make predictions of whether or not the loan should be approved without incorporating any bias based on gender, ethnicity, or other factors that might result in an unfair advantage or disadvantage to specific groups of applicants.

Azure Machine Learning includes the capability to interpret models and quantify the extent to which each feature of the data influences the model's prediction. This capability helps data scientists and developers identify and mitigate bias in the model.

Reliability and safety

AI systems should perform reliably and safely. For example, consider an AI-based software system for an autonomous vehicle; or a machine learning model that diagnoses patient symptoms and recommends prescriptions. Unreliability in these kinds of system can result in substantial risk to human life.

AI-based software application development must be subjected to rigorous testing and deployment management processes to ensure that they work as expected before release.

Privacy and security

AI systems should be secure and respect privacy. The machine learning models on which AI systems are based rely on large volumes of data, which may contain personal details that must be kept private. Even after the models are trained and the system is in production, it uses new data to make predictions or take action that may be subject to privacy or security concerns.

Inclusiveness

AI systems should empower everyone and engage people. AI should bring benefits to all parts of society, regardless of physical ability, gender, sexual orientation, ethnicity, or other factors.

Transparency

AI systems should be understandable. Users should be made fully aware of the purpose of the system, how it works, and what limitations may be expected.

Accountability

People should be accountable for AI systems. Designers and developers of AI-based solution should work within a framework of governance and organizational principles that ensure the solution meets ethical and legal standards that are clearly defined.

Note: For more information about Microsoft's principles for responsible AI, visit **the Microsoft responsible AI site**[1].

1 https://microsoft.com/ai/responsible-ai

Responsible AI

Azure basics

Microsoft Azure provides a scalable, reliable cloud platform for AI, including:

- **Data storage**: Azure Storage offers highly available, scalable, and secure storage for a variety of data objects in the cloud.
- **Compute**: Azure cloud compute provides the infrastructure to run applications and scale capacity on demand. A compute target is a designated compute resource or environment.
- **Services**: Azure services are delivered over the internet in a pay-as-you-go model. Services include servers, storage, databases, networking, software, analytics, and intelligence. You can learn more about **Azure services**[2] here.

Artificial Intelligence in Microsoft Azure

Some of the key AI-related services in Azure are described in this table:

Service	Description
Azure Machine Learning	A platform for training, deploying, and managing machine learning models
Cognitive Services	A suite of services with four main pillars: Vision, Speech, Language, Decision
Azure Bot Service	A cloud-based platform for developing and managing bots
Azure Cognitive Search	Data extraction, enrichment, and indexing for intelligent search and knowledge mining

Cognitive Services

In this lab, you will explore the Anomaly Detector cognitive service, which analyzes data over time to detect any unusual values.

1. Start the virtual machine for this lab or go to the exercise page at https://aka.ms/ai900-module-01.
2. Follow the instructions to complete the exercise on Microsoft Learn.

Explore Further on Microsoft Learn

To learn more about the concepts described in this module, review the **Get Started with Artificial Intelligence on Azure**[3] learning path on Microsoft Learn.

2 https://docs.microsoft.com/learn/modules/intro-to-azure-fundamentals/tour-of-azure-services
3 https://aka.ms/learn-artificial-intelligence

Module 2 Machine Learning

Introduction to Machine Learning

What is machine learning?

Machine Learning is the foundation for most AI solutions, and enables the creation of models that predict unknown values and infer insights from observed data.

So how do machines learn?

The answer is, from data. In today's world, we create huge volumes of data as we go about our everyday lives. From the text messages, emails, and social media posts we send to the photographs and videos we take on our phones, we generate massive amounts of information. More data still is created by millions of sensors in our homes, cars, cities, public transport infrastructure, and factories.

Data scientists can use all of that data to train machine learning models that can make predictions and inferences based on the relationships they find in the data.

For example, suppose an environmental conservation organization wants volunteers to identify and catalog different species of wildflower using a phone app. The following animation shows how machine learning can be used to enable this scenario.

1. A team of botanists and data scientists collects samples of wildflowers.
2. The team labels the samples with the correct species.
3. The labeled data is processed using an algorithm that finds relationships between the features of the samples and the labeled species.
4. The results of the algorithm are encapsulated in a model.
5. When new samples are found by volunteers, the model can identify the correct species label.

Types of machine learning

There are two general approaches to machine learning, supervised and unsupervised machine learning.

Both have a goal of capturing a model, or equation, that will try to predict a result, or *label*, that is as close to the actual result as possible.

Both approaches require an input of data. We call different groups of data, *features*.

Supervised machine learning relies on us having some data with known label values that we can fit to a model, which can then be applied to new data for which we don’t have the label values. Labels are what we want to predict, such as house price, an item in an image, or if someone has diabetes.

Two types of supervised machine learning tasks include:

- **Regression**: used predict a continuous numeric value; like a price, a sales total, or some other measure.
- **Classification**: used to determine the probability between 0 and 1 data belongs to a particular group, or class.

Unsupervised machine learning is an approach that trains a model to separate items only based on their features. There is no previously known cluster value, or label, from which to train the model.

One type of unsupervised machine learning task is:

- **Clustering**: used to determine a label based on how similar the item is to the items in the label group.

Model training and validation

The learning component of machine learning occurs during training. We try to capture the relationships between the features and label in a model. Training is the action of iterating on an algorithm to best fit, or encapsulate those relationships.

After training, we have a model that we can test. We can use some of the data set aside, validation data, to test how closely our model's predicted labels are to actual labels.

There are many types of evaluation metrics. The important thing to remember is that the goal of machine learning is to find a model that gets as close as possible to predicting the actual label. The best model can still have some margin of error.

The training and validation process:

1. Split the data into a training set and a validation set
2. Apply an algorithm to fit the training data to a model
3. The trained model encapsulates the relationships in the data
4. Use the model to generate predictions from the validation data
5. Use evaluation metrics to compare predicted vs actual labels (supervised) or measure cluster separation (unsupervised)
6. Repeat…

Azure Machine Learning

What is Azure Machine Learning?

Microsoft Azure provides the **Azure Machine Learning** service - a cloud-based platform for creating, managing, and publishing machine learning models. Azure Machine Learning provides the following features and capabilities:

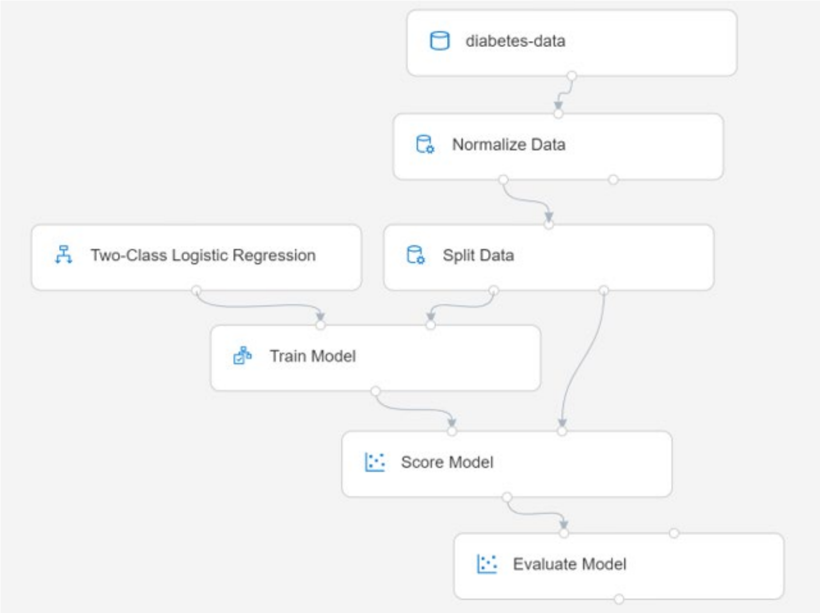
Feature	Capability
Automated machine learning	This feature enables non-experts to quickly create an effective machine learning model from data.
Azure Machine Learning designer	A graphical interface enabling no-code development of machine learning solutions.
Data and compute management	Cloud-based data storage and compute resources that professional data scientists can use to run data experiment code at scale.
Pipelines	Data scientists, software engineers, and IT operations professionals can define pipelines to orchestrate model training, deployment, and management tasks.

Automated Machine Learning

Automated Machine Learning in Azure Machine Learning provides the easiest way to train a machine learning model for regression or classification (or forecasting, which is really just regression with a time-series element). There's a visual interface for automated machine learning in the Azure Machine Learning studio web portal. You just need to supply the training data and select the required model type, and Azure Machine Learning does the rest.

Automated machine learning helps data scientists increase their efficiency by automating many of the time-consuming tasks associated with training models; and it enables them to use cloud-based compute resources that scale effectively to run multiple training experiments in parallel while incurring costs only when actually used.

Azure Machine Learning Designer



In Azure Machine Learning, multi-step workflows to prepare data, train models, and perform model management tasks are called pipelines. The *designer* tool in Azure Machine Learning studio enables you to create and run pipelines by using a drag & drop visual interface to connect modules that define the steps and data flow for the pipeline.

Lab: Explore Machine Learning

In this lab, you will explore the Azure Machine Learning service's Automated Machine Learning capability to train a machine learning model.

1. Start the virtual machine for this lab or go to the exercise page at https://aka.ms/ai900-module-02.
2. Follow the instructions to complete the exercise on Microsoft Learn.

Explore Further on Microsoft Learn

To learn more about the concepts described in this module, review the online modules in the **Create no-code predictive models with Azure Machine Learning**[1] learning path on Microsoft Learn.

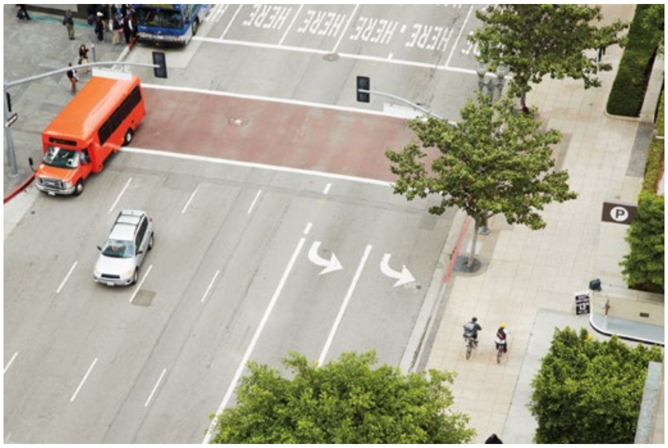
1 https://aka.ms/no-code-ml

Module 3 Computer Vision

Computer Vision Concepts

What is Computer Vision?

Computer vision is one of the core areas of artificial intelligence (AI), and focuses on creating solutions that enable AI-enabled applications to “see” the world and make sense of it.



Of course, computers don't have biological eyes that work the way ours do, but they are capable of processing images; either from a live camera feed or from digital photographs or videos. This ability to process images is the key to creating software that can emulate human visual perception.

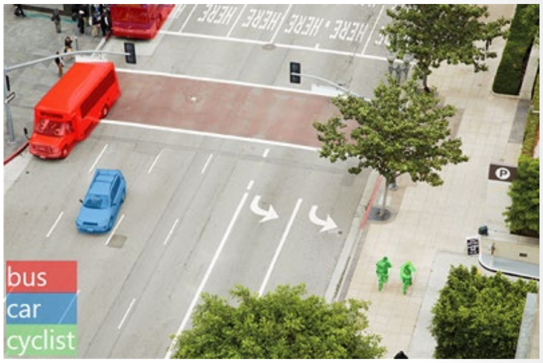
To an AI application, an image is just an array of pixel values. These numeric values can be used as *features* to train machine learning models that make predictions about the image and its contents.

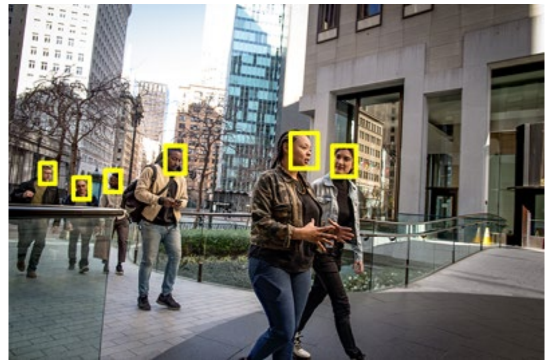
Applications of Computer Vision

Most computer vision solutions are based on machine learning models that can be applied to visual input from cameras, videos, or images.

The following table describes common applications of computer vision.

Task	Description
Image classification	 Image classification involves training a machine learning model to classify images based on their contents. For example, in a traffic monitoring solution you might use an image classification model to classify images based on the type of vehicle they contain, such as taxis, buses, cyclists, and so on.
Object detection	 Object detection machine learning models are trained to classify individual objects within an image, and identify their location with a bounding box. For example, a traffic monitoring solution might use object detection to identify the location of different classes of vehicle.

Task	Description
Semantic segmentation	 Semantic segmentation is an advanced machine learning technique in which individual pixels in the image are classified according to the object to which they belong. For example, a traffic monitoring solution might overlay traffic images with "mask" layers to highlight different vehicles using specific colors.
Image analysis	 You can create solutions that combine machine learning models with advanced image analysis techniques to extract information from images, including "tags" that could help catalog the image or even descriptive captions that summarize the scene shown in the image.

Task	Description
Face detection, analysis, and recognition	 Face detection is a specialized form of object detection that locates human faces in an image. This can be combined with classification and facial geometry analysis techniques to infer details such as age, and emotional state; and even recognize individuals based on their facial features.
Optical character recognition (OCR)	 Optical character recognition is a technique used to detect and read text in images. You can use OCR to read text in photographs (for example, road signs or store fronts) or to extract information from scanned documents such as letters, invoices, or forms.

Azure Computer Vision

Computer vision services in Azure are described in this table:

Service	Description
Computer Vision	- Image analysis – automated captioning and tagging - Common object detection - Face detection - Smart cropping - Optical character recognition

Service	Description
Custom Vision	- Custom image classification - Custom object detection
Face	- Face detection and analysis - Facial identification and recognition
Form Recognizer	- Data extraction from forms, invoices, and other documents

Computer Vision in Azure

Image analysis with the Computer Vision service

The Computer Vision service is a cognitive service in Microsoft Azure that provides pre-built computer vision capabilities. The service can analyze images, and return detailed information about an image and the objects it depicts.

Azure resources for Computer Vision

To use the Computer Vision service, you need to create a resource for it in your Azure subscription. You can use either of the following resource types:

- **Computer Vision**: A specific resource for the Computer Vision service. Use this resource type if you don't intend to use any other cognitive services, or if you want to track utilization and costs for your Computer Vision resource separately.
- **Cognitive Services**: A general cognitive services resource that includes Computer Vision along with many other cognitive services; such as Text Analytics, Translator Text, and others. Use this resource type if you plan to use multiple cognitive services and want to simplify administration and development.

Whichever type of resource you choose to create, it will provide two pieces of information that you will need to use it:

- A **key** that is used to authenticate client applications.
- An **endpoint** that provides the HTTP address at which your resource can be accessed.

Analyzing images with the Computer Vision service

After you've created a suitable resource in your subscription, you can submit images to the Computer Vision service to perform a wide range of analytical tasks.

Describing an image

Computer Vision has the ability to analyze an image, evaluate the objects that are detected, and generate a human-readable phrase or sentence that can describe what was detected in the image. Depending on the image contents, the service may return multiple results, or phrases. Each returned phrase will have an associated confidence score, indicating how confident the algorithm is in the supplied description. The highest confidence phrases will be listed first.

To help you understand this concept, consider the following image of the Empire State building in New York. The returned phrases are listed below the image in the order of confidence.

- A black and white photo of a city
- A black and white photo of a large city
- A large white building in a city

Tagging visual features

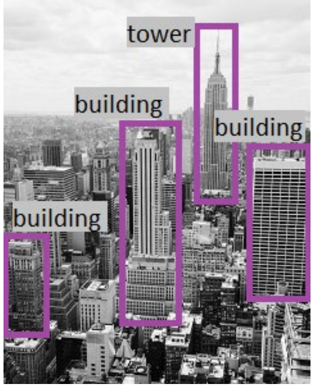
The image descriptions generated by Computer Vision are based on a set of thousands of recognizable objects, which can be used to suggest *tags* for the image. These tags can be associated with the image as metadata that summarizes attributes of the image; and can be particularly useful if you want to index an image along with a set of key terms that might be used to search for images with specific attributes or contents.

For example, the tags returned for the Empire State building image include:

- skyscraper
- tower
- building

Detecting objects

The object detection capability is similar to tagging, in that the service can identify common objects; but rather than tagging, or providing tags for the recognized objects only, this service can also return what is known as bounding box coordinates. Not only will you get the type of object, but you will also receive a set of coordinates that indicate the top, left, width, and height of the object detected, which you can use to identify the location of the object in the image, like this:



Detecting brands

This feature provides the ability to identify commercial brands. The service has an existing database of thousands of globally recognized logos from commercial brands of products.

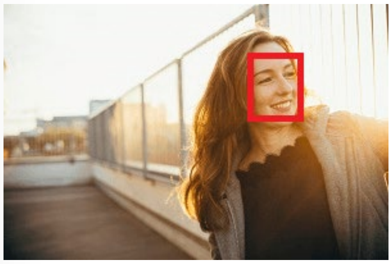
When you call the service and pass it an image, it performs a detection task and determine if any of the identified objects in the image are recognized brands. The service compares the brands against its database of popular brands spanning clothing, consumer electronics, and many more categories. If a known brand is detected, the service returns a response that contains the brand name, a confidence score (from 0 to 1 indicating how positive the identification is), and a bounding box (coordinates) for where in the image the detected brand was found.

For example, in the following image, a laptop has a Microsoft logo on its lid, which is identified and located by the Computer Vision service.

Detecting faces

The Computer Vision service can detect and analyze human faces in an image, including the ability to determine age and a bounding box rectangle for the location of the face(s). The facial analysis capabilities of the Computer Vision service are a subset of those provided by the dedicated **Face Service**[1]. If you need basic face detection and analysis, combined with general image analysis capabilities, you can use the Computer Vision service; but for more comprehensive facial analysis and facial recognition functionality, use the Face service.

The following example shows an image of a person with their face detected and approximate age estimated.



Categorizing an image

Computer Vision can categorize images based on their contents. The service uses a parent/child hierarchy with a "current" limited set of categories. When analyzing an image, detected objects are compared

1 https://docs.microsoft.com/azure/cognitive-services/face/

to the existing categories to determine the best way to provide the categorization. As an example, one of the parent categories is **people_**. This image of a person on a roof is assigned a category of **people_**.

A slightly different categorization is returned for the following image, which is assigned to the category **people_group** because there are multiple people in the image:

Review the 86-category list **here**[2].

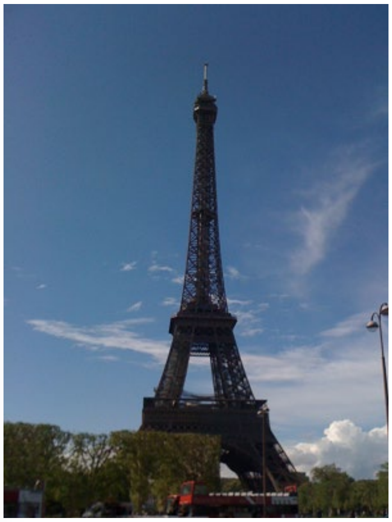
Detecting domain-specific content

When categorizing an image, the Computer Vision service supports two specialized domain models:

- **Celebrities** - The service includes a model that has been trained to identify thousands of well-known celebrities from the worlds of sports, entertainment, and business.
- **Landmarks** - The service can identify famous landmarks, such as the Taj Mahal and the Statue of Liberty.

For example, when analyzing the following image for landmarks, the Computer Vision service identifies the Eiffel Tower, with a confidence of 99.41%.

2 https://docs.microsoft.com/azure/cognitive-services/computer-vision/category-taxonomy



Training Models with the Custom Vision Service

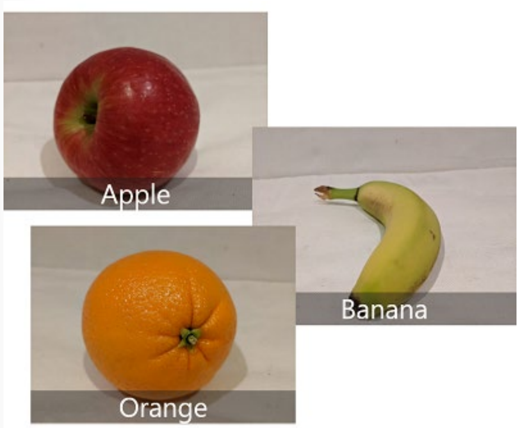
Most modern image classification solutions are based on *deep learning* techniques that make use of *convolutional neural networks* (CNNs) to uncover patterns in the pixels that correspond to particular classes. Training an effective CNN is a complex task that requires considerable expertise in data science and machine learning.

Common techniques used to train image classification models have been encapsulated into the **Custom Vision** cognitive service in Microsoft Azure; making it easy to train a model and publish it as a software service with minimal knowledge of deep learning techniques.

The Custom Vision cognitive services enables you to train and deploy a custom model for either image classification or object detection.

Image Classification

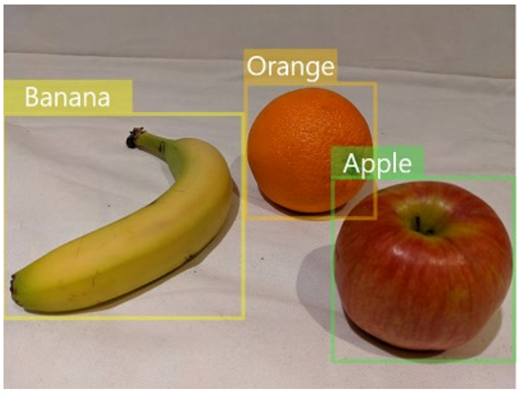
Image classification is a machine learning technique in which the object being classified is an image, such as a photograph.



As with any form of classification, creating an image classification solution involves training a model using a set of existing data for which the class is already known. In this case, the existing data consists of a set of categorized images, which you must upload to the Custom Vision service and tag with appropriate class labels. After training the model, you can publish it as a service for applications to use.

Object Detection

Object detection is a form of machine learning based computer vision in which a model is trained to recognize individual types of object in an image, and to identify their location in the image.



Creating an object detection solution with Custom Vision consists of three main tasks. First you must use upload and tag images, then you can train the model, and finally you must publish the model so that client applications can use it to locate objects in images.

Azure resources for Custom Vision

Creating an image classification solution with Custom Vision consists of two main tasks. First you must use existing images to train the model, and then you must publish the model so that client applications can use it to generate predictions.

For each of these tasks, you need a resource in your Azure subscription. You can use the following types of resource:

- **Custom Vision**: A dedicated resource for the custom vision service, which can be *training*, a *prediction*, or *both* resources.
- **Cognitive Services**: A general cognitive services resource that includes Custom Vision along with many other cognitive services. You can use this type of resource for *training*, *prediction*, or both.

The separation of training and prediction resources is useful when you want to track resource utilization for model training separately from client applications using the model to predict image classes. However, it can make development of an image classification solution a little confusing.

The simplest approach is to use a general Cognitive Services resource for both training and prediction. This means you only need to concern yourself with one *endpoint* (the HTTP address at which your service is hosted) and *key* (a secret value used by client applications to authenticate themselves).

If you choose to create a Custom Vision resource, you will be prompted to choose *training*, *prediction*, or *both* - and it's important to note that if you choose "both", then ***two*** resources are created - one for training and one for prediction.

It's also possible to take a mix-and-match approach in which you use a dedicated Custom Vision resource for training, but deploy your model to a Cognitive Services resource for prediction. For this to work, the training and prediction resources must be created in the same region.

Model training

To train a classification model, you must upload images to your training resource and label them with the appropriate class labels. Then, you must train the model and evaluate the training results.

You can perform these tasks in the *Custom Vision portal*, or if you have the necessary coding experience you can use one of the Custom Vision service programming language-specific software development kits (SDKs).

One of the key considerations when using images for classification, is to ensure that you have sufficient images of the objects in question and those images should be of the object from many different angles.

Model evaluation

Model training process is an iterative process in which the Custom Vision service repeatedly trains the model using some of the data, but holds some back to evaluate the model. At the end of the training process, the performance for the trained model is indicated by the following evaluation metrics:

- **Precision**: What percentage of the class predictions made by the model were correct? For example, if the model predicted that 10 images are oranges, of which eight were actually oranges, then the precision is 0.8 (80%).
- **Recall**: What percentage of class predictions did the model correctly identify? For example, if there are 10 images of apples, and the model found 7 of them, then the recall is 0.7 (70%).
- **Average Precision (AP)**: An overall metric that takes into account both precision and recall).

Using the model for prediction

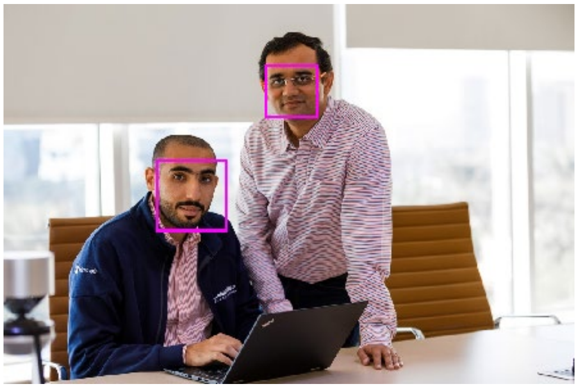
After you've trained the model, and you're satisfied with its evaluated performance, you can publish the model to your prediction resource. When you publish the model, you can assign it a name (the default is "Iteration*X*", where X is the number of times you have trained the model).

Analyzing Faces with the Face Service

Face detection and analysis is an area of artificial intelligence (AI) in which we use algorithms to locate and analyze human faces in images or video content.

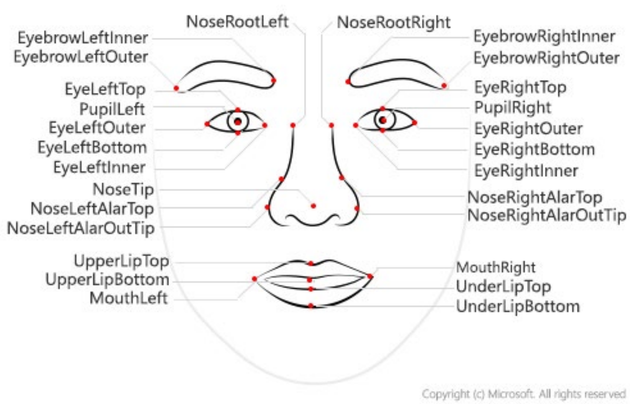
Face detection

Face detection involves identifying regions of an image that contain a human face, typically by returning *bounding box* coordinates that form a rectangle around the face, like this:

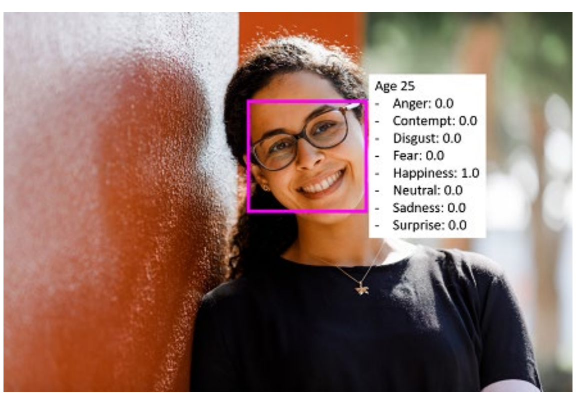


Facial analysis

Moving beyond simple face detection, some algorithms can also return other information, such as facial landmarks (nose, eyes, eyebrows, lips, and others).



These facial landmarks can be used as features with which to train a machine learning model from which you can infer information about a person, such as their age or peceived emotional state, like this:



Facial recognition

A further application of facial analysis is to train a machine learning model to identify known individuals from their facial features. This usage is more generally known as *facial recognition*, and involves using multiple images of each person you want to recognize to train a model so that it can detect those individuals in new images on which it wasn't trained.

Uses of face detection and analysis

There are many applications for face detection, analysis, and recognition. For example,

- Security - facial recognition can be used in building security applications, and increasingly it is used in smart phones operating systems for unlocking devices.
- Social media - facial recognition can be used to automatically tag known friends in photographs.
- Intelligent monitoring - for example, an automobile might include a system that monitors the driver's face to determine if the driver is looking at the road, looking at a mobile device, or shows signs of tiredness.
- Advertising - analyzing faces in an image can help direct advertisements to an appropriate demographic audience.
- Missing persons - using public cameras systems, facial recognition can be used to identify if a missing person is in the image frame.
- Identity validation - useful at ports of entry kiosks where a person holds a special entry permit.

Microsoft Azure's Face service

Microsoft Azure provides multiple cognitive services that you can use to detect and analyze faces, including:

- **Computer Vision**, which offers face detection and some basic face analysis, such as determining age.
- **Video Indexer**, which you can use to detect and identify faces in a video.
- **Face**, which offers pre-built algorithms that can detect, recognize, and analyze faces.

Of these, Face offers the widest range of facial analysis capabilities, so we'll focus on that service in this module.

Face

Face currently supports the following functionality:

- Face Detection
- Face Verification

- Find Similar Faces
- Group faces based on similarities
- Identify people

Face can return the rectangle coordinates for any human faces that are found in an image, as well as a series of attributes related to those faces such as:

- **Age**: a guess at an age
- **Blur**: how blurred the face is (which can be an indication of how likely the face is to be the main focus of the image)
- **Emotion**: what emotion is displayed
- **Exposure**: aspects such as underexposed or over exposed and applies to the face in the image and not the overall image exposure
- **Facial hair**: the estimated facial hair presence
- **Glasses**: if the person is wearing glasses
- **Hair**: the hair type and hair color
- **Head pose**: the face's orientation in a 3D space
- **Makeup**: whether the face in the image has makeup applied
- **Noise**: refers to visual noise in the image. If you have taken a photo with a high ISO setting for darker settings, you would notice this noise in the image. The image looks grainy or full of tiny dots that make the image less clear
- **Occlusion**: determines if there may be objects blocking the face in the image
- **Smile**: whether the person in the image is smiling

Azure resources for Face

To use Face, you must create one of the following types of resource in your Azure subscription:

- **Face**: Use this specific resource type if you don't intend to use any other cognitive services, or if you want to track utilization and costs for Face separately.
- **Cognitive Services**: A general cognitive services resource that includes Computer Vision along with many other cognitive services; such as Computer Vision, Text Analytics, Translator Text, and others. Use this resource type if you plan to use multiple cognitive services and want to simplify administration and development.

Whichever type of resource you choose to create, it will provide two pieces of information that you will need to use it:

- A **key** that is used to authenticate client applications.
- An **endpoint** that provides the HTTP address at which your resource can be accessed.

Reading Text with the Computer Vision Service

The ability for computer systems to process written or printed text is an area of artificial intelligence (AI) where *computer vision* intersects with *natural language processing*. You need computer vision capabilities to “read” the text, and then you need natural language processing capabilities to make sense of it.

The basic foundation of processing printed text is *optical character recognition* (OCR), in which a model can be trained to recognize individual shapes as letters, numerals, punctuation, or other elements of text. Much of the early work on implementing this kind of capability was performed by postal services to support automatic sorting of mail based on postal codes. Since then, the state-of-the-art for reading text has moved on, and it's now possible to build models that can detect printed or handwritten text in an image and read it line-by-line or even word-by-word.

Uses of OCR

The ability to recognize printed and handwritten text in images, is beneficial in many scenarios such as:

- note taking
- digitizing forms, such as medical records or historical documents
- scanning printed or handwritten checks for bank deposits

Use the Computer Vision service to read text

The ability to extract text from images is handled by the Computer Vision service, which also provides image analysis capabilities.

Azure resources for Computer Vision

The first step towards using the Computer Vision service is to create a resource for it in your Azure subscription. You can use either of the following resource types:

- **Computer Vision**: A specific resource for the Computer Vision service. Use this resource type if you don't intend to use any other cognitive services, or if you want to track utilization and costs for your Computer Vision resource separately.
- **Cognitive Services**: A general cognitive services resource that includes Computer Vision along with many other cognitive services; such as Text Analytics, Translator Text, and others. Use this resource type if you plan to use multiple cognitive services and want to simplify administration and development.

Whichever type of resource you choose to create, it will provide two pieces of information that you will need to use it:

- A **key** that is used to authenticate client applications.
- An **endpoint** that provides the HTTP address at which your resource can be accessed.

Use the Computer Vision service to read text

Many times an image contains text. It can be typewritten text or handwritten. Some common examples are images with road signs, scanned documents that are in an image format such as JPEG or PNG file formats, or even just a picture taken of a white board that was used during a meeting.

The Computer Vision service provides two application programming interfaces (APIs) that you can use to read text in images: the **OCR** API and the **Read** API.

The OCR API

The OCR API is designed for quick extraction of small amounts of text in images. It operates synchronously to provide immediate results, and can recognize text in numerous languages.

When you use the OCR API to process an image, it returns a hierarchy of information that consists of:

- **Regions** in the image that contain text
- **Lines** of text in each region
- **Words** in each line of text

For each of these elements, the OCR API also returns *bounding box* coordinates that define a rectangle to indicate the location in the image where the region, line, or word appears.

The Read API

The OCR method can have issues with false positives when the image is considered text-dominate. The Read API uses the latest recognition models and is optimized for images that have a significant amount of text or has considerable visual noise.

The Read API is a better option for scanned documents that have a lot of text. The Read API also has the ability to automatically determine the proper recognition model to use, taking into consideration lines of text and supporting images with printed text as well as recognizing handwriting.

Because the Read API can work with larger documents, it works asynchronously so as not to block your application while it is reading the content and returning results to your application. This means that to use the Read API, your application must use a three-step process:

1. Submit an image to the API, and retrieve an *operation ID* in response.
2. Use the operation ID to check on the status of the image analysis operation, and wait until it has completed.
3. Retrieve the results of the operation.

The results from the Read API are arranged into the following hierarchy:

- **Pages** - One for each page of text, including information about the page size and orientation.
- **Lines** - The lines of text on a page.
- **Words** - The words in a line of text.

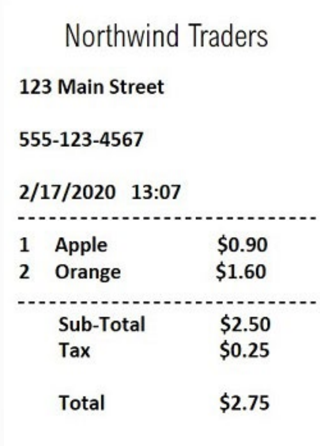
Each line and word includes bounding box coordinates indicating its position on the page.

Analyzing Forms with the Form Recognizer Service

A common problem in many organizations is the need to process receipt or invoice data. For example, a company might require expense claims to be submitted electronically with scanned receipts, or invoices might need to be digitized and routed to the correct accounts department.

It's relatively easy to scan receipts to create digital images or PDF documents, and it's possible to use optical character recognition (OCR) technologies to extract the text contents from the digitized documents. However, typically someone still needs to review the extracted text to make sense of the information it contains.

For example, consider the following receipt.


Northwind Traders

123 Main Street

555-123-4567

2/17/2020 13:07

1	Apple	$0.90
2	Orange	$1.60

Sub-Total	$2.50
Tax	$0.25
Total	$2.75

The receipt contains information that might be required for an expense claim, including:

- The name, address, and telephone number of the merchant.
- The date and time of the purchase.
- The quantity and price of each item purchased.
- The subtotal, tax, and total amounts.

Increasingly, organizations with large volumes of receipts and invoices to process are looking for artificial intelligence (AI) solutions that can not only extract the text data from receipts, but also intelligently interpret the information they contain.

Using the pre-built receipt model

The **Form Recognizer** in Azure provides intelligent form processing capabilities that you can use to automate the processing of data in documents such as forms, invoices, and receipts. It combines state-of-the-art optical character recognition (OCR) with predictive models that can interpret form data by:

- Matching field names to values.
- Processing tables of data.
- Identifying specific types of field, such as dates, telephone numbers, addresses, totals, and others.

Form Recognizer supports automated document processing through:

- **Pre-built models** that are provided out-of-the-box, and are trained to recognize and extract data from documents such as sales receipts.
- **Custom models**, which enable you to extract what are known as key/value pairs and table data from forms. Custom models are trained using your own data, which helps to tailor this model to your specific forms. Starting with only five samples of your forms, you can train the custom model. After the first training exercise, you can evaluate the results and consider if you need to add more samples and re-train.

Azure resources to access Form Recognizer services

To use the Form recognizer, you need to either create a **Form Recognizer** resource or a **Cognitive Services** resource in your Azure subscription. Both resource types give access to the Form Recognizer service.

After the resource has been created, you can create client applications that use its **key** and **endpoint** to connect submit forms for analysis.

Using the pre-built receipt model

Currently the pre-built receipt model is designed to recognize common receipts, in English, that are common to the USA. Examples are receipts used at restaurants, retail locations, and gas stations. The model is able to extract key information from the receipt slip:

- time of transaction
- date of transaction
- merchant information
- taxes paid
- receipt totals
- other pertinent information that may be present on the receipt
- all text on the receipt is recognized and returned as well

Lab: Analyze images with Computer Vision

In this lab, you will explore the Computer Vision cognitive service to analyze images.

1. Start the virtual machine for this lab or go to the exercise page at https://aka.ms/ai900-module-03.
2. Follow the instructions to complete the exercise on Microsoft Learn.

Explore Further on Microsoft Learn

To learn more about the concepts described in this module, review the modules in the **Explore computer vision in Microsoft Azure**[3] learning path on Microsoft Learn.

3 https://aka.ms/explore-computer-vision

Module 4 Natural Language Processing (NLP)

Introduction to Natural Language Processing

What is Natural Language Processing?

Natural language processing (NLP) is the area of AI that deals with creating software that understands written and spoken language.

NLP enables you to create software that can:

- Analyze text documents to extract key phrases and recognize entities (such as places, dates, or people).
- Perform *sentiment analysis* to determine how positive or negative the language used in a document is.
- Interpret spoken language, and synthesize speech responses.
- Automatically translate spoken or written phrases between languages.
- Interpret commands and determine appropriate actions.

Common NLP tasks:

Text analysis and entity recognition – Often you need to analyze a text document to determine its salient points or to identify entities it mentions, such as dates, places, people. For example, a company might use AI to analyze industry magazine articles to try to find articles that mention their products or executives or to determine the main subject of each article.

Sentiment analysis – This is a common form of text analysis that calculates a score indicating how positive (or negative) a text extract is. For example, a retailer might analyze reviews from customers to determine which ones are positive and which are negative.

Speech recognition and synthesis – It's increasingly common to encounter AI systems that can recognize spoken language as input and synthesize spoken output. For example, an in-car system might enable hands-free communication by reading incoming text messages aloud and enabling you to verbally dictate a response.

Machine translation – International and cross-cultural collaboration is often a key to success, and this requires the ability to eliminate language barriers. AI can be used to automate translation of written and spoken language. For example, an inbox add-in might be used to automatically translate incoming or outgoing emails, or a conference call presentation system might provide a simultaneous transcript of the speaker's words in multiple languages.

Semantic language modeling – Language can be complex and nuanced, so that multiple phrases might be used to mean the same thing. For example, a driver might ask "Where can I get gas near here?", "What's the location of the closest gas station?", or “Give me directions to a gas station.” All of these mean essentially the same thing, so a semantic understanding of the language being used is required to discern what the driver needs. An automobile manufacturer could train a language model to understand phrases like these and respond by displaying appropriate satellite navigation directions.

What is Conversational AI?

Conversational AI is a solution that enables a dialog between an AI agent and a human.

Generically, conversational AI agents are known as bots. Bots can engage over multiple channels:

- Web chat interfaces
- Email
- Social media platforms
- Voice

Conversational AI builds on other AI workloads, in particular natural language processing but also machine learning and potentially computer vision. In general, when people use the term “conversational AI”, they're referring to bots.

People often associate the term “bot” with a chat interface on a website, but actually this is just one (very common) way to interact with a bot. Bots can be connected to multiple channels, including email, social media, telephone and so on.

Natural Language Processing in Azure

In Microsoft Azure, you can use the following cognitive services to build natural language processing solutions:

Service	Capabilities
Language	- Language detection - Key phrase extraction - Entity detection - Sentiment analysis - Question answering - Conversational language understanding
Speech	- Text to speech - Speech to text - Speech translation
Translator	- Text Translation
Azure Bot Service	- Platform for conversational AI

Building Natural Language Processing Solutions in Azure

Analyzing Text

Analyzing text is a process where you evaluate different aspects of a document or phrase, in order to gain insights into the content of that text. For the most part, humans are able to read some text and understand the meaning behind it. Even without considering grammar rules for the language the text is written in, specific insights can be identified in the text.

As an example, you might read some text and identify some key phrases that indicate the main talking points of the text. You might also recognize names of people or well-known landmarks such as the Eiffel Tower. Although difficult at times, you might also be able to get a sense for how the person was feeling when they wrote the text, also commonly known as sentiment.

Text Analytics Techniques

Text analytics is a process where an artificial intelligence (AI) algorithm, running on a computer, evaluates these same attributes in text, to determine specific insights. A person will typically rely on their own experiences and knowledge to achieve the insights. A computer must be provided with similar knowledge to be able to perform the task. There are some commonly used techniques that can be used to build software to analyze text, including:

- Statistical analysis of terms used in the text. For example, removing common "stop words" (words like "the" or "a", which reveal little semantic information about the text), and performing *frequency analysis* of the remaining words (counting how often each word appears) can provide clues about the main subject of the text.
- Extending frequency analysis to multi-term phrases, commonly known as *N-grams* (a two-word phrase is a *bi-gram*, a three-word phrase is a *tri-gram*, and so on).
- Applying *stemming* or *lemmatization* algorithms to normalize words before counting them - for example, so that words like "power", "powered", and "powerful" are interpreted as being the same word.
- Applying linguistic structure rules to analyze sentences - for example, breaking down sentences into tree-like structures such as a *noun phrase*, which itself contains *nouns*, *verbs*, *adjectives*, and so on.
- Encoding words or terms as numeric features that can be used to train a machine learning model. For example, to classify a text document based on the terms it contains. This technique is often used to perform *sentiment analysis*, in which a document is classified as positive or negative.
- Creating *vectorized* models that capture semantic relationships between words by assigning them to locations in n-dimensional space. This modeling technique might, for example, assign values to the words "flower" and "plant" that locate them close to one another, while "skateboard" might be given a value that positions it much further away.

While these techniques can be used to great effect, programming them can be complex. In Microsoft Azure, the **Language** cognitive service can help simplify application development by using pre-trained models that can:

- Determine the language of a document or text (for example, French or English).
- Perform sentiment analysis on text to determine a positive or negative sentiment.
- Extract key phrases from text that might indicate its main talking points.

- Identify and categorize entities in the text. Entities can be people, places, organizations, or even everyday items such as dates, times, quantities, and so on.

Provisioning Azure resources

You can use the following resource to access this service:

- **Language Service**: A resource that enables you to build apps with industry-leading natural language understanding capabilities without machine learning expertise.

Speech Recognition and Synthesis

Increasingly, we expect artificial intelligence (AI) solutions to accept vocal commands and provide spoken responses. Consider the growing number of home and auto systems that you can control by speaking to them - issuing commands such as "turn off the lights", and soliciting verbal answers to questions such as "will it rain today?"

To enable this kind of interaction, the AI system must support two capabilities:

- **Speech recognition** - the ability to detect and interpret spoken input.
- **Speech synthesis** - the ability to generate spoken output.

Speech recognition

Speech recognition is concerned with taking the spoken word and converting it into data that can be processed - often by transcribing it into a text representation. The spoken words can be in the form of a recorded voice in an audio file, or live audio from a microphone. Speech patterns are analyzed in the audio to determine recognizable patterns that are mapped to words. To accomplish this feat, the software typically uses multiple types of models, including:

- An *acoustic* model that converts the audio signal into phonemes (representations of specific sounds).
- A *language* model that maps phonemes to words, usually using a statistical algorithm that predicts the most probable sequence of words based on the phonemes.

The recognized words are typically converted to text, which you can use for various purposes, such as.

- Providing closed captions for recorded or live videos
- Creating a transcript of a phone call or meeting
- Automated note dictation
- Determining intended user input for further processing

Speech synthesis

Speech synthesis is in many respects the reverse of speech recognition. It is concerned with vocalizing data, usually by converting text to speech. A speech synthesis solution typically requires the following information:

- The text to be spoken.
- The voice to be used to vocalize the speech.

To synthesize speech, the system typically *tokenizes* the text to break it down into individual words, and assigns phonetic sounds to each word. It then breaks the phonetic transcription into *prosodic* units (such as phrases, clauses, or sentences) to create phonemes that will be converted to audio format. These phonemes are then synthesized as audio by applying a voice, which will determine parameters such as pitch and timbre; and generating an audio wave form that can be output to a speaker or written to a file.

You can use the output of speech synthesis for many purposes, including:

- Generating spoken responses to user input.
- Creating voice menus for telephone systems.
- Reading email or text messages aloud in hands-free scenarios.
- Broadcasting announcements in public locations, such as railway stations or airports.

Provisioning Azure resources

To use the Speech service in an application, you must provision an appropriate resource in your Azure subscription. You can choose to provision either of the following types of resource:

- A **Speech** resource - choose this resource type if you only plan to use the Speech service, or if you want to manage access and billing for the resource separately from other services.
- A **Cognitive Services** resource - choose this resource type if you plan to use the Speech service in combination with other cognitive services, and you want to manage access and billing for these services together.

Translation

As organizations and individuals increasingly need to collaborate with people in other cultures and geographic locations, the removal of language barriers has become a significant problem.

One solution is to find bilingual, or even multilingual, people to translate between languages. However the scarcity of such skills, and the number of possible language combinations can make this approach difficult to scale. Increasingly, automated translation, sometimes known as *machine translation*, is being employed to solve this problem.

Literal and semantic translation

Early attempts at machine translation applied *literal* translations. A literal translation is where each word is translated to the corresponding word in the target language. This approach presents some issues. For one case, there may not be an equivalent word in the target language. Another case is where literal translation can change the meaning of the phrase or not get the context correct.

For example, the French phrase "*éteindre la lumière*" can be translated to English as "*turn off the light*". However, in French you might also say "*fermer la lumiere*" to mean the same thing. The French verb *fermer* literally means to "*close*", so a literal translation based only on the words would indicate, in English, "*close the light*"; which for the average English speaker, doesn't really make sense, so to be useful, a translation service should take into account the semantic context and return an English translation of "*turn off the light*".

Artificial intelligence systems must be able to understand, not only the words, but also the *semantic* context in which they are used. In this way, the service can return a more accurate translation of the input phrase or phrases. The grammar rules, formal versus informal, and colloquialisms all need to be considered.

Text and speech translation

Text translation can be used to translate documents from one language to another, translate email communications that come from foreign governments, and even provide the ability to translate web pages on the Internet. Many times you will see a *Translate* option for posts on social media sites, or the Bing search engine can offer to translate entire web pages that are turned in search results.

Speech translation is used to translate between spoken languages, sometimes directly (speech-to-speech translation) and sometimes by translating to an intermediary text format (speech-to-text translation).

Provisioning Azure resources

Microsoft Azure provides cognitive services that support translation. Specifically, you can use the following services:

- The **Translator** service, which supports text-to-text translation.
- The **Speech** service, which enables speech-to-text and speech-to-speech translation.

Alternatively, you can create a **Cognitive Services** resource that provides access to both services through a single Azure resource, consolidating billing and enabling applications to access both services through a single endpoint and authentication key.

Conversational Language Understanding

As artificial intelligence (AI) grows ever more sophisticated, conversational interactions with applications and digital assistants is becoming more and more common, and in specific scenarios can result in human-like interactions with AI agents. Common scenarios for this kind of solution include customer support applications, reservation systems, and home automation among others.

To enable these kinds of conversational solution, computers need not only to be able to accept language as input (either in text or audio format), but also to be able to interpret the semantic meaning of the input - in other words, *understand* what is being said.

On Microsoft Azure, language understanding is supported through the **Conversational Language Understanding** service. To work with the **Conversational Language Understanding** service, you need to take into account three core concepts: *utterances*, *entities*, and *intents*.

Utterances

An utterance is an example of something a user might say, and which your application must interpret. For example, when using a home automation system, a user might use the following utterances:

"*Switch the fan on.*"

"*Turn on the light.*"

Entities

An entity is an item to which an utterance refers. For example, **fan** and **light** in the following utterances:

"*Switch the **fan** on.*"

"*Turn on the **light**.*"

You can think of the **fan** and **light** entities as being specific instances of a general **device** entity.

Intents

An intent represents the purpose, or goal, expressed in a user's utterance. For example, for both of the previously considered utterances, the intent is to turn a device on; so in your Language Understanding application, you might define a **TurnOn** intent that is related to these utterances.

A Language Understanding application defines a model consisting of intents and entities. Utterances are used to train the model to identify the most likely intent and the entities to which it should be applied based on a given input. The home assistant application we've been considering might include multiple intents, like the following examples:

Intent	Related Utterances	Entities
Greeting	“Hello”	
	"Hi"	
	“Hey”	
	"Good morning"	
TurnOn	“Switch the fan on”	fan (device)
	"Turn the light on"	light (device)
	“Turn on the light”	light (device)
TurnOff	"Switch the fan off"	fan (device)
	“Turn the light off”	light (device)
	"Turn off the light"	light (device)
CheckWeather	“What is the weather for today?”	today (datetime)
	"Give me the weather forecast"	
	“What is the forecast for Paris?”	Paris (location)
	"What will the weather be like in Seattle tomorrow?"	Seattle (location), tomorrow (datetime)
None	“What is the meaning of life?”	
	"Is this thing on?"	

In this table there are numerous utterances used for each of the intents. The intent should be a concise way of grouping the utterance tasks. Of special interest is the ***None*** intent. You should consider always using the None intent to help handle utterances that do not map any of the utterances you have entered. The None intent is considered a fallback, and is typically used to provide a generic response to users when their requests don't match any other intent.

After defining the entities and intents with sample utterances in your application, you can train a language model to predict intents and entities from user input - even if it doesn't match the sample utterances exactly. You can then use the model from a client application to retrieve predictions and respond appropriately.

Creating intents

Define intents based on actions a user would want to perform with your application. For each intent, you should include a variety of utterances that provide examples of how a user might express the intent.

If an intent can be applied to multiple entities, be sure to include sample utterances for each potential entity; and ensure that each entity is identified in the utterance.

Training the model

After you have defined the intents and entities in your model, and included a suitable set of sample utterances; the next step is to train the model. Training is the process of using your sample utterances to teach your model to match natural language expressions that a user might say to probable intents and entities.

After training the model, you can test it by submitting text and reviewing the predicted intents. Training and testing is an iterative process. After you train your model, you test it with sample utterances to see if the intents and entities are recognized correctly. If they're not, make updates, retrain, and test again.

Predicting

When you are satisfied with the results from the training and testing, you can publish your Language Understanding application to a prediction resource for consumption.

Client applications can use the model by connecting to the endpoint for the prediction resource, specifying the appropriate authentication key; and submit user input to get predicted intents and entities. The predictions are returned to the client application, which can then take appropriate action based on the predicted intent.

Provisioning Azure Resources

Creating an application with Conversational Language Understanding consists of two main tasks. First you must define entities, intents, and utterances with which to train the language model - referred to as *authoring* the model. Then you must publish the model so that client applications can use it for intent and entity *prediction* based on user input.

For each of the authoring and prediction tasks, you need a resource in your Azure subscription. You can use the following types of resource:

- **Language Service**: A resource that enables you to build apps with industry-leading natural language understanding capabilities without machine learning expertise.
- **Cognitive Services**: A general cognitive services resource that includes Conversational Language Understanding along with many other cognitive services. You can only use this type of resource for *prediction*.

Custom Question Answering

You can use the Language Studio's custom question answering feature to create a knowledge base that consists of question-and-answer pairs. These questions and answers can be:

- Generated from an existing FAQ document or web page.
- Entered and edited manually.
- In many cases, a knowledge base is created using a combination of all of these techniques; starting with a base dataset of questions and answers from an existing FAQ document and extending the knowledge base with additional manual entries.

Questions in the knowledge base can be assigned alternative phrasing to help consolidate questions with the same meaning. For example, you might include a question like:

- *What is your head office location?*

You can anticipate different ways this question could be asked by adding an alternative phrasing such as:

- *Where is your head office located?*

Test the knowledge base

After creating a set of question-and-answer pairs, you must save it. This process analyzes your literal questions and answers and applies a built-in natural language processing model to match appropriate answers to questions, even when they are not phrased exactly as specified in your question definitions. Then you can use the built-in test interface in the Language Studio to test your knowledge base by submitting questions and reviewing the answers that are returned.

Use the knowledge base

When you're satisfied with your knowledge base, deploy it. Then you can use it over its REST interface. To access the knowledge base, client applications require:

- The knowledge base ID
- The knowledge base endpoint
- The knowledge base authorization key

Provisioning Azure resources

You can use the following resource to access this service:

- **Language**: A resource that enables you to build apps with industry-leading natural language under-standing capabilities without machine learning expertise.

Azure Bot Service

After you've created and published a knowledge base, you can use Azure Bot Service deliver it to users through a bot.

Create a bot for your knowledge base

You can create a custom bot by using the Microsoft Bot Framework SDK to write code that controls conversation flow and integrates with your QnA Maker knowledge base. However, an easier approach is to use the automatic bot creation functionality of QnA Maker, which enables you create a bot for your published knowledge base and publish it as an Azure Bot Service application with just a few clicks.

Extend and configure the bot

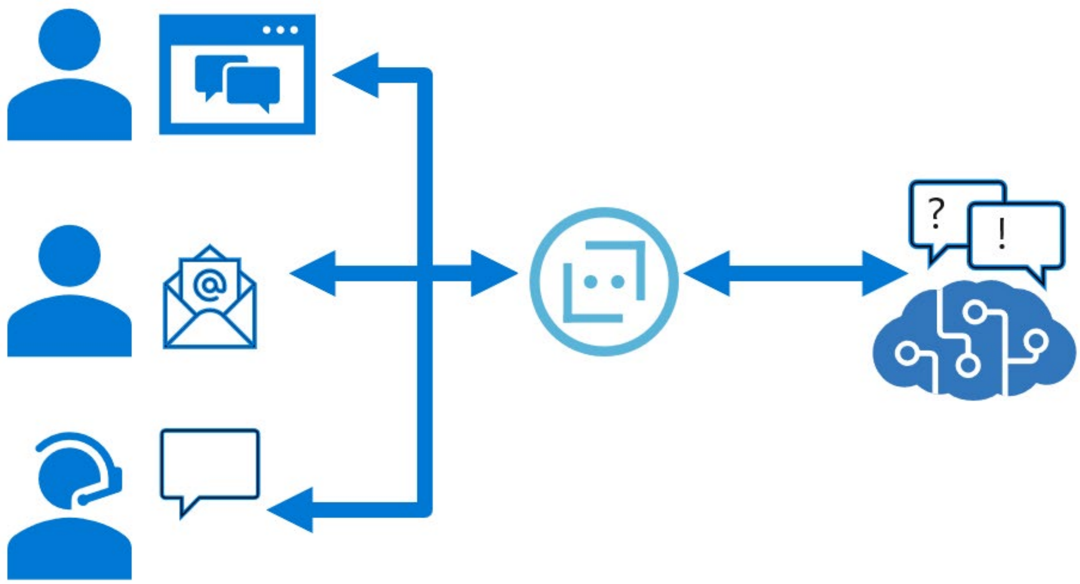
After creating your bot, you can manage it in the Azure portal, where you can:

- Extend the bot's functionality by adding custom code.
- Test the bot in an interactive test interface.
- Configure logging, analytics, and integration with other services.

For simple updates, you can edit bot code directly in the Azure portal. However, for more comprehensive customization, you can download the source code and edit it locally; republishing the bot directly to Azure when you're ready.

Connect channels

When your bot is ready to be delivered to users, you can connect it to multiple *channels*; making it possible for users to interact with it through web chat, email, Microsoft Teams, and other common communication media.



Users can submit questions to the bot through any of its channels, and receive an appropriate answer from the knowledge base on which the bot is based.

Provisioning Azure resources

You can use the following resource to access this service:

- **Azure Bot**: This service provides a framework for developing, publishing, and managing bots on Azure.

Lab: Analyze text with the Language Service

In this lab, you will use the Language cognitive service to analyze text.

1. Start the virtual machine for this lab or go to the exercise page at https://aka.ms/ai900-module-04.
2. Follow the instructions to complete the exercise on Microsoft Learn.

Explore Further on Microsoft Learn

To learn more about the concepts described in this module, review the modules in the **Explore natural language processing**[1] learning path on Microsoft Learn.

1 https://aka.ms/explore-nlp