In [35]:
```
import os
from dotenv import load_dotenv
load_dotenv()
```

Out[35]: True

In [36]:
```
google_api_key = os.getenv("GOOGLE_API_KEY")  # This will retrieve the API key from
```

In [37]:
```
if google_api_key:
    print("Google API Key loaded successfully.")
else:
    print("Failed to load Google API Key. Please check your .env file.")
```

```
Google API Key loaded successfully.
```

In [38]:
```
from llama_index.llms.gemini import Gemini
import google.generativeai as genai
from llama_index.core import SimpleDirectoryReader
from llama_index.llms.gemini import Gemini
from IPython.display import display, Markdown
from llama_index.core import VectorStoreIndex
from llama_index.core import ServiceContext
from llama_index.core import StorageContext, load_index_from_storage

from llama_index.embeddings.gemini import GeminiEmbedding
```

In [39]:
```
genai.configure(api_key=google_api_key)
```

In [40]:
```
for models in genai.list_models():
    print(models.name)
```

```
models/embedding-gecko-001
models/gemini-1.5-pro-latest
models/gemini-1.5-pro-002
models/gemini-1.5-pro
models/gemini-1.5-flash-latest
models/gemini-1.5-flash
models/gemini-1.5-flash-002
models/gemini-1.5-flash-8b
models/gemini-1.5-flash-8b-001
models/gemini-1.5-flash-8b-latest
models/gemini-2.5-pro-preview-03-25
models/gemini-2.5-flash-preview-05-20
models/gemini-2.5-flash
models/gemini-2.5-flash-lite-preview-06-17
models/gemini-2.5-pro-preview-05-06
models/gemini-2.5-pro-preview-06-05
models/gemini-2.5-pro
models/gemini-2.0-flash-exp
models/gemini-2.0-flash
models/gemini-2.0-flash-001
models/gemini-2.0-flash-exp-image-generation
models/gemini-2.0-flash-lite-001
models/gemini-2.0-flash-lite
models/gemini-2.0-flash-preview-image-generation
models/gemini-2.0-flash-lite-preview-02-05
models/gemini-2.0-flash-lite-preview
models/gemini-2.0-pro-exp
models/gemini-2.0-pro-exp-02-05
models/gemini-exp-1206
models/gemini-2.0-flash-thinking-exp-01-21
models/gemini-2.0-flash-thinking-exp
models/gemini-2.0-flash-thinking-exp-1219
models/gemini-2.5-flash-preview-tts
models/gemini-2.5-pro-preview-tts
models/learnlm-2.0-flash-experimental
models/gemma-3-1b-it
models/gemma-3-4b-it
models/gemma-3-12b-it
models/gemma-3-27b-it
models/gemma-3n-e4b-it
models/gemma-3n-e2b-it
models/gemini-2.5-flash-lite
models/embedding-001
models/text-embedding-004
models/gemini-embedding-exp-03-07
models/gemini-embedding-exp
models/gemini-embedding-001
models/aqa
models/imagen-3.0-generate-002
models/imagen-4.0-generate-preview-06-06
models/imagen-4.0-ultra-generate-preview-06-06
models/veo-2.0-generate-001
models/veo-3.0-generate-preview
models/veo-3.0-fast-generate-preview
models/gemini-2.5-flash-preview-native-audio-dialog
models/gemini-2.5-flash-exp-native-audio-thinking-dialog
```

```
models/gemini-2.0-flash-live-001
models/gemini-live-2.5-flash-preview
models/gemini-2.5-flash-live-preview
```

In [41]:
```
for models in genai.list_models():
    if 'generatecontent' in models.supported_generation_methods:
        print(models.name)
```

In [42]:
```
documents = SimpleDirectoryReader("D:\Generative AI(by sunny savita)\qafolder\Data"
```

In [43]:
```
documents
```

Out[43]:
```
<llama_index.core.readers.file.base.SimpleDirectoryReader at 0x1811af498e0>
```

In [44]:
```
doc = documents.load_data()
```

In [45]:
```
doc[0].text
```

Out[45]: 'What is machine learning?\nMachine learning is a branch of artificial intelligenc
e (AI) and computer science which\nfocuses on the use of data and algorithms to im
itate the way that humans learn,\ngradually improving its accuracy.\nIBM has a ric
h history with machine learning. One of its own, Arthur Samuel, is credited\nfor c
oining the term, “machine learning” with his research (link resides outside ibm.co
m)\naround the game of checkers. Robert Nealey, the self-proclaimed checkers maste
r,\nplayed the game on an IBM 7094 computer in 1962, and he lost to the compute
r.\nCompared to what can be done today, this feat seems trivial, but it’s consider
ed a major\nmilestone in the field of artificial intelligence.\nOver the last coup
le of decades, the technological advances in storage and processing\npower have en
abled some innovative products based on machine learning, such as\nNetflix’s recom
mendation engine and self-driving cars.\nMachine learning is an important componen
t of the growing field of data science.\nThrough the use of statistical methods, a
lgorithms are trained to make classifications or\npredictions, and to uncover key
insights in data mining projects. These insights\nsubsequently drive decision maki
ng within applications and businesses, ideally\nimpacting key growth metrics. As b
ig data continues to expand and grow, the market\ndemand for new data scientists w
ill increase. They will be required to help identify the\nmost relevant business q
uestions and the data to answer them.\nMachine learning algorithms are typically c
reated using frameworks such as Python that\naccelerate solution development by us
ing platforms like TensorFlow or PyTorch.\nNow available: watsonx.ai\nThe all-new
enterprise studio that brings together traditional machine learning along\nwith ne
w generative AI capabilities powered by foundation models.\nTry watsonx.ai\nBegin
your journey to AI\nLearn how to scale AI\nExplore the AI Academy\nMachine Learnin
g vs. Deep Learning vs. Neural Networks\nSince deep learning and machine learning
tend to be used interchangeably, it’s worth\nnoting the nuances between the two. M
achine learning, deep learning, and neural\nnetworks are all sub-fields of artific
ial intelligence. However, neural networks is actually\na sub-field of machine lea
rning, and deep learning is a sub-field of neural networks.\nThe way in which deep
learning and machine learning differ is in how each algorithm\nlearns. "Deep" mach
ine learning can use labeled datasets, also known as supervised\nlearning, to info
rm its algorithm, but it doesn’t necessarily require a labeled dataset. The\ndeep
learning process can ingest unstructured data in its raw form (e.g., text or image
s),\nand it can automatically determine the set of features which distinguish diff
erent\ncategories of data from one another. This eliminates some of the human inte
rvention\nrequired and enables the use of large amounts of data. You can think of
deep learning\nas "scalable machine learning" as Lex Fridman notes in this MIT lec
ture (link resides\noutside ibm.com).\nClassical, or "non-deep," machine learning
is more dependent on human intervention to\nlearn. Human experts determine the set
of features to understand the differences\nbetween data inputs, usually requiring
more structured data to learn.\nNeural networks, or artificial neural networks (AN
Ns), are comprised of node layers,\ncontaining an input layer, one or more hidden
layers, and an output layer. Each node, or\nartificial neuron, connects to another
and has an associated weight and threshold. If the\noutput of any individual node
is above the specified threshold value, that node is\nactivated, sending data to t
he next layer of the network. Otherwise, no data is passed\nalong to the next laye
r of the network by that node. The “deep” in deep learning is just\nreferring to t
he number of layers in a neural network. A neural network that consists of\nmore t
han three layers—which would be inclusive of the input and the output—can be\ncons
idered a deep learning algorithm or a deep neural network. A neural network that\n
only has three layers is just a basic neural network.\nDeep learning and neural ne
tworks are credited with accelerating progress in areas\nsuch as computer vision,
natural language processing, and speech recognition.\nSee the blog post “AI vs. Ma
chine Learning vs. Deep Learning vs. Neural Networks:\nWhat’s the Difference?” for
a closer look at how the different concepts relate.\nRelated content\nExplore the
watsonx.ai interactive demo\nDownload “Machine learning for Dummies”\n- This link

downloads a pdf\nExplore Gen AI for developers\nHow does machine learning work?\nU C Berkeley (link resides outside ibm.com) breaks out the learning system of a\nmac hine learning algorithm into three main parts.\nA Decision Process: In general, ma chine learning algorithms are used to make a\nprediction or classification. Based on some input data, which can be labeled or\nunlabeled, your algorithm will produc e an estimate about a pattern in the data.\nAn Error Function: An error function e valuates the prediction of the model. If\nthere are known examples, an error funct ion can make a comparison to assess\nthe accuracy of the model.\nA Model Optimizat ion Process: If the model can fit better to the data points in the\ntraining set, then weights are adjusted to reduce the discrepancy between the\nknown example and the model estimate. The algorithm will repeat this iterative\n“evaluate and optimi ze” process, updating weights autonomously until a\nthreshold of accuracy has been met.\nMachine learning methods\nMachine learning models fall into three primary ca tegories.\nSupervised machine learning\nSupervised learning, also known as supervi sed machine learning, is defined by its use\nof labeled datasets to train algorith ms to classify data or predict outcomes accurately.\nAs input data is fed into the model, the model adjusts its weights until it has been fitted\nappropriately. This occurs as part of the cross validation process to ensure that the\nmodel avoids ov erfitting or underfitting. Supervised learning helps organizations solve a\nvariet y of real-world problems at scale, such as classifying spam in a separate folder\n from your inbox. Some methods used in supervised learning include neural network s,\nnaïve bayes, linear regression, logistic regression, random forest, and suppor t vector\nmachine (SVM).\nUnsupervised machine learning\nUnsupervised learning, al so known as unsupervised machine learning, uses machine\nlearning algorithms to an alyze and cluster unlabeled datasets (subsets called clusters).\nThese algorithms discover hidden patterns or data groupings without the need for\nhuman interventio n. This method’s ability to discover similarities and differences in\ninformation make it ideal for exploratory data analysis, cross-selling strategies,\ncustomer s egmentation, and image and pattern recognition. It’s also used to reduce the\nnumb er of features in a model through the process of dimensionality reduction. Princip al\ncomponent analysis (PCA) and singular value decomposition (SVD) are two common \napproaches for this. Other algorithms used in unsupervised learning include neur al\nnetworks, k-means clustering, and probabilistic clustering methods.\nSemi-supe rvised learning\nSemi-supervised learning offers a happy medium between supervised and\nunsupervised learning. During training, it uses a smaller labeled data set to guide\nclassification and feature extraction from a larger, unlabeled data set. Se mi-supervised\nlearning can solve the problem of not having enough labeled data fo r a supervised\nlearning algorithm. It also helps if it’s too costly to label enou gh data.\nFor a deep dive into the differences between these approaches, check out "Supervised\nvs. Unsupervised Learning: What\'s the Difference?"\nReinforcement ma chine learning\nReinforcement machine learning is a machine learning model that is similar to\nsupervised learning, but the algorithm isn’t trained using sample dat a. This model learns\nas it goes by using trial and error. A sequence of successfu l outcomes will be reinforced\nto develop the best recommendation or policy for a given problem.\nThe IBM Watson® system that won the Jeopardy! challenge in 2011 is a good example.\nThe system used reinforcement learning to learn when to attempt a n answer (or\nquestion, as it were), which square to select on the board, and how much to\nwager—especially on daily doubles.\nLearn more about reinforcement learni ng\nCommon machine learning algorithms\nA number of machine learning algorithms ar e commonly used. These include:\nNeural networks: Neural networks simulate the way the human brain works, with\na huge number of linked processing nodes. Neural netw orks are good at\nrecognizing patterns and play an important role in applications including natural\nlanguage translation, image recognition, speech recognition, an d image creation.\nLinear regression: This algorithm is used to predict numerical values, based on a\nlinear relationship between different values. For example, the technique could be\nused to predict house prices based on historical data for the

area.\nLogistic regression: This supervised learning algorithm makes predictions f
or\ncategorical response variables, such as “yes/no” answers to questions. It can
be\nused for applications such as classifying spam and quality control on a\nprodu
ction line.\nClustering: Using unsupervised learning, clustering algorithms can id
entify\npatterns in data so that it can be grouped. Computers can help data scient
ists by\nidentifying differences between data items that humans have overlooked.\n
Decision trees: Decision trees can be used for both predicting numerical values\n
(regression) and classifying data into categories. Decision trees use a branching
\nsequence of linked decisions that can be represented with a tree diagram. One of
\nthe advantages of decision trees is that they are easy to validate and audit,\nu
nlike the black box of the neural network.\nRandom forests: In a random forest, th
e machine learning algorithm predicts a\nvalue or category by combining the result
s from a number of decision trees.\nAdvantages and disadvantages of machine learni
ng algorithms\nDepending on your budget, need for speed and precision required, ea
ch algorithm\ntype—supervised, unsupervised, semi-supervised, or reinforcement—has
its own\nadvantages and disadvantages. For example, decision tree algorithms are u
sed for both\npredicting numerical values (regression problems) and classifying da
ta into categories.\nDecision trees use a branching sequence of linked decisions t
hat may be represented\nwith a tree diagram. A prime advantage of decision trees i
s that they are easier to\nvalidate and audit than a neural network. The bad news
is that they can be more\nunstable than other decision predictors.\nOverall, there
are many advantages to machine learning that businesses can leverage\nfor new effi
ciencies. These include machine learning identifying patterns and trends in\nmassi
ve volumes of data that humans might not spot at all. And this analysis requires\n
little human intervention: just feed in the dataset of interest and let the machin
e learning\nsystem assemble and refine its own algorithms—which will continually i
mprove with\nmore data input over time. Customers and users can enjoy a more perso
nalized\nexperience as the model learns more with every experience with that perso
n.\nOn the downside, machine learning requires large training datasets that are ac
curate\nand unbiased. GIGO is the operative factor: garbage in / garbage out. Gath
ering\nsufficient data and having a system robust enough to run it might also be a
drain on\nresources. Machine learning can also be prone to error, depending on the
input. With\ntoo small a sample, the system could produce a perfectly logical algo
rithm that is\ncompletely wrong or misleading. To avoid wasting budget or displeas
ing customers,\norganizations should act on the answers only when there is high co
nfidence in the\noutput.\nReal-world machine learning use cases\nHere are just a f
ew examples of machine learning you might encounter every day:\nSpeech recognitio
n: It is also known as automatic speech recognition (ASR), computer\nspeech recogn
ition, or speech-to-text, and it is a capability which uses natural language\nproc
essing (NLP) to translate human speech into a written format. Many mobile devices
\nincorporate speech recognition into their systems to conduct voice search—e.g. S
iri—or\nimprove accessibility for texting.\nCustomer service: Online chatbots are
replacing human agents along the customer\njourney, changing the way we think abou
t customer engagement across websites and\nsocial media platforms. Chatbots answer
frequently asked questions (FAQs) about\ntopics such as shipping, or provide perso
nalized advice, cross-selling products or\nsuggesting sizes for users. Examples in
clude virtual agents on e-commerce sites;\nmessaging bots, using Slack and Faceboo
k Messenger; and tasks usually done by\nvirtual assistants and voice assistants.\n
Computer vision: This AI technology enables computers to derive meaningful\ninform
ation from digital images, videos, and other visual inputs, and then take the\napp
ropriate action. Powered by convolutional neural networks, computer vision has\nap
plications in photo tagging on social media, radiology imaging in healthcare, and
\nself-driving cars in the automotive industry.\nRecommendation engines: Using pas
t consumption behavior data, AI algorithms can\nhelp to discover data trends that
can be used to develop more effective cross-selling\nstrategies. Recommendation en
gines are used by online retailers to make relevant\nproduct recommendations to cu

stomers during the checkout process.\nRobotic process automation (RPA): Also known as software robotics, RPA uses\nintelligent automation technologies to perform repetitive manual tasks.\nAutomated stock trading: Designed to optimize stock portfolios, AI-driven\nhigh-frequency trading platforms make thousands or even millions of trades per day\nwithout human intervention.\nFraud detection: Banks and other financial institutions can use machine learning to spot\nsuspicious transactions. Supervised learning can train a model using information about\nknown fraudulent transactions. Anomaly detection can identify transactions that look\natypical and deserve further investigation.\nChallenges of machine learning\nAs machine learning technology has developed, it has certainly made our lives easier.\nHowever, implementing machine learning in businesses has also raised a number of\nethical concerns about AI technologies. Some of these include:\nTechnological singularity\nWhile this topic garners a lot of public attention, many researchers are not concerned\nwith the idea of AI surpassing human intelligence in the near future. Technological \nsingularity is also referred to as strong AI or superintelligence. Philosopher Nick\nBostrum defines superintelligence as “any intellect that vastly outperforms the best\nhuman brains in practically every field, including scientific creativity, general wisdom,\nand social skills.” Despite the fact that superintelligence is not imminent in society, the\nidea of it raises some interesting questions as we consider the use of autonomous\nsystems, like self-driving cars. It’s unrealistic to think that a driverless car would never\nhave an accident, but who is responsible and liable under those circumstances? Should\nwe still develop autonomous vehicles, or do we limit this technology to\nsemi-autonomous vehicles which help people drive safely? The jury is still out on this,\nbut these are the types of ethical debates that are occurring as new, innovative AI\ntechnology develops.\nAI impact on jobs\nWhile a lot of public perception of artificial intelligence centers around job losses, this\nconcern should probably be reframed. With every disruptive, new technology, we see\nthat the market demand for specific job roles shifts. For example, when we look at the\nautomotive industry, many manufacturers, like GM, are shifting to focus on electric\nvehicle production to align with green initiatives. The energy industry isn’t going away,\nbut the source of energy is shifting from a fuel economy to an electric one.\nIn a similar way, artificial intelligence will shift the demand for jobs to other areas. There\nwill need to be individuals to help manage AI systems. There will still need to be people\nto address more complex problems within the industries that are most likely to be\naffected by job demand shifts, such as customer service. The biggest challenge with\nartificial intelligence and its effect on the job market will be helping people to transition\nto new roles that are in demand.\nPrivacy\nPrivacy tends to be discussed in the context of data privacy, data protection, and data\nsecurity. These concerns have allowed policymakers to make more strides in recent\nyears. For example, in 2016, GDPR legislation was created to protect the personal data\nof people in the European Union and European Economic Area, giving individuals more\ncontrol of their data. In the United States, individual states are developing policies, such\nas the California Consumer Privacy Act (CCPA), which was introduced in 2018 and\nrequires businesses to inform consumers about the collection of their data. Legislation\nsuch as this has forced companies to rethink how they store and use personally\nidentifiable information (PII). As a result, investments in security have become an\nincreasing priority for businesses as they seek to eliminate any vulnerabilities and\nopportunities for surveillance, hacking, and cyberattacks.\nBias and discrimination\nInstances of bias and discrimination across a number of machine learning systems have\nraised many ethical questions regarding the use of artificial intelligence. How can we \nsafeguard against bias and discrimination when the training data itself may be\ngenerated by biased human processes? While companies typically have good\nintentions for their automation efforts, Reuters (link resides outside ibm.com) highlights \nsome of the unforeseen consequences of incorporating AI into hiring practices. In their\neffort to automate and simplify a process, Amazon unintentionally discrim

```
inated against\njob candidates by gender for technical roles, and the company ulti
mately had to scrap\nthe project. Harvard Business Review (link resides outside ib
m.com) has raised other\npointed questions about the use of AI in hiring practice
s, such as what data you should\nbe able to use when evaluating a candidate for a
role.\nBias and discrimination aren’t limited to the human resources function eith
er; they can\nbe found in a number of applications from facial recognition softwar
e to social media\nalgorithms.\nAs businesses become more aware of the risks with
AI, they’ve also become more\nactive in this discussion around AI ethics and value
s. For example, IBM has sunset its\ngeneral purpose facial recognition and analysi
s products. IBM CEO Arvind Krishna\nwrote: “IBM firmly opposes and will not condon
e uses of any technology, including facial\nrecognition technology offered by othe
r vendors, for mass surveillance, racial profiling,\nviolations of basic human rig
hts and freedoms, or any purpose which is not consistent\nwith our values and Prin
ciples of Trust and Transparency.”\nAccountability\nSince there isn’t significant
legislation to regulate AI practices, there is no real\nenforcement mechanism to e
nsure that ethical AI is practiced. The current incentives for\ncompanies to be et
hical are the negative repercussions of an unethical AI system on the\nbottom lin
e. To fill the gap, ethical frameworks have emerged as part of a collaboration\nbe
tween ethicists and researchers to govern the construction and distribution of AI
\nmodels within society. However, at the moment, these only serve to guide. Some\n
research (link resides outside ibm.com) shows that the combination of distributed
\nresponsibility and a lack of foresight into potential consequences aren’t conduc
ive to\npreventing harm to society.\nRead more about IBM\'s position on AI Ethics
\nHow to choose the right AI platform for machine learning\nSelecting a platform c
an be a challenging process, as the wrong system can drive up\ncosts, or limit the
use of other valuable tools or technologies. When reviewing multiple\nvendors to s
elect an AI platform, there is often a tendency to think that more features =\na b
etter system. Maybe so, but reviewers should start by thinking through what the AI
\nplatform will be doing for their organization. What machine learning capabilitie
s need to\nbe delivered and what features are important to accomplish them? One mi
ssing feature\nmight doom the usefulness of an entire system. Here are some featur
es to consider.\nMLOps capabilities. Does the system have:\na unified interface fo
r ease of management?\nautomated machine learning tools for faster model creation
with low-code\nand no-code functionality?\ndecision optimization to streamline the
selection and deployment of\noptimization models?\nvisual modeling to combine visu
al data science with open-source libraries\nand notebook-based interfaces on a uni
fied data and AI studio?\nautomated development for beginners to get started quick
ly and more\nadvanced data scientists to experiment?\nsynthetic data generator as
an alternative or supplement to real-world data\nwhen real-world data is not readi
ly available?\nGenerative AI capabilities. Does the system have:\na content genera
tor that can generate text, images and other content\nbased on the data it was tra
ined on?\nautomated classification to read and classify written input, such as\nev
aluating and sorting customer complaints or reviewing customer\nfeedback sentimen
t?\na summary generator that can transform dense text into a high-quality\nsummar
y, capture key points from financial reports, and generate meeting\ntranscription
s?\na data extraction capability to sort through complex details and quickly pull
\nthe necessary information from large documents?'
```

In [46]:
```
print(doc[0].text)
```

What is machine learning?
Machine learning is a branch of artificial intelligence (AI) and computer science wh
ich
focuses on the use of data and algorithms to imitate the way that humans learn,
gradually improving its accuracy.
IBM has a rich history with machine learning. One of its own, Arthur Samuel, is cred
ited
for coining the term, “machine learning” with his research (link resides outside ib
m.com)
around the game of checkers. Robert Nealey, the self-proclaimed checkers master,
played the game on an IBM 7094 computer in 1962, and he lost to the computer.
Compared to what can be done today, this feat seems trivial, but it’s considered a m
ajor
milestone in the field of artificial intelligence.
Over the last couple of decades, the technological advances in storage and processin
g
power have enabled some innovative products based on machine learning, such as
Netflix’s recommendation engine and self-driving cars.
Machine learning is an important component of the growing field of data science.
Through the use of statistical methods, algorithms are trained to make classificatio
ns or
predictions, and to uncover key insights in data mining projects. These insights
subsequently drive decision making within applications and businesses, ideally
impacting key growth metrics. As big data continues to expand and grow, the market
demand for new data scientists will increase. They will be required to help identify
the
most relevant business questions and the data to answer them.
Machine learning algorithms are typically created using frameworks such as Python th
at
accelerate solution development by using platforms like TensorFlow or PyTorch.
Now available: watsonx.ai
The all-new enterprise studio that brings together traditional machine learning alon
g
with new generative AI capabilities powered by foundation models.
Try watsonx.ai
Begin your journey to AI
Learn how to scale AI
Explore the AI Academy
Machine Learning vs. Deep Learning vs. Neural Networks
Since deep learning and machine learning tend to be used interchangeably, it’s worth
noting the nuances between the two. Machine learning, deep learning, and neural
networks are all sub-fields of artificial intelligence. However, neural networks is
actually
a sub-field of machine learning, and deep learning is a sub-field of neural network
s.
The way in which deep learning and machine learning differ is in how each algorithm
learns. "Deep" machine learning can use labeled datasets, also known as supervised
learning, to inform its algorithm, but it doesn’t necessarily require a labeled data
set. The
deep learning process can ingest unstructured data in its raw form (e.g., text or im
ages),
and it can automatically determine the set of features which distinguish different
categories of data from one another. This eliminates some of the human intervention
required and enables the use of large amounts of data. You can think of deep learnin
g
as "scalable machine learning" as Lex Fridman notes in this MIT lecture (link reside

s
outside ibm.com).
Classical, or "non-deep," machine learning is more dependent on human intervention t
o
learn. Human experts determine the set of features to understand the differences
between data inputs, usually requiring more structured data to learn.
Neural networks, or artificial neural networks (ANNs), are comprised of node layers,
containing an input layer, one or more hidden layers, and an output layer. Each nod
e, or
artificial neuron, connects to another and has an associated weight and threshold. I
f the
output of any individual node is above the specified threshold value, that node is
activated, sending data to the next layer of the network. Otherwise, no data is pass
ed
along to the next layer of the network by that node. The “deep” in deep learning is
just
referring to the number of layers in a neural network. A neural network that consist
s of
more than three layers—which would be inclusive of the input and the output—can be
considered a deep learning algorithm or a deep neural network. A neural network that
only has three layers is just a basic neural network.
Deep learning and neural networks are credited with accelerating progress in areas
such as computer vision, natural language processing, and speech recognition.
See the blog post “AI vs. Machine Learning vs. Deep Learning vs. Neural Networks:
What’s the Difference?” for a closer look at how the different concepts relate.
Related content
Explore the watsonx.ai interactive demo
Download “Machine learning for Dummies”
- This link downloads a pdf
Explore Gen AI for developers
How does machine learning work?
UC Berkeley (link resides outside ibm.com) breaks out the learning system of a
machine learning algorithm into three main parts.
A Decision Process: In general, machine learning algorithms are used to make a
prediction or classification. Based on some input data, which can be labeled or
unlabeled, your algorithm will produce an estimate about a pattern in the data.
An Error Function: An error function evaluates the prediction of the model. If
there are known examples, an error function can make a comparison to assess
the accuracy of the model.
A Model Optimization Process: If the model can fit better to the data points in the
training set, then weights are adjusted to reduce the discrepancy between the
known example and the model estimate. The algorithm will repeat this iterative
“evaluate and optimize” process, updating weights autonomously until a
threshold of accuracy has been met.
Machine learning methods
Machine learning models fall into three primary categories.
Supervised machine learning
Supervised learning, also known as supervised machine learning, is defined by its us
e
of labeled datasets to train algorithms to classify data or predict outcomes accurat
ely.
As input data is fed into the model, the model adjusts its weights until it has been
fitted
appropriately. This occurs as part of the cross validation process to ensure that th
e
model avoids overfitting or underfitting. Supervised learning helps organizations so

lve a
variety of real-world problems at scale, such as classifying spam in a separate fold
er
from your inbox. Some methods used in supervised learning include neural networks,
naïve bayes, linear regression, logistic regression, random forest, and support vect
or
machine (SVM).
Unsupervised machine learning
Unsupervised learning, also known as unsupervised machine learning, uses machine
learning algorithms to analyze and cluster unlabeled datasets (subsets called cluste
rs).
These algorithms discover hidden patterns or data groupings without the need for
human intervention. This method’s ability to discover similarities and differences i
n
information make it ideal for exploratory data analysis, cross-selling strategies,
customer segmentation, and image and pattern recognition. It’s also used to reduce t
he
number of features in a model through the process of dimensionality reduction. Princ
ipal
component analysis (PCA) and singular value decomposition (SVD) are two common
approaches for this. Other algorithms used in unsupervised learning include neural
networks, k-means clustering, and probabilistic clustering methods.
Semi-supervised learning
Semi-supervised learning offers a happy medium between supervised and
unsupervised learning. During training, it uses a smaller labeled data set to guide
classification and feature extraction from a larger, unlabeled data set. Semi-superv
ised
learning can solve the problem of not having enough labeled data for a supervised
learning algorithm. It also helps if it’s too costly to label enough data.
For a deep dive into the differences between these approaches, check out "Supervised
vs. Unsupervised Learning: What's the Difference?"
Reinforcement machine learning
Reinforcement machine learning is a machine learning model that is similar to
supervised learning, but the algorithm isn’t trained using sample data. This model l
earns
as it goes by using trial and error. A sequence of successful outcomes will be reinf
orced
to develop the best recommendation or policy for a given problem.
The IBM Watson® system that won the Jeopardy! challenge in 2011 is a good example.
The system used reinforcement learning to learn when to attempt an answer (or
question, as it were), which square to select on the board, and how much to
wager—especially on daily doubles.
Learn more about reinforcement learning
Common machine learning algorithms
A number of machine learning algorithms are commonly used. These include:
Neural networks: Neural networks simulate the way the human brain works, with
a huge number of linked processing nodes. Neural networks are good at
recognizing patterns and play an important role in applications including natural
language translation, image recognition, speech recognition, and image creation.
Linear regression: This algorithm is used to predict numerical values, based on a
linear relationship between different values. For example, the technique could be
used to predict house prices based on historical data for the area.
Logistic regression: This supervised learning algorithm makes predictions for
categorical response variables, such as “yes/no” answers to questions. It can be
used for applications such as classifying spam and quality control on a
production line.

Clustering: Using unsupervised learning, clustering algorithms can identify patterns in data so that it can be grouped. Computers can help data scientists by identifying differences between data items that humans have overlooked.
Decision trees: Decision trees can be used for both predicting numerical values (regression) and classifying data into categories. Decision trees use a branching sequence of linked decisions that can be represented with a tree diagram. One of the advantages of decision trees is that they are easy to validate and audit, unlike the black box of the neural network.
Random forests: In a random forest, the machine learning algorithm predicts a value or category by combining the results from a number of decision trees.
Advantages and disadvantages of machine learning algorithms
Depending on your budget, need for speed and precision required, each algorithm type—supervised, unsupervised, semi-supervised, or reinforcement—has its own advantages and disadvantages. For example, decision tree algorithms are used for both predicting numerical values (regression problems) and classifying data into categories.
Decision trees use a branching sequence of linked decisions that may be represented with a tree diagram. A prime advantage of decision trees is that they are easier to validate and audit than a neural network. The bad news is that they can be more unstable than other decision predictors.
Overall, there are many advantages to machine learning that businesses can leverage for new efficiencies. These include machine learning identifying patterns and trends in massive volumes of data that humans might not spot at all. And this analysis requires little human intervention: just feed in the dataset of interest and let the machine learning system assemble and refine its own algorithms—which will continually improve with more data input over time. Customers and users can enjoy a more personalized experience as the model learns more with every experience with that person.
On the downside, machine learning requires large training datasets that are accurate and unbiased. GIGO is the operative factor: garbage in / garbage out. Gathering sufficient data and having a system robust enough to run it might also be a drain on resources. Machine learning can also be prone to error, depending on the input. With too small a sample, the system could produce a perfectly logical algorithm that is completely wrong or misleading. To avoid wasting budget or displeasing customers, organizations should act on the answers only when there is high confidence in the output.
Real-world machine learning use cases
Here are just a few examples of machine learning you might encounter every day:
Speech recognition: It is also known as automatic speech recognition (ASR), computer speech recognition, or speech-to-text, and it is a capability which uses natural language processing (NLP) to translate human speech into a written format. Many mobile devices incorporate speech recognition into their systems to conduct voice search—e.g. Siri—or improve accessibility for texting.
Customer service: Online chatbots are replacing human agents along the customer journey, changing the way we think about customer engagement across websites and social media platforms. Chatbots answer frequently asked questions (FAQs) about topics such as shipping, or provide personalized advice, cross-selling products or suggesting sizes for users. Examples include virtual agents on e-commerce sites; messaging bots, using Slack and Facebook Messenger; and tasks usually done by virtual assistants and voice assistants.

Computer vision: This AI technology enables computers to derive meaningful information from digital images, videos, and other visual inputs, and then take the appropriate action. Powered by convolutional neural networks, computer vision has applications in photo tagging on social media, radiology imaging in healthcare, and self-driving cars in the automotive industry.
Recommendation engines: Using past consumption behavior data, AI algorithms can help to discover data trends that can be used to develop more effective cross-sellin g
strategies. Recommendation engines are used by online retailers to make relevant product recommendations to customers during the checkout process.
Robotic process automation (RPA): Also known as software robotics, RPA uses intelligent automation technologies to perform repetitive manual tasks.
Automated stock trading: Designed to optimize stock portfolios, AI-driven high-frequency trading platforms make thousands or even millions of trades per day without human intervention.
Fraud detection: Banks and other financial institutions can use machine learning to spot
suspicious transactions. Supervised learning can train a model using information abo ut
known fraudulent transactions. Anomaly detection can identify transactions that look atypical and deserve further investigation.
Challenges of machine learning
As machine learning technology has developed, it has certainly made our lives easie r.
However, implementing machine learning in businesses has also raised a number of ethical concerns about AI technologies. Some of these include:
Technological singularity
While this topic garners a lot of public attention, many researchers are not concern ed
with the idea of AI surpassing human intelligence in the near future. Technological singularity is also referred to as strong AI or superintelligence. Philosopher Nick Bostrum defines superintelligence as “any intellect that vastly outperforms the best human brains in practically every field, including scientific creativity, general wi sdom,
and social skills.” Despite the fact that superintelligence is not imminent in socie ty, the
idea of it raises some interesting questions as we consider the use of autonomous systems, like self-driving cars. It’s unrealistic to think that a driverless car wou ld never
have an accident, but who is responsible and liable under those circumstances? Shoul d
we still develop autonomous vehicles, or do we limit this technology to semi-autonomous vehicles which help people drive safely? The jury is still out on th is,
but these are the types of ethical debates that are occurring as new, innovative AI technology develops.
AI impact on jobs
While a lot of public perception of artificial intelligence centers around job losse s, this
concern should probably be reframed. With every disruptive, new technology, we see that the market demand for specific job roles shifts. For example, when we look at t he
automotive industry, many manufacturers, like GM, are shifting to focus on electric vehicle production to align with green initiatives. The energy industry isn’t going away,
but the source of energy is shifting from a fuel economy to an electric one.

In a similar way, artificial intelligence will shift the demand for jobs to other areas. There will need to be individuals to help manage AI systems. There will still need to be people to address more complex problems within the industries that are most likely to be affected by job demand shifts, such as customer service. The biggest challenge with artificial intelligence and its effect on the job market will be helping people to transition to new roles that are in demand.

Privacy

Privacy tends to be discussed in the context of data privacy, data protection, and data security. These concerns have allowed policymakers to make more strides in recent years. For example, in 2016, GDPR legislation was created to protect the personal data of people in the European Union and European Economic Area, giving individuals more control of their data. In the United States, individual states are developing policies, such as the California Consumer Privacy Act (CCPA), which was introduced in 2018 and requires businesses to inform consumers about the collection of their data. Legislation such as this has forced companies to rethink how they store and use personally identifiable information (PII). As a result, investments in security have become an increasing priority for businesses as they seek to eliminate any vulnerabilities and opportunities for surveillance, hacking, and cyberattacks.

Bias and discrimination

Instances of bias and discrimination across a number of machine learning systems have raised many ethical questions regarding the use of artificial intelligence. How can we safeguard against bias and discrimination when the training data itself may be generated by biased human processes? While companies typically have good intentions for their automation efforts, Reuters (link resides outside ibm.com) highlights some of the unforeseen consequences of incorporating AI into hiring practices. In their effort to automate and simplify a process, Amazon unintentionally discriminated against job candidates by gender for technical roles, and the company ultimately had to scrap the project. Harvard Business Review (link resides outside ibm.com) has raised other pointed questions about the use of AI in hiring practices, such as what data you should be able to use when evaluating a candidate for a role.

Bias and discrimination aren’t limited to the human resources function either; they can be found in a number of applications from facial recognition software to social media algorithms.

As businesses become more aware of the risks with AI, they’ve also become more active in this discussion around AI ethics and values. For example, IBM has sunset its general purpose facial recognition and analysis products. IBM CEO Arvind Krishna wrote: “IBM firmly opposes and will not condone uses of any technology, including facial recognition technology offered by other vendors, for mass surveillance, racial profi

```
ling,
violations of basic human rights and freedoms, or any purpose which is not consisten
t
with our values and Principles of Trust and Transparency.”
Accountability
Since there isn’t significant legislation to regulate AI practices, there is no real
enforcement mechanism to ensure that ethical AI is practiced. The current incentives
for
companies to be ethical are the negative repercussions of an unethical AI system on
the
bottom line. To fill the gap, ethical frameworks have emerged as part of a collabora
tion
between ethicists and researchers to govern the construction and distribution of AI
models within society. However, at the moment, these only serve to guide. Some
research (link resides outside ibm.com) shows that the combination of distributed
responsibility and a lack of foresight into potential consequences aren’t conducive
to
preventing harm to society.
Read more about IBM's position on AI Ethics
How to choose the right AI platform for machine learning
Selecting a platform can be a challenging process, as the wrong system can drive up
costs, or limit the use of other valuable tools or technologies. When reviewing mult
iple
vendors to select an AI platform, there is often a tendency to think that more featu
res =
a better system. Maybe so, but reviewers should start by thinking through what the A
I
platform will be doing for their organization. What machine learning capabilities ne
ed to
be delivered and what features are important to accomplish them? One missing feature
might doom the usefulness of an entire system. Here are some features to consider.
MLOps capabilities. Does the system have:
a unified interface for ease of management?
automated machine learning tools for faster model creation with low-code
and no-code functionality?
decision optimization to streamline the selection and deployment of
optimization models?
visual modeling to combine visual data science with open-source libraries
and notebook-based interfaces on a unified data and AI studio?
automated development for beginners to get started quickly and more
advanced data scientists to experiment?
synthetic data generator as an alternative or supplement to real-world data
when real-world data is not readily available?
Generative AI capabilities. Does the system have:
a content generator that can generate text, images and other content
based on the data it was trained on?
automated classification to read and classify written input, such as
evaluating and sorting customer complaints or reviewing customer
feedback sentiment?
a summary generator that can transform dense text into a high-quality
summary, capture key points from financial reports, and generate meeting
transcriptions?
a data extraction capability to sort through complex details and quickly pull
the necessary information from large documents?
```

In [47]:
```
model = Gemini(model="gemini-1.5-flash", api_key=google_api_key)
```

```
C:\Users\amits_b2zbl1r\AppData\Local\Temp\ipykernel_21024\2469896082.py:1: Deprecati
onWarning: Call to deprecated class Gemini. (Should use `llama-index-llms-google-gen
ai` instead, using Google's latest unified SDK. See: https://docs.llamaindex.ai/en/s
table/examples/llm/google_genai/)
  model = Gemini(model="gemini-1.5-flash", api_key=google_api_key)
```

In [48]:
```
GeminiEmbedding = GeminiEmbedding(model="embedding-gecko-001")
```

```
C:\Users\amits_b2zbl1r\AppData\Local\Temp\ipykernel_21024\2179559832.py:1: Deprecati
onWarning: Call to deprecated class GeminiEmbedding. (Should use `llama-index-embedd
ings-google-genai` instead, using Google's latest unified SDK. See: https://docs.lla
maindex.ai/en/stable/examples/embeddings/google_genai/)
  GeminiEmbedding = GeminiEmbedding(model="embedding-gecko-001")
```

In [49]:
```
ServiceContext = ServiceContext.from_defaults(llm=model, embed_model=GeminiEmbeddin
```

```
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
Cell In[49], line 1
----> 1 ServiceContext = ServiceContext.from_defaults(llm=model, embed_model=GeminiE
mbedding, chunk_size=500, chunk_overlap=20)

File d:\Generative AI(by sunny savita)\qafolder\venv\lib\site-packages\llama_index\c
ore\service_context.py:33, in ServiceContext.from_defaults(cls, **kwargs)
     21 @classmethod
     22 def from_defaults(
     23     cls,
     24     **kwargs: Any,
     25 ) -> "ServiceContext":
     26     """
     27     Create a ServiceContext from defaults.
     28
   (...)
     31
     32     """
---> 33     raise ValueError(
     34         "ServiceContext is deprecated. Use llama_index.settings.Settings ins
tead, "
     35         "or pass in modules to local functions/methods/interfaces.\n"
     36         "See the docs for updated usage/migration: \n"
     37         "https://docs.llamaindex.ai/en/stable/module_guides/supporting_modul
es/service_context_migration/"
     38     )

ValueError: ServiceContext is deprecated. Use llama_index.settings.Settings instead,
or pass in modules to local functions/methods/interfaces.
See the docs for updated usage/migration:
https://docs.llamaindex.ai/en/stable/module_guides/supporting_modules/service_contex
t_migration/
```

In [50]:
```
index = VectorStoreIndex.from_documents(
    doc,
    llm=model,
    embed_model=GeminiEmbedding,
    chunk_size=500,
```

```
    chunk_overlap=20
)
```

In [51]:
```
index
```

Out[51]:
```
<llama_index.core.indices.vector_store.base.VectorStoreIndex at 0x1811a48b6a0>
```

In [52]:
```
index.storage_context.persist(persist_dir="index_storage")
```

In [53]:
```
query_engine = index.as_query_engine()
```

```
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
File d:\Generative AI(by sunny savita)\qafolder\venv\lib\site-packages\llama_index\c
ore\llms\utils.py:42, in resolve_llm(llm, callback_manager)
     41     llm = OpenAI()
---> 42     validate_openai_api_key(llm.api_key)  # type: ignore
     43 except ImportError:

File d:\Generative AI(by sunny savita)\qafolder\venv\lib\site-packages\llama_index\l
lms\openai\utils.py:790, in validate_openai_api_key(api_key)
    789 if not openai_api_key:
--> 790     raise ValueError(MISSING_API_KEY_ERROR_MESSAGE)

ValueError: No API key found for OpenAI.
Please set either the OPENAI_API_KEY environment variable or openai.api_key prior to
initialization.
API keys can be found or created at https://platform.openai.com/account/api-keys


During handling of the above exception, another exception occurred:

ValueError                                Traceback (most recent call last)
Cell In[53], line 1
----> 1 query_engine = index.as_query_engine()

File d:\Generative AI(by sunny savita)\qafolder\venv\lib\site-packages\llama_index\c
ore\indices\base.py:511, in BaseIndex.as_query_engine(self, llm, **kwargs)
    503 from llama_index.core.query_engine.retriever_query_engine import (
    504     RetrieverQueryEngine,
    505 )
    507 retriever = self.as_retriever(**kwargs)
    508 llm = (
    509     resolve_llm(llm, callback_manager=self._callback_manager)
    510     if llm
--> 511     else Settings.llm
    512 )
    514 return RetrieverQueryEngine.from_args(
    515     retriever,
    516     llm=llm,
    517     **kwargs,
    518 )

File d:\Generative AI(by sunny savita)\qafolder\venv\lib\site-packages\llama_index\c
ore\settings.py:36, in _Settings.llm(self)
     34 """Get the LLM."""
     35 if self._llm is None:
---> 36     self._llm = resolve_llm("default")
     38 if self._callback_manager is not None:
     39     self._llm.callback_manager = self._callback_manager

File d:\Generative AI(by sunny savita)\qafolder\venv\lib\site-packages\llama_index\c
ore\llms\utils.py:49, in resolve_llm(llm, callback_manager)
     44         raise ImportError(
     45             "`llama-index-llms-openai` package not found, "
     46             "please run `pip install llama-index-llms-openai`"
     47         )
```

```
     48     except ValueError as e:
---> 49         raise ValueError(
     50             "\n******\n"
     51             "Could not load OpenAI model. "
     52             "If you intended to use OpenAI, please check your OPENAI_API_KE
Y.\n"
     53             "Original error:\n"
     54             f"{e!s}"
     55             "\nTo disable the LLM entirely, set llm=None."
     56             "\n******"
     57         )
     59 if isinstance(llm, str):
     60     splits = llm.split(":", 1)

ValueError:
******
Could not load OpenAI model. If you intended to use OpenAI, please check your OPENAI
_API_KEY.
Original error:
No API key found for OpenAI.
Please set either the OPENAI_API_KEY environment variable or openai.api_key prior to
initialization.
API keys can be found or created at https://platform.openai.com/account/api-keys

To disable the LLM entirely, set llm=None.
******
```

In [54]:
```
query_engine = index.as_query_engine(llm=model)
```

In [56]:
```
response = query_engine.query("What is machine learning?  Explain in simple terms.
```

In [57]:
```
response.response
```

Out[57]:
```
"Machine learning uses data and algorithms to mimic human learning, improving accu
racy over time.  It's a type of artificial intelligence.\n\nOne example is a model
that predicts numerical values based on a linear relationship between different va
lues; this is called linear regression.\n"
```

In [58]:
```
print(response.response)
```

```
Machine learning uses data and algorithms to mimic human learning, improving accurac
y over time.  It's a type of artificial intelligence.

One example is a model that predicts numerical values based on a linear relationship
between different values; this is called linear regression.
```

In [60]:
```
response = query_engine.query("What is mahindra singh dhoni?")
```

In [62]:
```
print(response.response)
```

```
This question cannot be answered from the given source.
```

In [ ]:
```

```