simplilearn | PURDUE UNIVERSITY

Natural Language Processing and Speech Recognition

©Simplilearn. All rights reserved.

DATA AND
ARTIFICIAL INTELLIGENCE

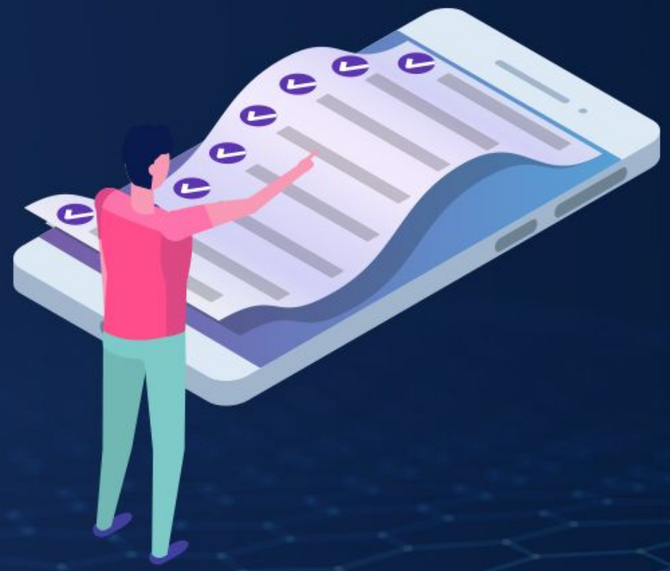
Natural Language Generation

©Simplilearn. All rights reserved.

Learning Objectives

By the end of this lesson, you will be able to:

- Explain various Natural Language Generation models
- Define Language Modeling
- Explain the challenges in NLP and how sentence correction works
- Create AIML patterns
- Predict the next suitable word in a sentence



©Simplilearn. All rights reserved.

DATA AND
ARTIFICIAL INTELLIGENCE

Introduction to NLG

©Simplilearn. All rights reserved.

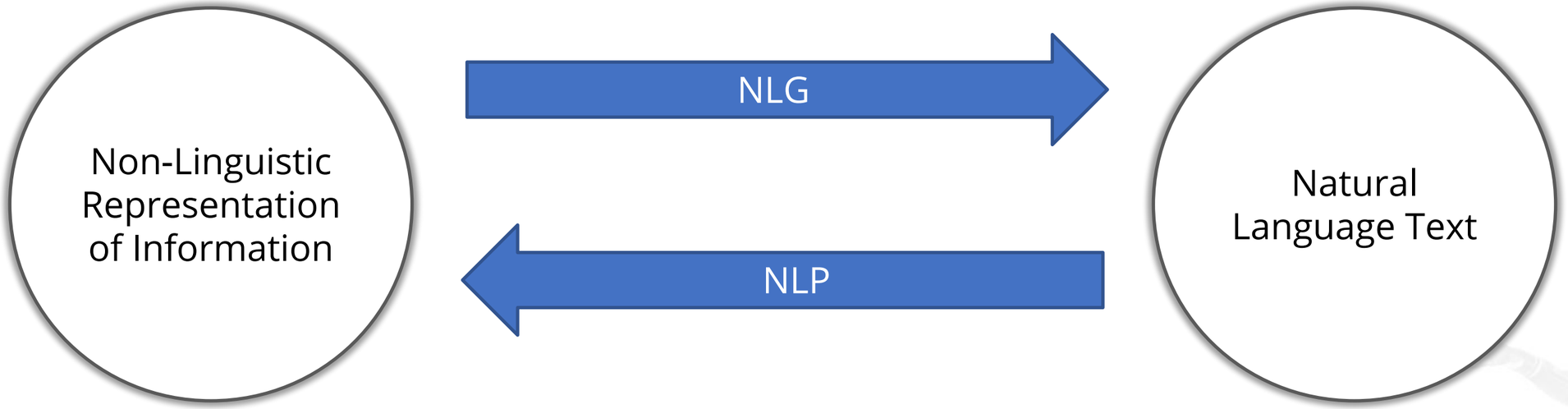
What Is NLG?

It is a part of artificial intelligence and computational linguistics that mainly focuses on computer systems which can produce understandable text in human languages.

It converts a computer-based representation into natural language representation which is the opposite process of Natural Language Understanding.

©Simplilearn. All rights reserved.

NLU and NLG Conversion



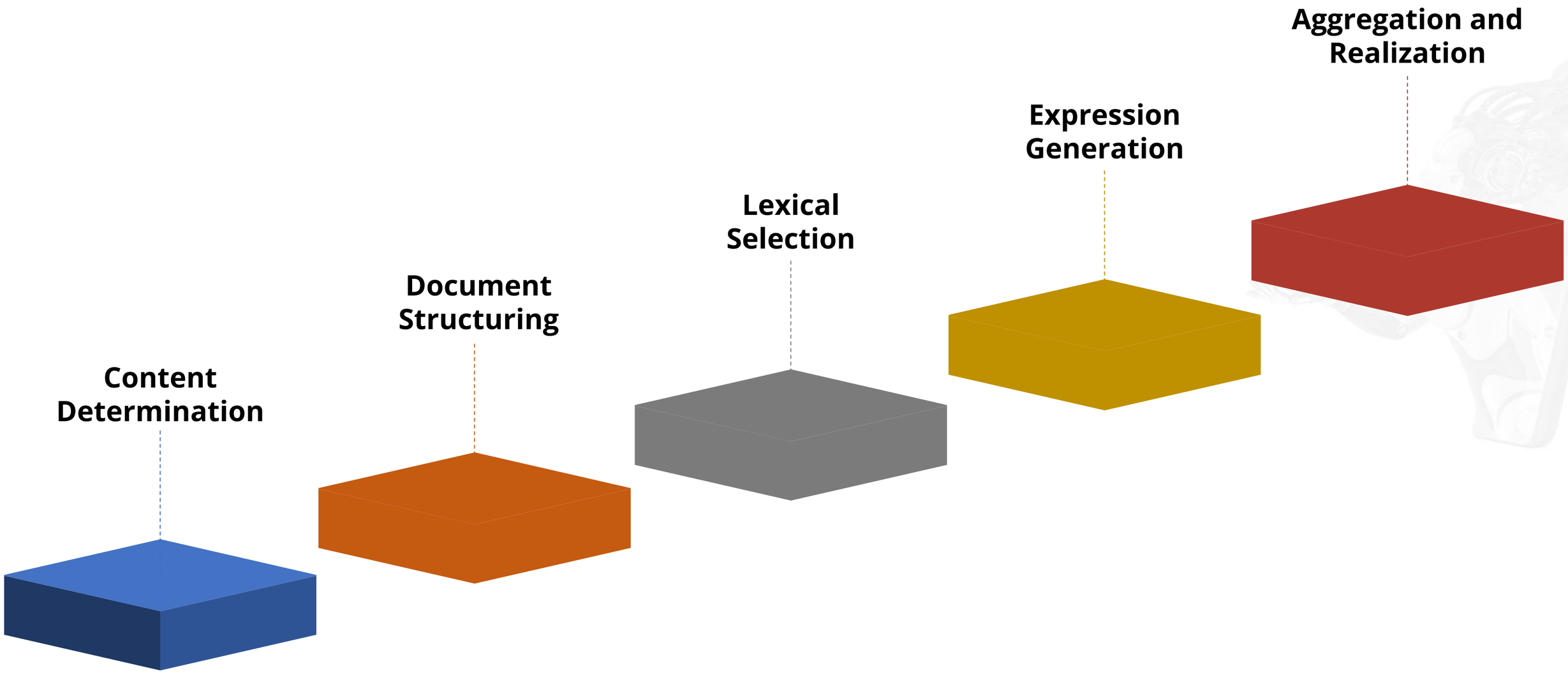
Data Stream, Log Files, Semantic Information

Reports, Explanations, Summaries, Recommendations, and Narratives

©Simplilearn. All rights reserved.

simplilearn

Stages in NLG



©Simplilearn. All rights reserved.

simplilearn

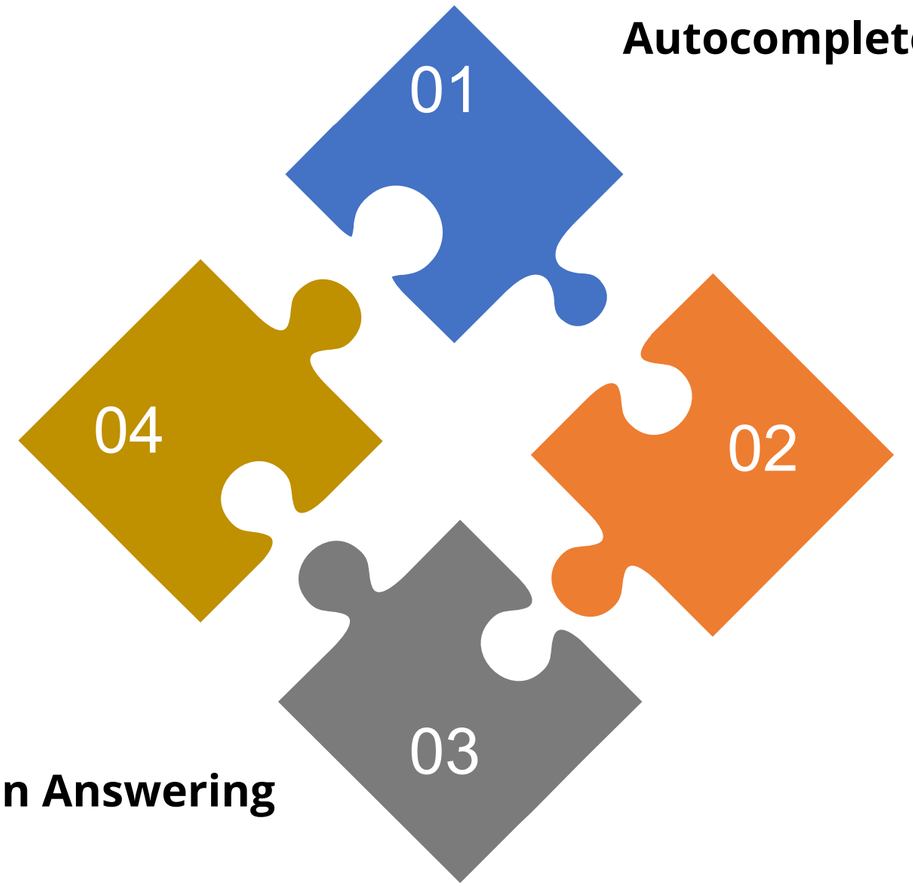
Response Generation Mechanism

Generative-Based Model

Retrieval-Based Model

©Simplilearn. All rights reserved.

Applications of NLG



Autocomplete

Translation

Question Answering

Dialogue System

©Simplilearn. All rights reserved.

simplilearn

DATA AND
ARTIFICIAL INTELLIGENCE

Retrieval-Based Model

©Simplilearn. All rights reserved.

Retrieval-Based Model: Introduction

1. The model creates responses from a bunch of predefined patterns
2. Input and context are important parameters to pick the responses
3. It uses heuristics to fetch the best result from the available responses
4. The score is generated for picking the relevant responses

©Simplilearn. All rights reserved.

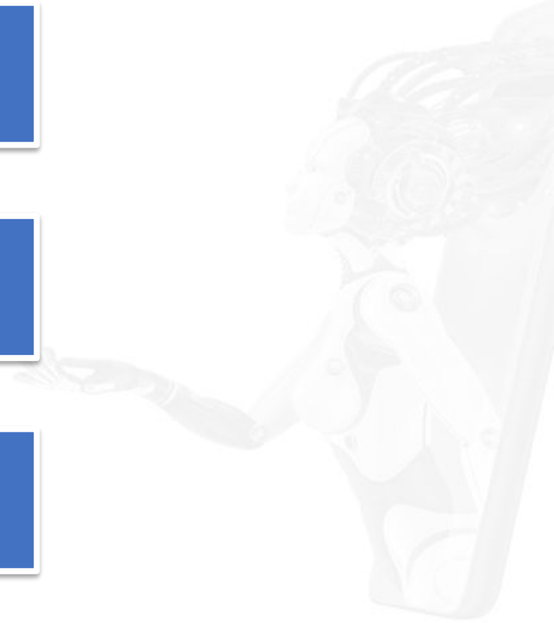
Retrieval-Based Model: Pros

- Less chances of error as the system consists of grammatically correct responses
- Suitable for customer satisfaction and business problems
- Requires less effort and data

©Simplilearn. All rights reserved.

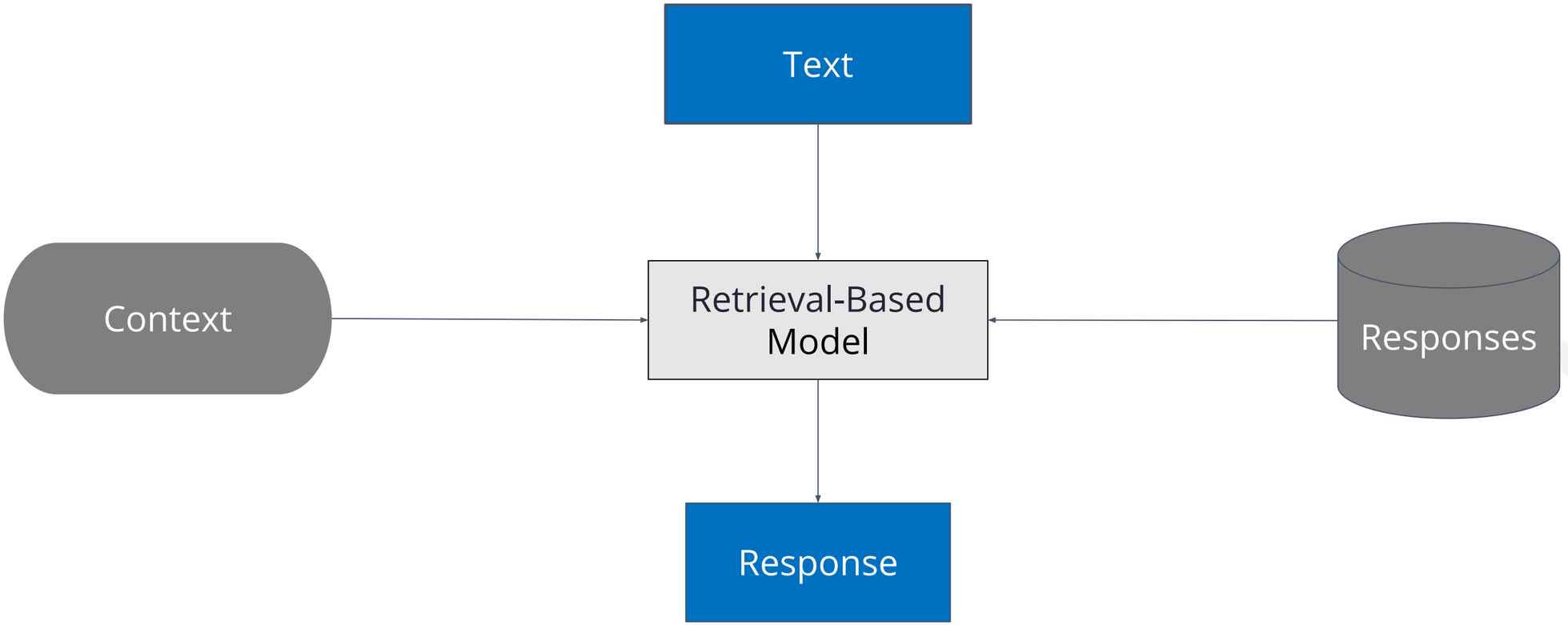
Retrieval-Based Model: Cons

- No new text generation possible due to its fixed nature
- Lots of heuristics are written due to which the system is not intelligent
- Can handle only predefined scenarios



©Simplilearn. All rights reserved.

Retrieval-Based Model: Architecture



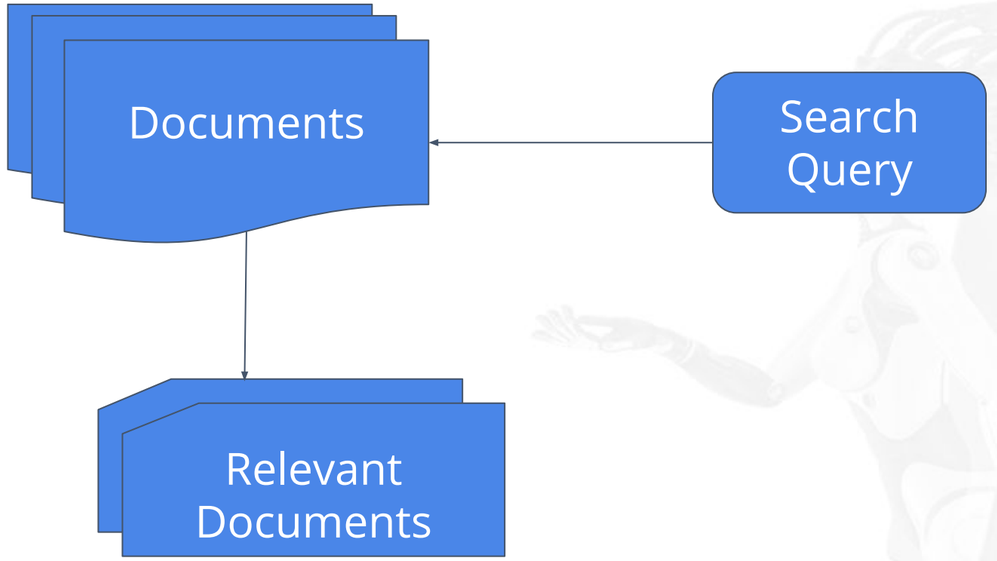
©Simplilearn. All rights reserved.

simplilearn

Retrieval-Based Model: Example

Search Engine or Document Retrieval System:

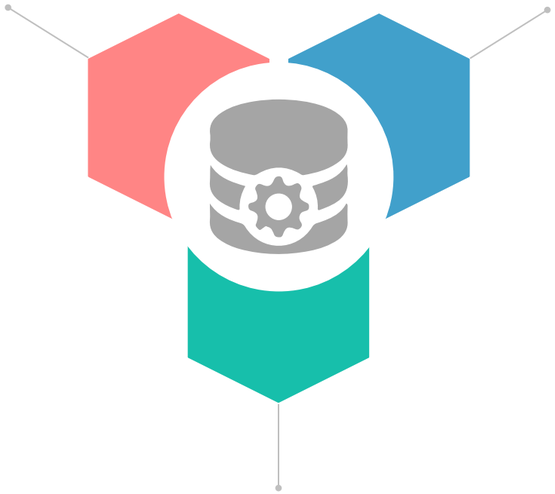
- It is used in information retrieval systems
- Knowledge base is set of documents and input is a search term or query
- Task is to retrieve documents that are most relevant to the search query



©Simplilearn. All rights reserved.

Retrieval-Based Model: Tools

Vinyals and Le Approach



ELIZA

AIML

©Simplilearn. All rights reserved.

DATA AND
ARTIFICIAL INTELLIGENCE

Artificial Intelligence Markup Language (AIML)

©Simplilearn. All rights reserved.

AIML: Introduction

1. XML based markup language
2. Pattern-based heuristics
3. Easy to understand and highly maintainable
4. Programming language, an Extensible Markup Language (XML)
5. Useful in creating artificially intelligent applications

©Simplilearn. All rights reserved.

AIML: Introduction

AIML is used in:

©Simplilearn. All rights reserved.

AIML: History

Developed by the Alicebot free software community and Dr. Richard S. Wallace

1995

- It is used to create Alicebot.
- A.L.I.C.E. (Artificial Linguistic Internet Computer Entity) is a chat box application.
- AIML interpreters are available in Java, Ruby, Python, C++, C#, and Pascal.

©Simplilearn. All rights reserved.

AIML: Elements

©Simplilearn. All rights reserved.

AIML: Tags

Sr.No	AIML Tags
1	<star> Used to match wildcard * characters in the <pattern> Tag
2	<srai> Multipurpose tag, used to call or match the other categories
3	<random> Used <random> to get random responses
4	<li> Used to represent multiple responses
5	<set> Used to set value in an AIML variable
6	<get> Used to get value, stored in an AIML variable
7	<that> Used to respond, based on the context in AIML
8	<topic> Used in AIML to store a context
9	<think> Used to store a variable in AIML without acknowledging the user
10	<condition> Helps ALICE to respond to the matching input

©Simplilearn. All rights reserved.

AIML: Example

```
<category>
  <pattern>WHAT IS YOUR DOB?</pattern>
  <template>My DOB is 15/08/1990</template>
</category>

<category>
  <pattern>WHEN IS YOUR BIRTHDAY?</pattern>
  <template>
    <srai>What is your DOB?</srai>
  </template>
 </category>
```


©Simplilearn. All rights reserved.

Creating AIML Patterns

Problem Statement: A retrieval-based model is used to create response for questions asked by the user. These are predefined responses in the addition to the slots. Create AIML patterns for QnA on mental wellness.

Access: Click on the **Practice Labs** tab on the left side panel of the LMS. Copy or note the username and password that is generated. Click on the **Launch Lab** button. On the page that appears, enter the username and password in the respective fields, and click **Login**.

ASSISTED PRACTICE

©Simplilearn. All rights reserved.

DATA AND
ARTIFICIAL INTELLIGENCE

Generative-Based Model

©Simplilearn. All rights reserved.

Generative-Based Model: Introduction

- Generative model is a statistical model of the joint probability distribution of X and Y:
  - X is observable
  - Y is target
- Describes how a dataset is created, in terms of a probabilistic model
- Learns any kind of data distribution, using unsupervised learning models

©Simplilearn. All rights reserved.

Generative-Based Model: Approaches

Variational Autoencoders (VAE)

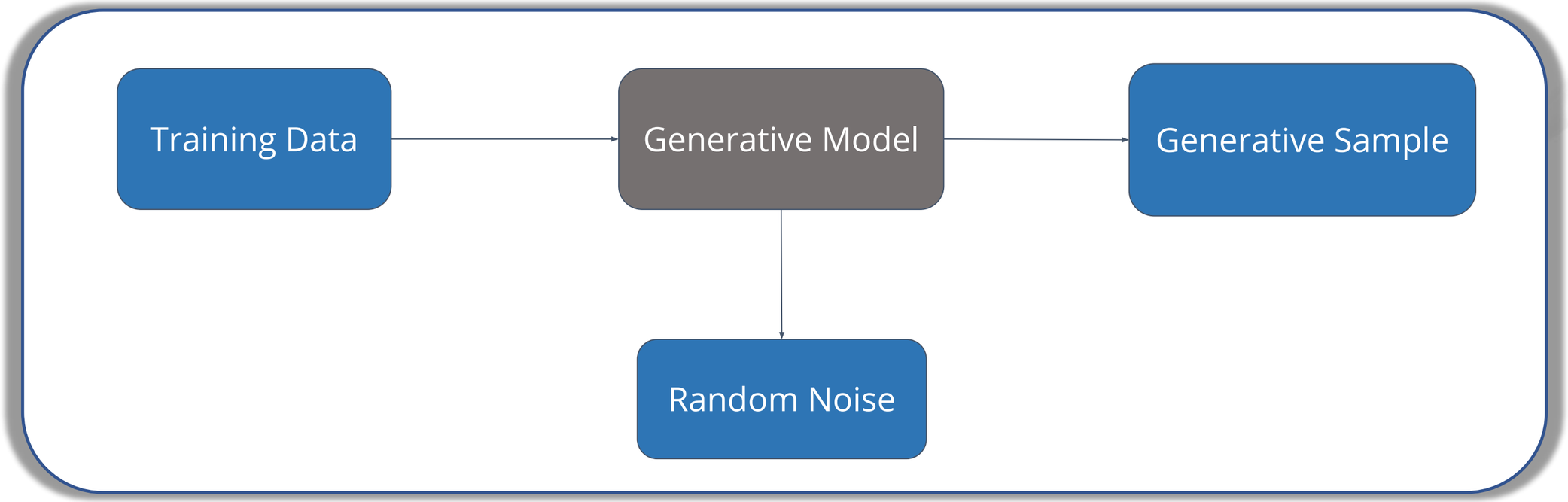
Generative Adversarial Networks (GAN)

©Simplilearn. All rights reserved.

Generative-Based Model: Example

Create a model that can generate a new image of dog:

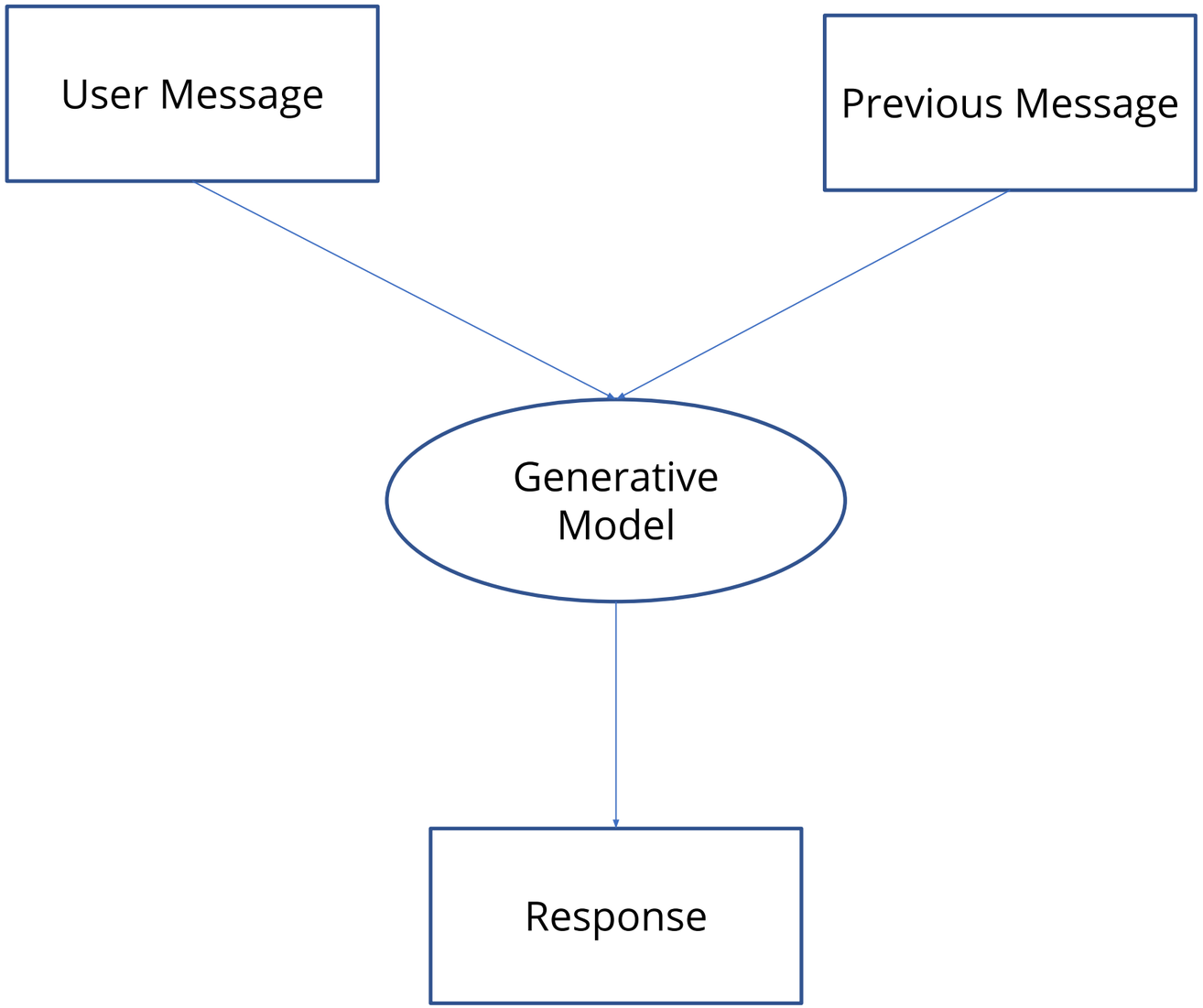
- Input: Dataset of dogs
- Model: To learn generic rules to create new outputs
- Output: New images of dogs



©Simplilearn. All rights reserved.

simplilearn

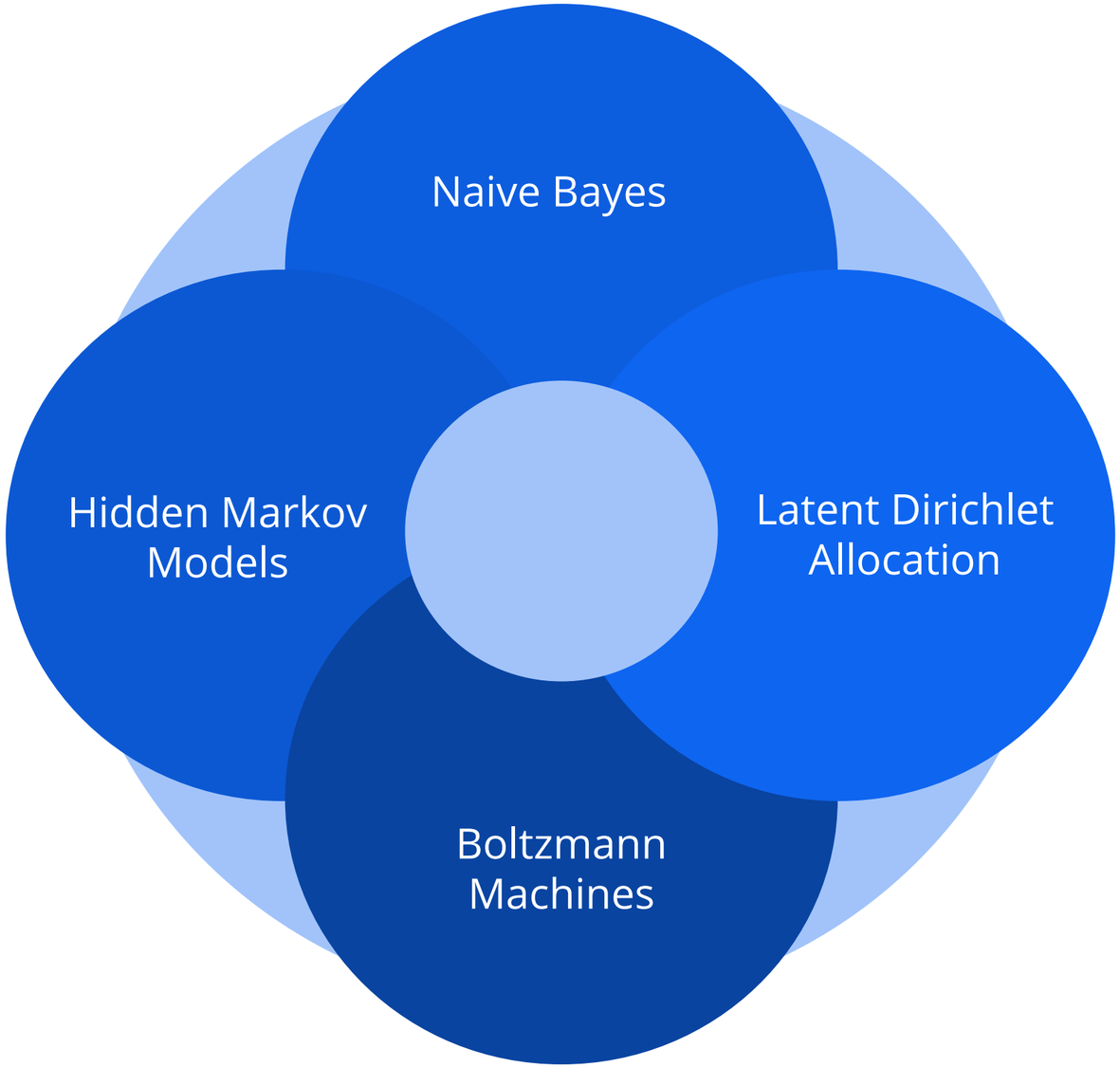
Generative-Based Model: Architecture



©Simplilearn. All rights reserved.

simplilearn

Generative-Based Model: Techniques



©Simplilearn. All rights reserved.

simplilearn

Generative-Based Model: Working

A generative model can be used to perform prediction: **argmaxy** P(Y=y|X=x)= **argmaxy** P(Y=y,X=x)/P(X=x) and since P(X=x) is constant on the RHS, this equals to **argmaxy** P(Y=y,X=x)

Generative models are capable of more than just prediction, i.e. maximizing P(Y|X=x). By estimating P(Y, X) and able to sample X, Y pairs

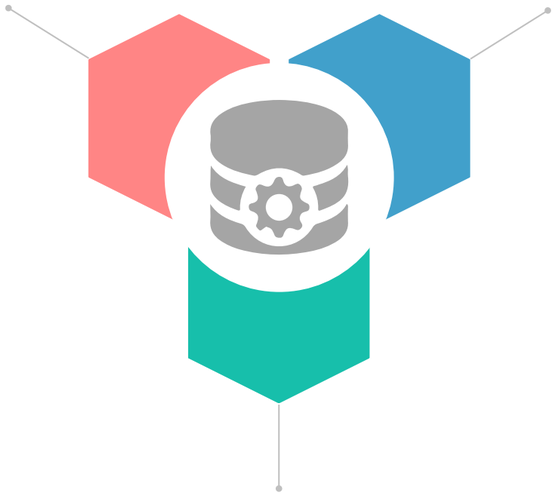
©Simplilearn. All rights reserved.

simplilearn

Generative-Based Model: Working

Generative model can be used to:

Impute missing data



Compress the dataset

Generate unseen data

©Simplilearn. All rights reserved.

DATA AND
ARTIFICIAL INTELLIGENCE

Language Modeling

©Simplilearn. All rights reserved.

Introduction to Language Modeling

One of the fundamental tasks of NLP that has many applications

Used to compute a probability of a token

Way of statistical analysis of natural language

©Simplilearn. All rights reserved.

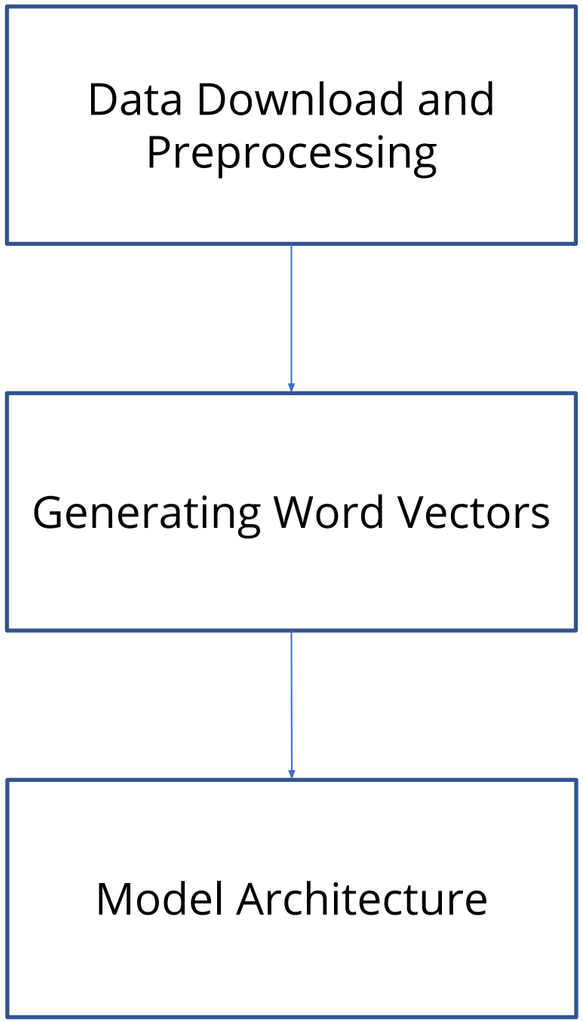
Introduction to Language Modeling

Language modeling is used for:

©Simplilearn. All rights reserved.

simplilearn

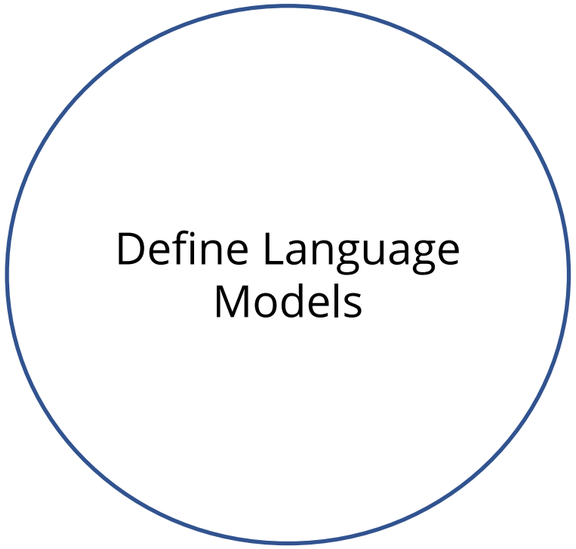
Next Word Prediction

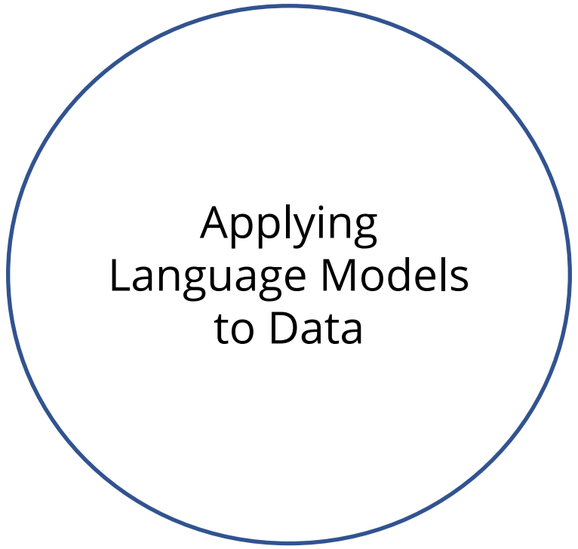


©Simplilearn. All rights reserved.

simplilearn

Process of Language Modeling





©Simplilearn. All rights reserved.

Process of Language Modeling

Define Language Models

Calculate the probability of a sentence of sequence of words:

$$P(W) = P(w1,w2,w3,....,wn)$$

Conditional probability or Chain rule

©Simplilearn. All rights reserved.

Process of Language Modeling

Define Language Models

Markov Assumption:

The conditional probability distribution of future states depend upon the present state

P(wi | w1, w2,..., wi-1) ⏩ P(wi | wi-k, ..., wi-1)

K is number of words

©Simplilearn. All rights reserved.

Process of Language Modeling

Define Language Models

N Grams:

- Unigram model
- Bi-Gram model

©Simplilearn. All rights reserved.

Process of Language Modeling

Applying Language Models to Data

Data Preprocessing:-

- Removing any punctuation and lowercase all words
- Forming sentences with probabilities

©Simplilearn. All rights reserved.

Process of Language Modeling

Training and Testing

- Training:
  - Model fitting
  - Model saving
- Testing is performed in the second step

©Simplilearn. All rights reserved.

Predict Next Word in a Sentence

Problem Statement: While writing something there are some systems that give you the prediction of next word based on the previous context. These are made by using huge trained data. Write a script to identify the next word in an email written to the manager for leave application.

Access: Click on the **Practice Labs** tab on the left side panel of the LMS. Copy or note the username and password that is generated. Click on the **Launch Lab** button. On the page that appears, enter the username and password in the respective fields, and click **Login**.

ASSISTED PRACTICE

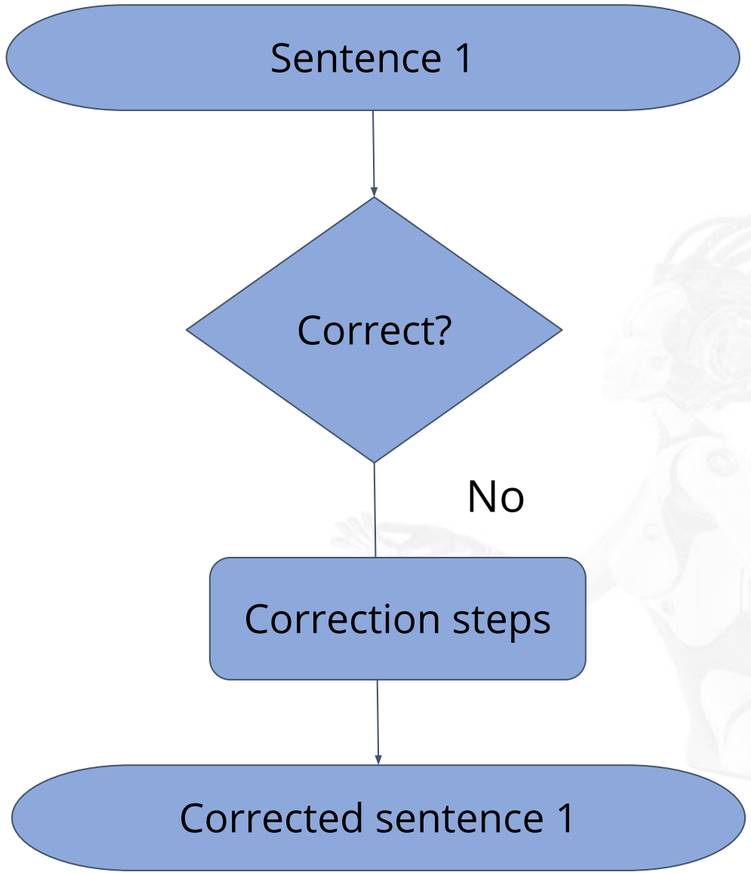
©Simplilearn. All rights reserved.

Sentence Correction

©Simplilearn. All rights reserved.

Introduction to Sentence Correction

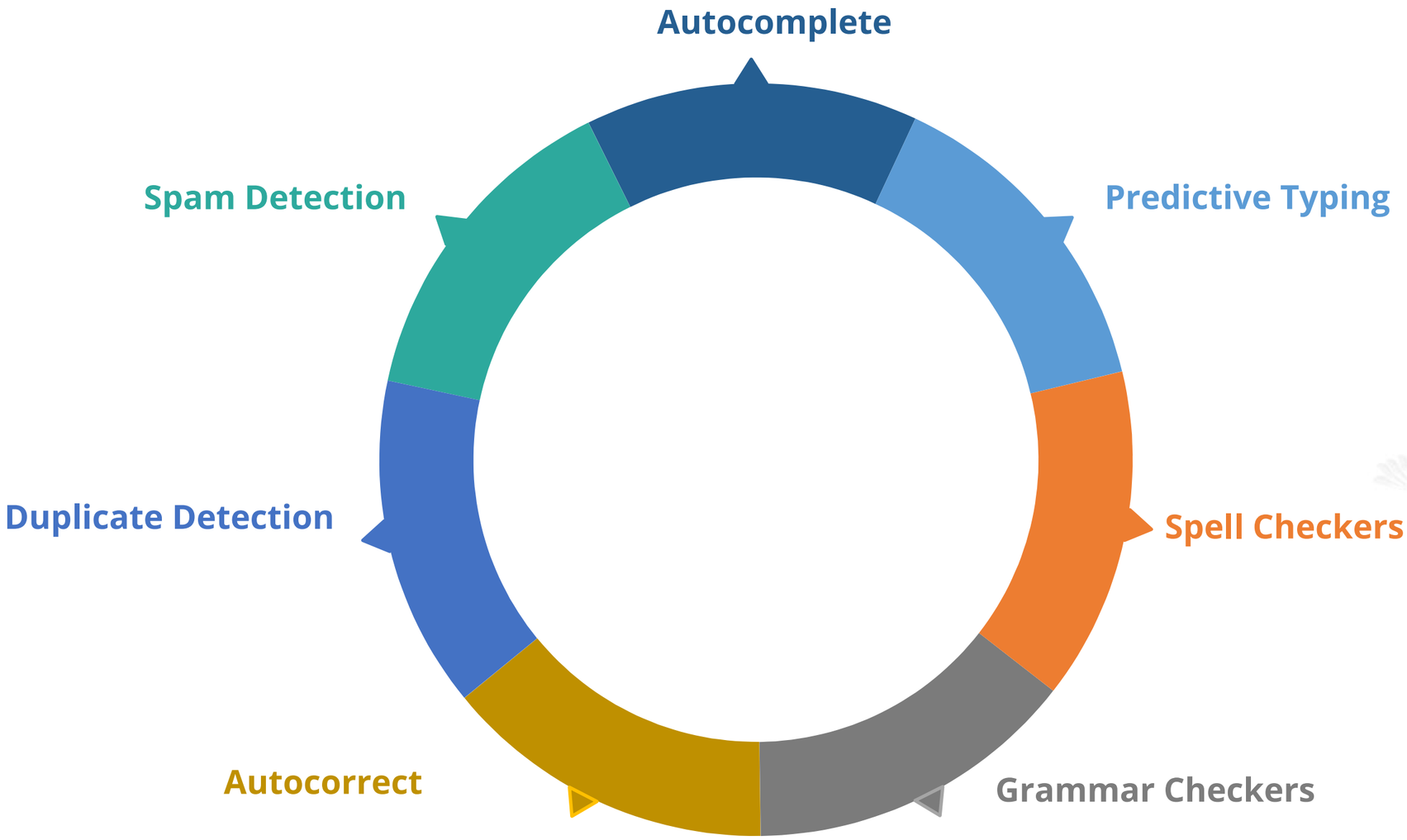
- Sentence correction is a task of correcting different kinds of errors in incoming text
- These errors are classified into:
  - Spelling
  - Punctuation
  - Grammatical
- Various types of error correction systems are available
- It is an important process of NLU and NLG



©Simplilearn. All rights reserved.

simplilearn

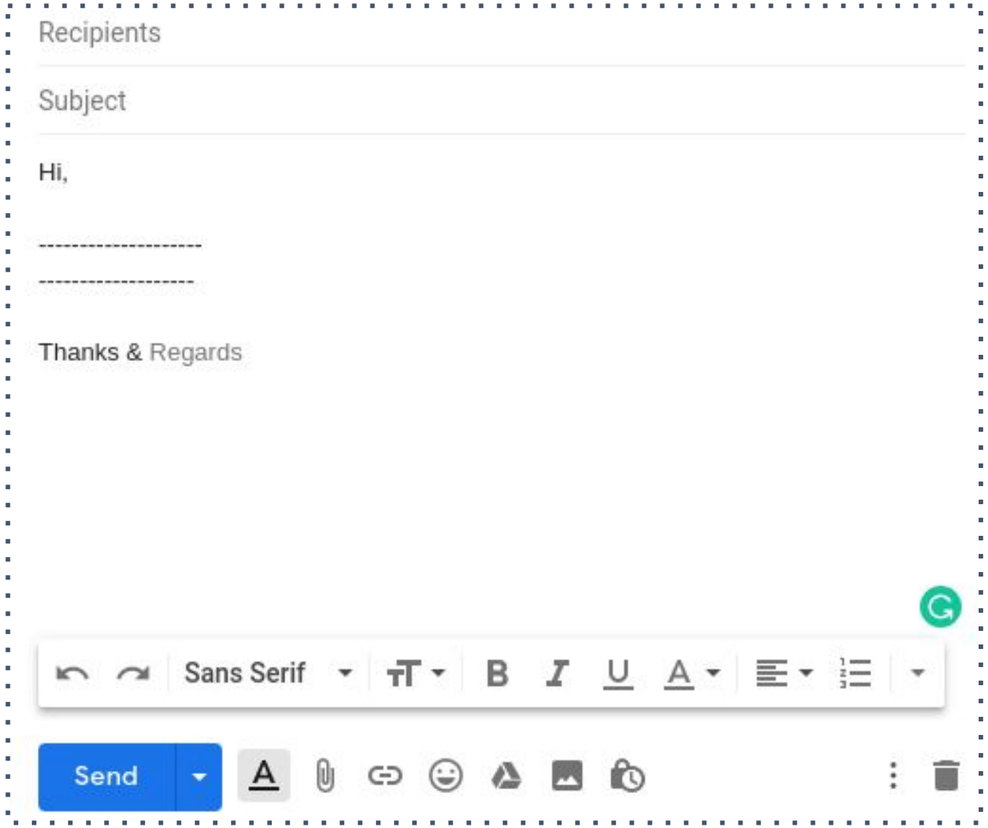
Error Correction Systems

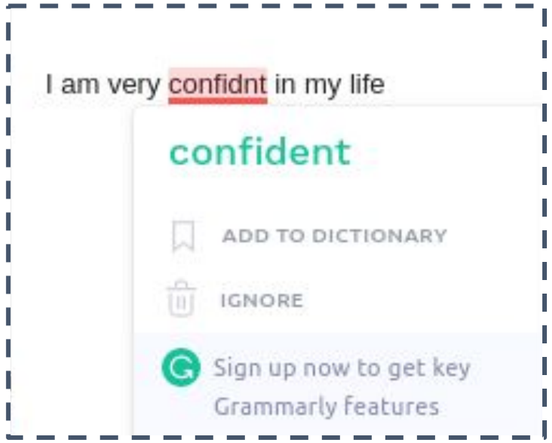


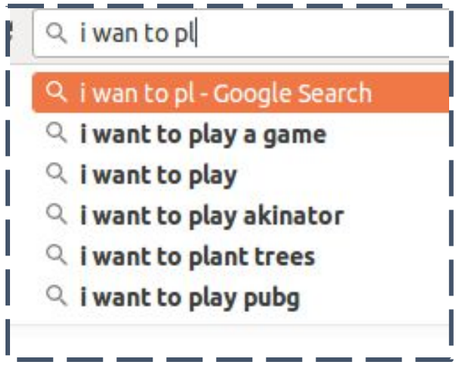
©Simplilearn. All rights reserved.

simplilearn

Example of Sentence Correction







©Simplilearn. All rights reserved.

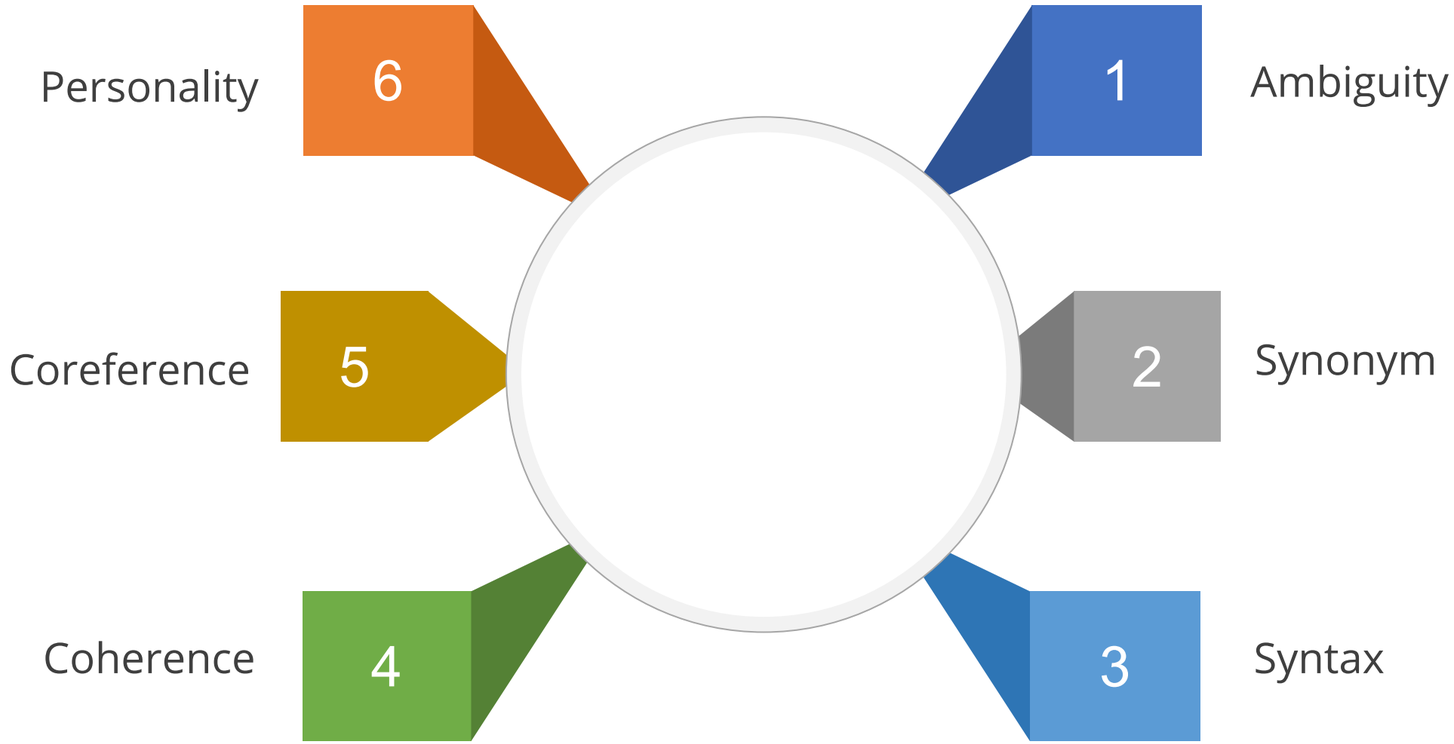
Basic Principles of Sentence Correction

There are two basic principles behind the spelling correction algorithms:

1. Choose the nearest one. It expects the proximity between a pair of queries.
2. Select the one that is more common when correctly spelled queries. Example: grunt and grant both seem equally plausible.

©Simplilearn. All rights reserved.

Challenges in Sentence Correction

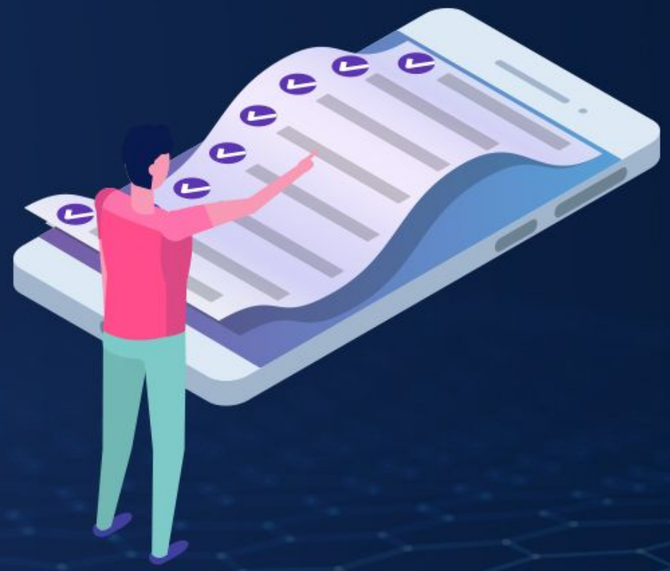


©Simplilearn. All rights reserved.

Key Takeaways

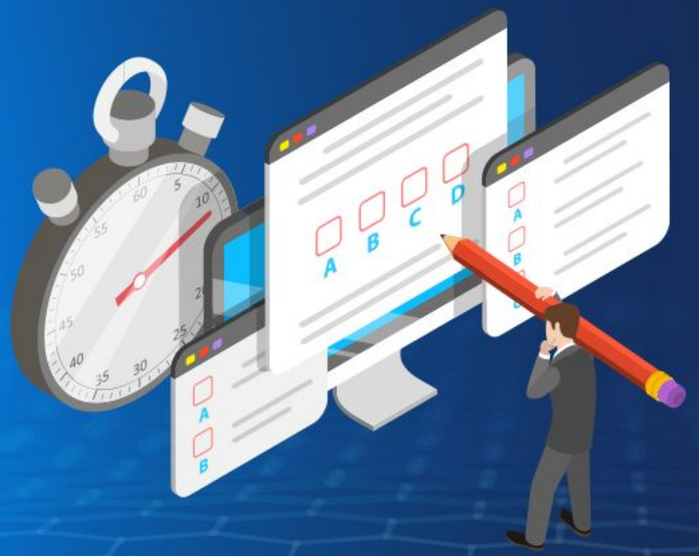
You are now able to:

- Explain various Natural Language Generation models
- Define Language Modeling
- Explain the challenges in NLP and how sentence correction works
- Create AIML patterns
- Predict the next suitable word in a sentence



©Simplilearn. All rights reserved.

DATA AND
ARTIFICIAL INTELLIGENCE



Knowledge Check

©Simplilearn. All rights reserved.

Knowledge Check

1

In which of the following models, we produce predefined responses for a query?

a. Retrieval-Based Model

b. Generative-Based Model

c. Statistical Model

d. Topic Modeling

©Simplilearn. All rights reserved.

Knowledge Check

1

In which of the following models, we produce predefined responses for a query?

a. Retrieval-Based Model

b. Generative-Based Model

c. Statistical Model

d. Topic Modeling

The correct answer is **a.**

Predefined patterns are used to generate the responses for a query in Retrieval-based model.

©Simplilearn. All rights reserved.

Knowledge Check

2

Which of the following are the uses of Natural Language Generation?

a. Creating answers

b. Prediction of next or previous word

c. Autocomplete

d. All of the above

©Simplilearn. All rights reserved.

Knowledge Check

2

Which of the following are the uses of Natural Language Generation?

a. Creating answers

b. Prediction of next or previous word

c. Autocomplete

d. All of the above

The correct answer is **d.**

NLG has all the capabilities for answer generation and correction.

©Simplilearn. All rights reserved.

Knowledge Check

3

Which of the following techniques are used in NLG?

a. Trend analysis

b. Market-based analysis

c. TF-IDF

d. RNN

©Simplilearn. All rights reserved.

Knowledge Check

3

Which of the following techniques are used in NLG?

a. Trend analysis

b. Market-based analysis

c. TF-IDF

d. RNN

The correct answer is **d.**

RNN is used for seq2seq, Markove is for pattern analysis, N-gram for next word prediction.

©Simplilearn. All rights reserved.

Knowledge Check

4

What is AIML?

a. Artificial Intelligence Markup Language

b. Artificial Intelligent Modeling Language

c. Artificial Intelligence Machine Language

d. Artificial Intelligence Morphology Language

©Simplilearn. All rights reserved.

Knowledge Check

4

What is AIML?

a. Artificial Intelligence Markup Language

b. Artificial Intelligent Modeling Language

c. Artificial Intelligence Machine Language

d. Artificial Intelligence Morphology Language

The correct answer is **a.**

AIM stands for Artificial Intelligence Markup Language.

©Simplilearn. All rights reserved.

Knowledge Check

5

Which one of the following creates a problem in NLG?

a. Data ambiguity

b. Autocomplete

c. Creating answers

d. Diversified answers

©Simplilearn. All rights reserved.

Knowledge Check

5

Which one of the following creates a problem in NLG?

a. Data ambiguity

b. Autocomplete

c. Creating answers

d. Diversified answers

The correct answer is **a.**

Data ambiguity creates problem in NLG.

©Simplilearn. All rights reserved.