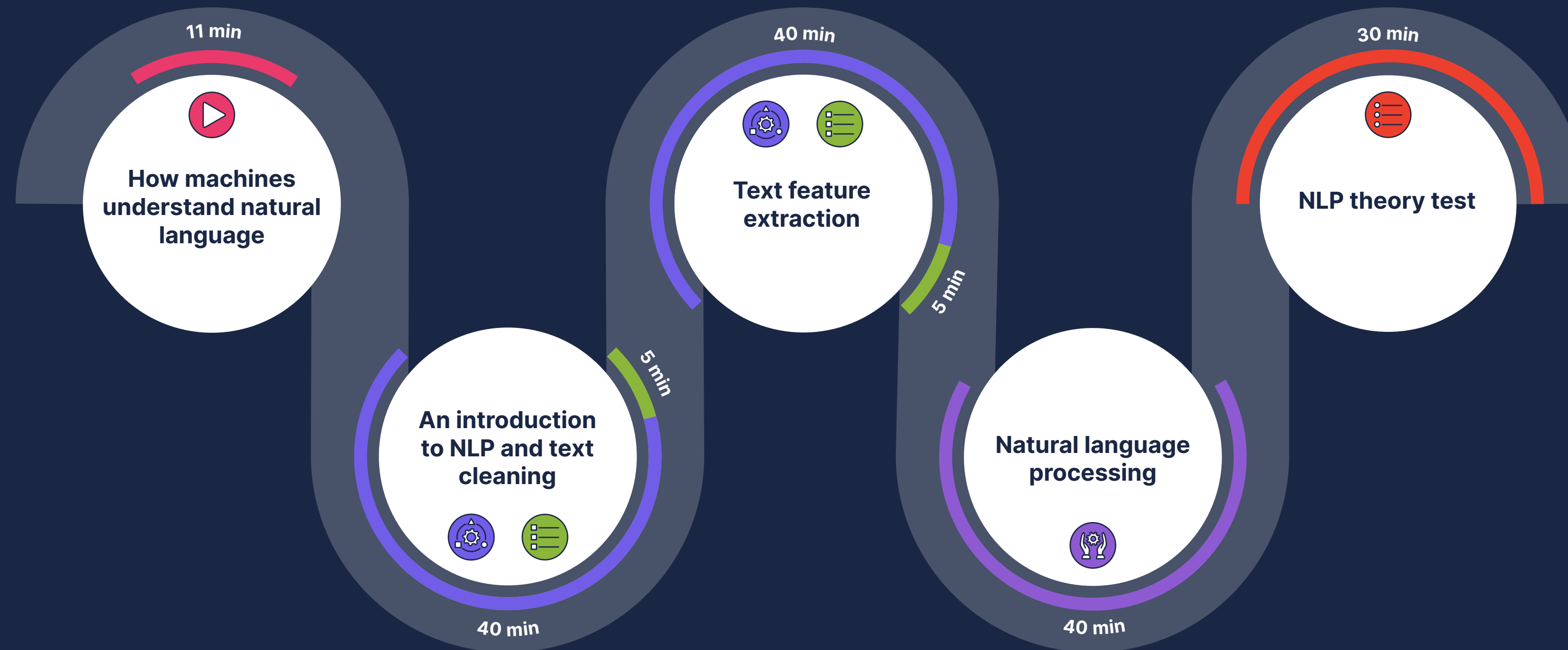


# Natural language processing – Lesson overview

Understanding **text preprocessing** and **feature extraction techniques** is crucial for effectively working with **textual data** in various natural language processing tasks. These techniques enable us to **transform raw text** into a format suitable for **analysis, modelling, and machine learning algorithms**.

In this lesson, we will explore text preprocessing methods such as **tokenisation, removing stopwords, stemming, lemmatisation, and removing punctuation**. We will also delve into feature extraction techniques such as **bag-of-words** and **n-grams** to convert text data into numerical representations.

 Video	 Knowledge questions
 Examples	 Multiple choice questions
 Exercise	



## Learning objectives

- Understand the importance of text preprocessing in natural language processing tasks.
- Create a bag-of-words representation to quantify the occurrence of words in text data.
- Apply stemming and lemmatisation to extract the root forms of words.
- Learn how to tokenise text data into words or tokens.
- Explore the concept of n-grams to capture combinations of words in text data.

