

[Print to PDF](#)

Latex Text Splitter

LatexTextSplitter splits text along Latex headings, headlines, enumerations and more. It's implemented as a simple subclass of RecursiveCharacterSplitter with Latex-specific separators. See the source code to see the Latex syntax expected by default.

1. How the text is split: by list of latex specific tags
2. How the chunk size is measured: by length function passed in (defaults to number of characters)

```
from langchain.text_splitter import LatexTextSplitter
```

```
latex_text = """
\documentclass{article}

\begin{document}

\maketitle

\section{Introduction}
Large language models (LLMs) are a type of machine learning model that can be
trained on vast amounts of text data to generate human-like language. In recent
years, LLMs have made significant advances in a variety of natural language
processing tasks, including language translation, text generation, and sentiment
analysis.

\subsection{History of LLMs}
The earliest LLMs were developed in the 1980s and 1990s, but they were limited by
the amount of data that could be processed and the computational power available
at the time. In the past decade, however, advances in hardware and software have
made it possible to train LLMs on massive datasets, leading to significant
improvements in performance.

\subsection{Applications of LLMs}
LLMs have many applications in industry, including chatbots, content creation, and
virtual assistants. They can also be used in academia for research in linguistics,
psychology, and computational linguistics.

\end{document}
"""

latex_splitter = LatexTextSplitter(chunk_size=400, chunk_overlap=0)
```

[Skip to main content](#)

```
docs = latex_splitter.create_documents([latex_text])
```

```
docs
```

```
[Document(page_content='\\documentclass{article}\\n\\n\\x08egin{document}\\n\\n\\maketitle', lookup_str='', metadata={}, lookup_index=0),  
 Document(page_content='Introduction}\\nLarge language models (LLMs) are a type of machine learning model that can be trained on vast amounts of text data to generate human-like language. In recent years, LLMs have made significant advances in a variety of natural language processing tasks, including language translation, text generation, and sentiment analysis.', lookup_str='', metadata={}, lookup_index=0),  
 Document(page_content='History of LLMs}\\nThe earliest LLMs were developed in the 1980s and 1990s, but they were limited by the amount of data that could be processed and the computational power available at the time. In the past decade, however, advances in hardware and software have made it possible to train LLMs on massive datasets, leading to significant improvements in performance.', lookup_str='', metadata={}, lookup_index=0),  
 Document(page_content='Applications of LLMs}\\nLLMs have many applications in industry, including chatbots, content creation, and virtual assistants. They can also be used in academia for research in linguistics, psychology, and computational linguistics.\\n\\n\\end{document}', lookup_str='', metadata={}, lookup_index=0)]
```