# [Title] Document Loader Input Example

[Subtitle] This document serves as a comprehensive example of an input for PDF, and DOCX Loader.

**[Paragraph]** This document is intended to provide **a comprehensive set of test cases for evaluating the functionality and robustness of our SDK Loaders**, which are responsible for extracting content from various document elements. The goal is to ensure that the loaders are capable of accurately processing a diverse range of document components. However, this DOES NOT imply that the current loader can accurately extract all the information in this document. To see the capability of each Loader, please check the Exploration Documentation:

- PDF: GenAI Dynamic Pipeline: DPO PDF Loader Parser
- DOCX: GenAI Dynamic Pipeline: DPO DOCX Loader Parser

This document contains a diverse range of document components, including[1] :

1. Header
2. Title
3. Subtitle
4. Heading
    a. Heading 1
    b. Heading 2
    c. Heading 3, etc
5. Paragraph
6. Table
7. Multiple Column layout
8. Footer
9. Footnotes
10. Endnote
11. Image
12. Hyperlink
13. List
14. Watermark

Document examples facilitate better testing and validation of the loaders, helping developers understand how different document elements are handled by the SDK. By providing diverse input examples, they allow developers to identify edge cases, troubleshoot issues, and refine their implementations. Ultimately, the examples help ensure that loaders perform consistently across various document formats and structures, leading to improved accuracy and reliability in document processing[2].

## [Heading 1] Content Types

---

[1] [Footnote] This is Footnote Document Example
[2] [Footnote] F. N. E. This is Footnote Document Example

The Content Types section highlights the potential components within a document. By accounting for this variety, developers can better equip their loaders to handle the complexities of real-world documents effectively.

## [Heading 2] Text

Text is the foundational element of a document and serves to convey the main content. Text can vary in several ways, including:

1. Structure: It can be organized into headers, footers, footnotes, titles, multiple heading levels, and paragraphs.
2. Text Properties: It may include attributes like font color, font family, and font size.

## [Heading 2] Table

Tables are structured layouts that organize data into rows and columns. They can range from simple, standard tables to more complex designs with merged cells or missing borders. In this document, we will provide several examples of different table types. The aim is to illustrate how various table formats processed by our SDK[i].

### [Heading 3] Table Variation 1: Standard Table

A standard table organizes data into rows and columns with clear borders, making it easy to view and distinguish information.

*Table 1. Standard Table Caption*

| Keterangan | Online | BI-FAST | LLG | RTGS |
|---|---|---|---|---|
| Melalui Cabang BCA* | - | Rp 2.500 | Rp 2.900 | Rp 30.000 |
| Melalui m-BCA** | Rp 6.500 | Rp 2.500 | - | - |
| Melalui Internet Banking dan myBCA | Rp 6.500 | Rp 2.500 | Rp 2.900 | Rp 25.000 |

### [Heading 3] Table Variation 2: Table with Merged Cell

A merged cell table features cells that span across multiple rows or columns, combining them into a single larger cell.

*Table 2. Caption for Merged Cell Table*

| Jenis Transaksi | Paspor Blue | Paspor Gold | Paspor Platinum |
|---|---|---|---|
| Melalui ATM BCA | | | |
| Tarik Tunai | Rp 10.000.000 | Rp 75.000.000 | Rp 10.000.000 |

| Transfer antar Rekening BCA | Rp 50.000.000 | | Rp 100.000.000 |
|---|---|---|---|
| Transfer antar Rekening BCA (ke rekening valas) | | | Rp 100.000.000 |
| Transfer antar Bank | Rp 15.000.000 | Rp 20.000.000 | Rp 25.000.000 |

### [Heading 3] Table Variation 3: Table with Missing Border

A table with missing borders is a table where some or all of the cell borders are invisible or incomplete, making it harder to visually distinguish the table cell.

*Table 3. Caption for Table with Missing Border 1*

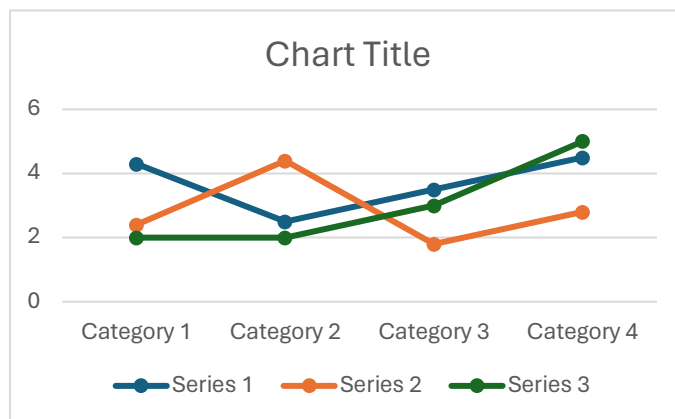| Keterangan | Online | BI-FAST | LLG | RTGS |
|---|---|---|---|---|
| Melalui Cabang BCA* | - | Rp 2.500 | Rp 2.900 | Rp 30.000 |
| Melalui m-BCA** | Rp 6.500 | Rp 2.500 | - | - |
| Melalui Internet Banking dan myBCA | Rp 6.500 | Rp 2.500 | Rp 2.900 | Rp 25.000 |

*Table 4. Caption for Table with Missing Border 2*

| Keterangan | Online | BI-FAST | LLG | RTGS |
|---|---|---|---|---|
| Melalui Cabang BCA* | - | Rp 2.500 | Rp 2.900 | Rp 30.000 |
| Melalui m-BCA** | Rp 6.500 | Rp 2.500 | - | - |
| Melalui Internet Banking dan myBCA | Rp 6.500 | Rp 2.500 | Rp 2.900 | Rp 25.000 |

## [Heading 2] Image

Image is a graphical element embedded within the text that can include photos, diagrams, charts, or illustrations. These images are typically stored in formats like PNG or JPEG and may vary in quality, resolution, and size. Here are some image example

### [Heading 3] Image Variation 1: Generated using the Microsoft Office, insert Chart feature



### [Heading 3] Image Variation 2: Downloaded from Internet

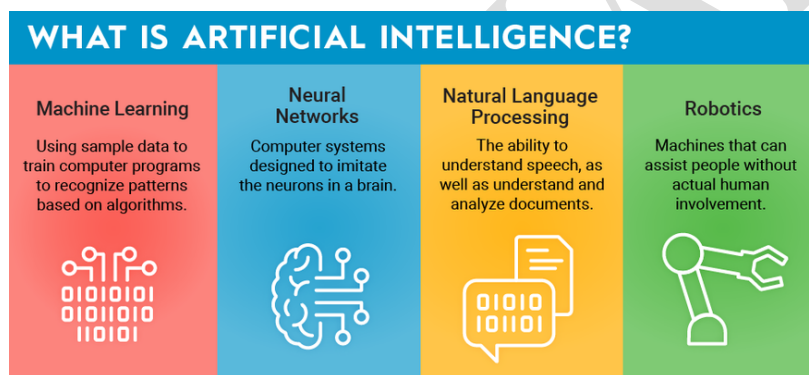"What is Artificial Intelligence?" image was retrieved from this link[ii].
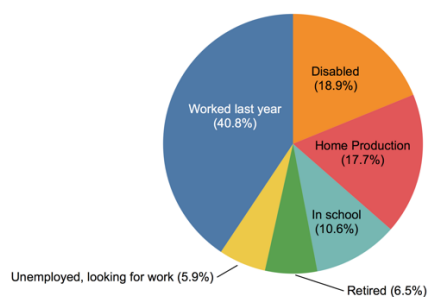


*Figure 2. what is Artificial Intelligence*

"The Working-Age Poor" image was retrieved from this link.



*Figure 1. The Working-Age Poor*

## [Heading 2] Hyperlink

<u>Hyperlinks</u> are clickable text or image elements that redirect to other locations within the document or to external resources (e.g., web pages).

## [Heading 2] Watermark

A watermark is typically a faint image or text embedded into the <span style="color:red">BACKGROUND</span> of the pages. Watermark placed in the background, meaning it should not be obstructed by other content like text or images. Since the watermark is embedded behind the primary document content, it remains visible but non-intrusive.

## [Heading 2] Math Equations

A math equation is a formula or expression included in a document, either placed within the text or displayed separately. These equations are usually formatted using typesetting tools like LaTeX or MathML for precision.

### [Heading 3] Physical Constants

speed of light in a vacuum $\rightarrow c = 2 \cdot 99792458 \times 10^8 \; m \; s^{-1}$

permeability of a vacuum $\rightarrow \mu_0 = 4\pi \times 10^{-7} \; H \; m^{-1}$

### [Heading 3] Arithmetic and Geometric Progression

A.P. $\quad S_n = a + (a+d) + (a+2d) + \cdots + [a+(n-1)d] = \frac{n}{2}[2a+(n-1)d]$

G.P. $\quad S_n = a + ar^2 = a\frac{1-r^2}{1-r}, \qquad \left(S_\infty = \frac{a}{1-r} \quad for \; |r| < 1\right)$

### [Heading 3] Convergence of series: the ratio test

$S_n = u_1 + u_2 + u_3 + \cdots + u_n \quad coverges \; as \; n \rightarrow \infty \; if \; \lim\limits_{n\to\infty}\left|\frac{u_{n+1}}{u_n}\right| < 1$

### [Heading 3] Integer Series

$\sum_1^\infty \frac{(-1)^{n+1}}{2n-1} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots = \frac{\pi}{4}$

### [Heading 3] Vector Algebra

$A \cdot B = |A||B|\cos\theta = A_x B_x + A_y B_y + A_z B_z = [A_x \quad A_y \quad A_z]\begin{bmatrix} B_x \\ B_y \\ B_z \end{bmatrix}$

### [Heading 3] Equation of a plane

$$r \equiv (x, y, z)$$

$$r \cdot \hat{d} = |d|, where\ d\ is\ the\ normal\ from\ the\ origin\ to\ the\ plane$$

---

[i] [Endnote] This is Endnote Example
[ii] [Endnote] E. N. This is Endnote Example