Requirements:

1. No password protection: Ensure that PDF files are not password-protected or encrypted. The extraction process requires free access to the content, which is not possible if the PDF is secured with a password.
2. Single-view pages: Pages should be in single-view format. Double-page spreads can confuse the extraction process, leading to inaccurate text extraction and summaries.

Best Practices:

1. Consistent formatting: Maintain consistent formatting across all pages. Inconsistent formatting can lead to errors in text extraction and summarisation.
2. High-quality scans: If dealing with scanned documents, ensure they are of high quality, as low-resolution scans can result in poor accuracy.
3. Consistent page sizes: Ensure that all pages in the PDF have consistent sizes. Varying page sizes can complicate the extraction process.
4. Remove watermarks and annotations: Watermarks, annotations, and other non-text elements can interfere with text extraction. Remove these elements, if possible.
5. Metadata and bookmarks: Ensure that the PDF includes proper metadata and bookmarks if available. While not directly affecting text extraction, metadata can help in organising and referencing the documents.
6. Use standard fonts: Use standard fonts that are easily recognisable by text extraction tools. Custom fonts can sometimes lead to incorrect character recognition.