1. The PDF is actually a scanned image

- Some PDFs are not “text PDFs” but just images of text (like scanned resumes).
- The PDF Extract node can’t read text from images—it only works on PDFs with real text layers.

2. The PDF is malformed or encrypted

- Some PDFs (like those exported from certain apps) might have an unusual structure or encryption.
- n8n’s PDF extraction library cannot parse them.

How to fix / work around it

Option A: Use OCR for scanned PDFs

- Use an **OCR service** to extract text from images inside PDFs.
- n8n doesn’t have a built-in OCR node, but you can:
  - Use **Tesseract OCR** in a **Function node** with Node.js (or Python if you have Python node).
  - Use an **external API** like Google Cloud Vision or AWS Textract:
    - Send the PDF (or convert to images per page)
    - Get text output

Option B: Convert PDF to text externally

- Use a CLI tool like pdftotext or poppler-utils if your server allows it.
- Then feed the plain text back into n8n.