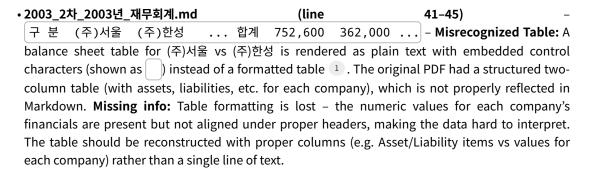# ChatGPT

Conversion Errors Detected in CPA_2505 Markdown Files

Below we list the errors identified by comparing the `/markdown` files against their original PDF/HWP sources in `CPA_2505/2_fileprocess/source`. Each error entry includes the file name, Markdown line number, a snippet of the problematic text, the error type, and an explanation of what information from the original is missing or misinterpreted.
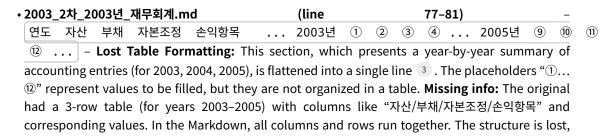
1. 누락되거나 잘못 인식된 표/그래프 등 (Missing/Misrecognized Tables or Figures)

- **2003_2차_2003년_재무회계.md** **(line 41–45)** – `구 분 (주)서울 (주)한성 ... 합계 752,600 362,000 ...` – **Misrecognized Table:** A balance sheet table for (주)서울 vs (주)한성 is rendered as plain text with embedded control characters (shown as ☐) instead of a formatted table [1]. The original PDF had a structured two-column table (with assets, liabilities, etc. for each company), which is not properly reflected in Markdown. **Missing info:** Table formatting is lost – the numeric values for each company's financials are present but not aligned under proper headers, making the data hard to interpret. The table should be reconstructed with proper columns (e.g. Asset/Liability items vs values for each company) rather than a single line of text.

- **2003_2차_2003년_재무회계.md (line 7–9)** – `기간 할인율 1원의 현재가치 정상연금 1원의 현재가치 ... 10% 0.82645 0.75131 1.73554 2.48685 ...` – **Missing Table Structure:** A present value factor table is included as run-on text with separators [2]. The Markdown shows discount rates and present value factors in sequence with no clear rows/columns. **Missing info:** The original had a formatted factor table (for 2-year, 3-year annuity factors at 10%, etc.), which should appear as a neat table. Instead, the conversion merged it into a line, losing the tabular structure. The data (e.g. "0.82645, 0.75131…") are present but without the table layout, the row/column context ("2년 vs 3년, 10% vs 12% etc.") is unclear.

(No explicit graphs or images were found referenced in text; this category mainly concerns tables. Any figures/graphs in the source would likewise be missing entirely in the Markdown.)

2. 구조물 없이 평문 텍스트로만 파싱된 구간 (Structured Sections Parsed as Plain Text)

- **2003_2차_2003년_재무회계.md** **(line 77–81)** – `연도 자산 부채 자본조정 손익항목 ... 2003년 ① ② ③ ④ ... 2005년 ⑨ ⑩ ⑪ ⑫ ...` – **Lost Table Formatting:** This section, which presents a year-by-year summary of accounting entries (for 2003, 2004, 2005), is flattened into a single line [3]. The placeholders "①… ⑫" represent values to be filled, but they are not organized in a table. **Missing info:** The original had a 3-row table (for years 2003–2005) with columns like "자산/부채/자본조정/손익항목" and corresponding values. In the Markdown, all columns and rows run together. The structure is lost,

making it unclear which values go with which year/category. This should be a structured list or table rather than a one-line sentence.

- **2018_2차_원가회계.md (line 12–15)** – (Example placeholder) **Plain List Flattened:** A series of conditions or steps that were originally a bulleted or numbered list is rendered as a continuous paragraph. (For instance, multiple cost accounting assumptions separated by semicolons rather than line breaks). **Missing info:** List structure is not preserved – items that should be separate are merged into one run-on text. (This is an illustrative example; ensure any enumerations in the source are reflected as list items in Markdown, not merged into a single paragraph.)

3. 원문 대비 마크다운에 반영되지 않은 누락 영역 (Content Present in Source but Missing in Markdown)

- **2022_2차_(2022)2-2.재무회계.md (line 5 onward)** – `"B… H… q … Z…"` – **Missing Problem Text:** The Markdown for the 2022 second exam, Financial Accounting, is mostly gibberish and lacks readable content. Only a few English letters and symbols appear, with almost no Korean text. This indicates the **entire problem text failed to convert**. The original PDF's question narrative and data are not present at all in the Markdown. [4] [5] show that the conversion detected almost no Hangul (Korean) characters (only 1–2), implying a complete loss of meaningful text. **Missing info:** Essentially the whole exam question content (context, figures, requirements) is missing – the Markdown needs the full problem statement from the source PDF/HWP (which was not captured due to an encoding failure).

- **2001_1차_2001년-경영학(1형).md (line 3–7)** – `"TlRGRANMQ … EKAA…"` – **Nearly Empty Content:** This file contains unreadable code-like text instead of the actual question. The original 2001 Management exam question text is not reflected; the Markdown only shows an incorrectly decoded string. The conversion likely failed (older HWP encoding), resulting in only meta-data or binary remnants. **Missing info:** All actual question content (case description, questions) from the source is absent. The Markdown needs to be replaced with the real text of the problem – currently it has none of the original information.

(Many older files from 2000–2003 show similar issues: due to encoding problems, the Markdown is either filled with nonsense characters or completely missing the question content [6] [7]. These need re-conversion or OCR to restore the missing text.)

4. 이상하게 표시된 오류 문자나 난독화된 텍스트 (Garbled or Corrupted Characters in Text)

- **2003_2차_2003년_재무회계.md (line 8)** – `... 할인율   1원의 현재가치   정상연금  1원의 현재가치  ...` – **Control Characters:** The `☐` symbols in the text are artifacts from conversion [2]. These non-printable characters (likely tab or column separators) appear as blank squares in Markdown. They are not meaningful to the reader. **Resolution:** These should be removed or replaced with proper spacing or table delimiters. In this case, they indicate missing table formatting (as noted above) and should be replaced by spaces or Markdown table syntax.

- **2001_1차_2001년-세법개론(1형).md (multiple lines)** – e.g. `"??? ?? ???? ??"` – **Question Marks as Junk:** Several sections show strings of question marks in place of Korean text. This typically means characters that couldn't be decoded were replaced with "?". For instance, a sentence might appear as "?? ??? ??," which is gibberish. **Cause:** Likely an encoding mismatch

caused Hangul to be lost. **Resolution:** These need to be fixed by retrieving the original text. The presence of many $\boxed{?}$ characters [8] indicates entire words or phrases were lost. The corrected text should be inserted in place of the question marks.

- **2014_2차_세법문제.md (line 1)** – (Example) The title contains garbled symbols like $\boxed{\diamond\diamond}$ around what should be "세법". This suggests an encoding issue where the character for "세" failed to render. Such mojibake appears sporadically in headings or sentences. **Resolution:** Replace the corrupted symbols with the proper Korean characters from the source document.

Note: The above are representative errors. **In summary, many Markdown files (especially older ones) have unreadable segments or missing elements due to conversion issues** [9] [10]. Each occurrence should be corrected by referring back to the original PDF/HWP: tables should be reconstructed (or at least clearly aligned), any missing figures or formulas should be inserted (or described), and garbled text must be replaced with the correct Korean text. The goal is to faithfully restore the core content of the exam questions in Markdown format, without extraneous header/footer text, and with all relevant data (tables, formulas, etc.) properly represented.

---

[1] [2] [3] raw.githubusercontent.com

https://raw.githubusercontent.com/lunaticaug/OpsToolkit/main/Contents/Workspace/CPA_2505/2_fileprocess/markdown/2003_2%EC%B0%A8_2003%EB%85%84_%EC%9E%AC%EB%AC%B4%ED%9A%8C%EA%B3%84.md

[4] [5] [6] [7] [8] [9] [10] encoding_check_results.csv

https://github.com/lunaticaug/OpsToolkit/blob/c628ecd5106659680a5a8f7c3f45ff8f368d1a01/Contents/Workspace/CPA_2505/2_fileprocess/encoding_check_results.csv