

LLMで3GPP標準化提案文書を読み切る

~要約させたら突然ハングルが出てきたので工夫した話~

NTTドコモ 6Gテック部 野崎航平

3GPPとは？

世界中のキャリア・ベンダーが集まり、スマホがどの国でも繋がるよう共通ルールを決める国際会議。

各社が技術提案を「寄書」として持ち寄り、合意形成・仕様化していく。

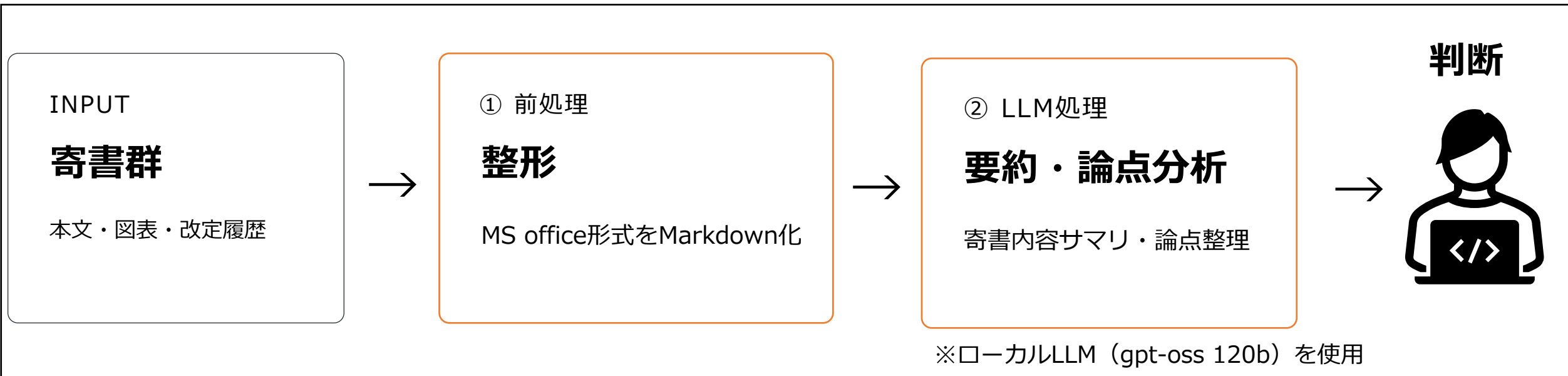
苦労点：会合1週間前、大量の提案文書の分析タイムアタックが発生

- ・ 合意形成には事前の分析が重要



LLMで標準化担当者のサポート

- 他社提案の分析と対処案策定にLLMを活用
- 結果、寄書分析時間は半分に ※社内効果アンケート結果
 - 内容把握・論点分析を補助



取り組み概要

工夫①：プロンプトを作り込む

- プロンプトでしっかりと出力指定することで、毎回同じ出力形式に。
- さらに、異なる出力形式だった場合、再生成するロジックも。

実際のプロンプトの要点

- 出力は日本語
- サマリ・主張・結論を必ず出力
- 原文にない情報は足さない

要約イメージ

- **サマリ**：6Gの～～を定義
- **主張**：UPFの～を規定
- **結論**：提案手順によって～を保証

つまづきポイント：出力が突如外国語になる

- 数十ページの英文書の重要なProposalを、日本語要約したい
- LLMに全文入力したが、正しく日本語出力できなかった
 - プロンプトを頑張ってもどうにも直らない...

- **Proposal 17**
- *Background:* バンド幅適応の信頼性・遅延を抑えるため、RAN1 による検証と「最低限」要件提示が必要です。

- **Proposal 18**

- *Background:* RAN4 の早期参画で UE 能力や 전력소비, 전파 손실 등 초기 설계 단계에서 반영해 향후 스펙 조정 시 재조정 필요 최소화.

突然ハングルに...

なぜ起きた？

- 長文のため「日本語で」の指示が忘れられる
- テキストの区切りもなく、尚更LLMは指示を見失う

⇒整形して区切りを持たせた結果、上手く行くように

工夫②：LLMに頼る前に、入力情報を整形

- 3GPPで利用されるdocx, pptxを素朴に抽出すると、**見出しや注記が区別されず**、LLMが意味を取りにくくなる

⇒**構造を明示する形に整形すると、出力ミスはなくなった**

整形前（素朴にテキスト抽出）

```
6.22.2.2.2.1      Computing service session
  establishment Procedure ← 見出し? 本文?
Figure 6.22.2.2.2.1-1: ...procedure
1.The UE sends UP Session
  Establishment Request to the eSMF.
  By including the "CC Profile ID" ← 文の続き? 別段落?
  indication, the UE...
NOTE:      The UP Session Establishment ← 注記が埋没
  Request is transported by a NAS...
2a.       The eSMF performs CCCE selection.
```



整形後（Markdown構造化）

```
#### 6.22.2.2.2.1 Procedure ← 見出しが明示
![Figure 6.22...1-1](fig1.png) ← 図も保持
1. The UE sends UP Session
  Establishment Request to eSMF.
  - By including the "CC Profile ← 従属関係も明示
  ID" indication, the UE...
> **NOTE:** The UP Session ← 注記も強調
> Establishment Request is...
2a. The eSMF performs CCCE selection.
```

- 大量・長大な標準化提案文書をLLMで要約・分析に成功
- 長文に対して出力がハングルになる現象に見舞われながらも、工夫して解消

明日から使える3つのコツ

- ・**プロンプト**
 - ・プロンプトでしつこく指定し、出力を安定させる
- ・**データ整形**
 - ・LLM入力前に、データ整形してLLMが認識しやすく
- ・**とはいえ、人も大事**
 - ・ハルシネーションや情報の欠落リスクを利用者に周知