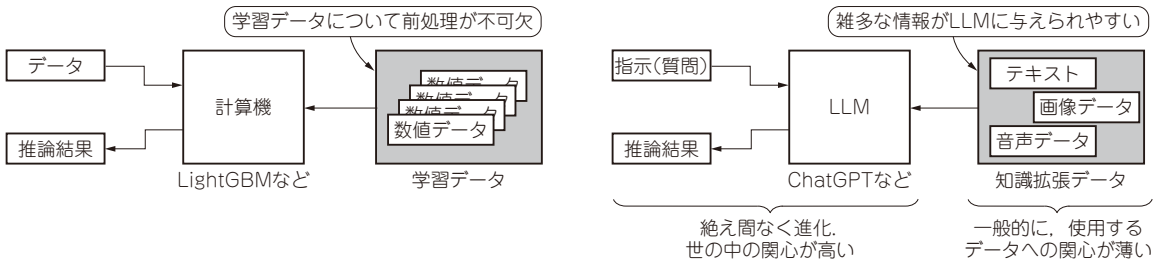
第3章 信頼できる生成結果は地道な基礎作業から始まる

ハルシネーション抑制のためのデータ整理，運用のコツ

S.K.



(a) 数値を取り扱う場合
数値データを扱うデータ・サイエンスの領域では計算機の性能やアルゴリズムがいくら進化しても，学習データに対する統計的理解に基づく分析や前処理などの地道な作業は不可欠

(b) LLMの場合
適切な前処理がなされていない雑多な情報が与えられることで，ハルシネーションが発生しやすくなっている可能がある

図1 LLMの場合は雑多な情報が使用されやすい

AI分野でのハルシネーション(hallucination，幻覚)とは，AI(生成モデル)が事実とは異なる情報や存在しない内容を，事実であるかのようにもっともらしく生成してしまう現象を指します．

本稿ではハルシネーション抑制を目的として，データの扱い方に注目し，次の3つの対策方法を紹介します．

①データを適切に整理する
②MCP(Model Context Protocol)　でLLM(Large Language Model)にデータを効果的に参照させる
③適切なプロンプトでLLMにデータを効果的に参照させる

これらはより低コストで汎用性が高く，誰でも実践可能な方法です．

ハルシネーション抑制が必要な理由

近年，ChatGPTをはじめとするLLMの出力は，より洗練され，説得力を増し，幅広い分野で活用が進んでいます．しかし，LLMの能力に依存しすぎると危険も伴います．

図1に示すように，数値データを扱うデータ・サイエンスの世界では，計算機やアルゴリズムが進化しても，データに対する統計的理解や前処理といった基礎作業が欠かせません．LLMもデータを扱う以上，同様のことが当てはまります．LLMに対して学習データが不足している事柄の質問をする場合や，学習データにはない知識を拡張する(Retrieval Augmented Generation：RAG)ためのデータそのものに対する理解や管理が不十分なまま利用されると，適切でない情報やノイズが混在し，事実と異なる内容をもっともらしく提示するハルシネーションが発生しやすくなります．

特に意思決定，研究支援，業務支援の場面では，この誤情報が大きな影響を及ぼす可能性があります．現状，ハルシネーションを完全に防ぐ決定的な方法は報告されておらず，対策が困難です[1][2]．だからこそ，ハルシネーションの存在を理解し，LLMが参照するデータの理解，前処理，管理，参照設計に注目することで，その発生を抑え，LLMをより安全かつ有効に活用することが求められます[3]．

ハルシネーションの分類と発生原因

● ハルシネーションにもさまざまな種類がある

次のように分類できます．

- **事実矛盾型(Fact-conflicting)**
  実在する事実や既知の知識に関して，誤った内容を生成するパターン．例として，歴史的日付や統計値の誤りなど