

LLMの弱点克服を目指すソフトウェアプラットフォームWISDOM-LLM

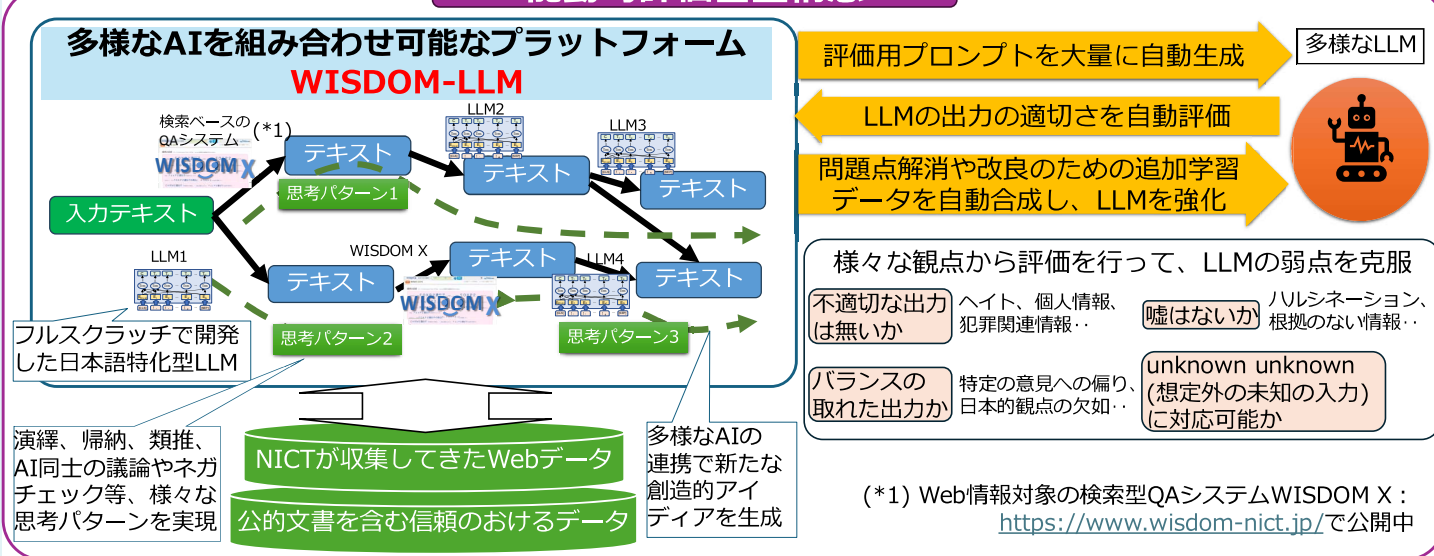
WISDOM-LLM: a Software Platform for Overcoming LLM's Weaknesses

概要 Overview

- ✓ 44TBの高品質日本語学習データを整備し、日本語特化型の独自LLMを試作
- ✓ LLMやWISDOM X(*1)等の様々なAIを組み合わせるWISDOM-LLMを開発
- ✓ LLMを自動評価し、安全性強化・高性能化を図る能動的評価基盤を構想中

- ・ 約20年にわたり、700億ページを超える日本語Webデータ(CommonCrawlの日本語データの5倍程度の規模)を独自に収集し、**高品質なLLM学習用の日本語データ(最大44TB)**を構築
- ・ **構築したデータを民間企業に提供開始**。また、このデータを使い、**日本語特化型LLM**を試作
- ・ 様々なLLMと検索型QAシステムであるWISDOM X等の多様なAIを組み合わせ可能なプラットフォーム**WISDOM-LLM**を開発。WISDOM-LLM上で、LLM生成結果の根拠情報をWebから抽出、ハルシネーションやバランスを欠いた一方的情報等を検知するサービスも開発中
- ・ 様々なLLMに関して、自動合成した評価用プロンプトで出力を自動評価し、そのリスクや弱点を特定、追加学習データを自動合成して安全性強化・高性能化を図る**能動的評価基盤**構想中

能動的評価基盤構想



(*2) Preferred Networks、さくらインターネット、情報通信研究機構、安心安全で日本社会と調和する国産生成AIのエコシステム構築に向け、基本合意を締結(2025年9月18日プレスリリース) <https://www.nict.go.jp/press/2025/09/18-1.html>

社会実装イメージ Use Cases

- 整備した日本語データや独自LLMを民間企業等に提供、信頼性の高い国産LLMを共同開発(*2)
- 多様なAIが互いに議論・評価することで不適切な情報やシステム・攻撃に民主的に対抗
- 能動的評価基盤により、国産LLMの安全性担保や、高機能化、高精度化にも貢献



国立研究開発法人情報通信研究機構 (NICT)
ユニバーサルコミュニケーション研究所 データ駆動知能システム研究センター
Mail : wisdom-contact@khn.nict.go.jp <https://direct.nict.go.jp>



LLMの弱点克服を目指すソフトウェアプラットフォームWISDOM-LLM

WISDOM-LLM: a Software Platform for Overcoming LLM's Weaknesses

概要 Overview

- ✓ 44TBの高品質日本語学習データを整備し、日本語特化型の独自LLMを試作
- ✓ LLMやWISDOM X等の様々なAIを組み合わせるWISDOM-LLMを開発
- ✓ LLMを自動評価し、安全性強化・高性能化を図る能動的評価基盤を構想中

WISDOM-CE(Credible Evidence):LLM生成テキスト検証サービス

- ・ LLMが生成したテキストに対して、根拠となりうる支持情報や、矛盾する情報・反論などを、WISDOM X 等を活用して収集済みのWeb情報から探索し、「裏取り」結果として表示
- ・ ハルシネーションの可能性を検知・提示することで、信頼性の欠如というLLMの弱点を克服

入力：乳幼児におすすめの飲み物を複数教えてください。

SEND

RESUME

SAVE

①乳幼児におすすめの飲み物を複数挙げると以下の通りになります。②・麦茶：麦茶はノンカフェインなので、赤ちゃんにも安心して与えることができます。③・白湯：母乳やミルク以外の味に慣れさせるために、白湯やお茶を飲ませることがあります。④・ルイボスティー：ルイボスティーはノンカフェインでミネラルが豊富なので、赤ちゃんにも優しい飲み物です。⑤・ほうじ茶：ほうじ茶はカフェインが少なく、苦みや渋みもほとんどないので、赤

オレンジの下線：支持情報

緑の点線の下線：矛盾情報

文②の裏取り結果

ID	裏取り対象文	裏取り関係	裏取り重要文	evidenceとなるWebページ	裏取りスコア
5	・麦茶：麦茶はノンカフェインなので、赤ちゃんにも安心して与えることができます。	支持	麦茶 ノンカフェイン 赤ちゃん 安心		0.917

文③の裏取り結果

ID	裏取り対象文	裏取り関係	裏取り重要文	evidenceとなるWebページ	裏取りスコア
1	・白湯：母乳やミルク以外の味に慣れさせるために、白湯やお茶を飲ませることがあります。	支持	母乳 ミルク以外 味に慣れ お茶 お白湯 飲ませる		0.797
2	・白湯：母乳やミルク以外の味に慣れさせるために、白湯やお茶を飲ませることがあります。	矛盾	お茶 白湯 栄養価ない いない		0.516
6	・白湯：母乳やミルク以外の味に慣れさせるために、	矛盾	白湯 推奨さ		0.708

下線が引かれていない箇所や点線の下線しか引かれていない箇所は、ハルシネーションの疑いがあると判断可能。加えてユーザは、具体的な支持情報・矛盾情報を見て、バランスの取れた総合的判断が可能に



国立研究開発法人情報通信研究機構（NICT）
ユニバーサルコミュニケーション研究所 データ駆動知能システム研究センター
Mail : wisdom-contact@khn.nict.go.jp https://direct.nict.go.jp

