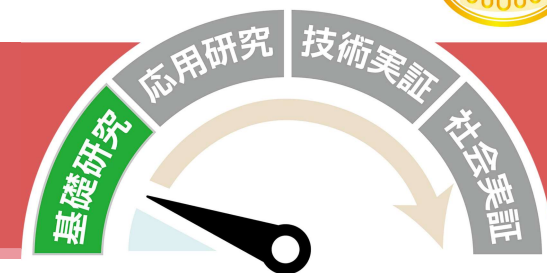


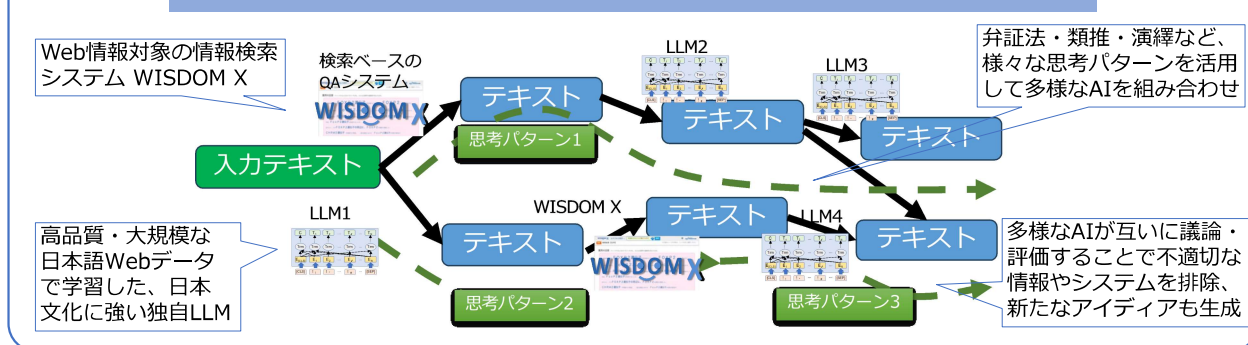
大規模言語モデル(LLM)のリスクに対する取り組み



概要

- ・「独自LLM」×「WISDOM X」でLLMがもたらし得るリスクに対抗
- ・26.2TBの独自日本語データを整備、日本文化にも強いLLMを開発
- ・整備した日本語データや開発したLLMを民間企業等へ提供開始

多様なAIを組み合わせ可能なプラットフォームWISDOM-LLM



対抗

ハルシネーション

事実ではない情報を生成

悪意を持った利用

意図的なフェイクニュースや一方的な主張を大量に生成・拡散。攻撃に活用

LLMがもたらし得る
様々なリスク

unknown unknown

今後、誰も想定していない
未知のリスクも出現

画一化

多様性、新規性を欠く類似の出力
ばかりでイノベーションを阻害

海外文化の過度の浸透

日本の主張、アイデンティティが
かき消され、日本文化の危機に。
コンテンツ産業の独自性も消滅

特徴

- ・過去15年に亘り収集した日本語中心の660億ページのWebデータから、高品質・大規模なLLM学習用の日本語データ(最大26.2TB)を整備
- ・このデータを活用し、130億から日本語特化型では国内最大規模の3,110億パラメータまで、日本文化にも強い様々なLLMを独自開発
- ・複数LLMと情報検索システム等の多様なAIを組み合わせ可能なプラットフォームWISDOM-LLMを開発

ユースケース

- ・LLM生成テキストの「裏取り」情報をWISDOM XがWebから抽出し、ハルシネーションを抑制
- ・多様な視点を持つ多様なAIが互いに議論・評価することで信頼性の高いテキストを出力、悪意を持った情報や攻撃に民主的に対抗
- ・複数のAIや様々な思考パターンの組み合わせにより創造的なシナリオを生成しイノベーションを実現。未知のリスクにも対応

今後の展開

- ・外部計算資源も活用し、LLMを大規模・高性能化
- ・民間企業等との共同研究を通じてLLMの強化を図りAll Japanで信頼性の高い国産AIを実現

【お問合せ先】

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@ml.nict.go.jp