（資料２）

「個人情報保護法のいわゆる３年ごと見直しの検討の充実に向けた視点」へのコメント（生成AIの学習用Webデータの充実という観点から）

情報通信研究機構(NICT)
鳥澤 健太郎

2024年12月3日

日本のデータで学習した国産LLMの必要性

- **海外製生成AIの学習データは英語中心で、日本語データはわずか**
  - → 日本の主張、文化、アイデンティティが海外のLLMによってかき消される可能性
  - → 使用料金流出でデジタル「農奴」（アイデンティティもなくなれば小作人というよりも農奴）
- **安全保障上や社会の健全性を担保する上でも重要な問題**
  - LLMは偽情報、セキュリティに関連して認知戦の主要兵器となるだろう
  - **海外のLLMからやってくるフェイクニュース等には国産の生成AIで対抗するより他ない**
  - 海外製LLMを日本の安全保障で活用することで思わぬ不利益を被る可能性もある
  - 国産のデータ、LLM自体が海外に流出しないようにすることも重要
- センシティブな意思決定等に使う生成AIは日本の中で閉じたシステムとすべき
  - 日本の組織のデータを海外製生成AIが学習すれば日本の打ち手の予測が容易になるかもしれない
- そもそも日本人が皆少数の海外製生成AIの言うことばかりを気にするようになれば言論、イノベーションも阻害→国産生成AIは多様性確保と言う意味でも重要
  - 生成AIはこれからますます賢くなる→多くの国民が生成AIの言うことを鵜呑みにしてしまうリスク
- 対外ビジネス上も重要な可能性
  - 日本初コンテンツ、インバウンドの人気は日本文化に立脚
  - 海外製生成AIがコンテンツ作成の支援等で使われるようになれば、日本初コンテンツの魅力も薄れるのでは？

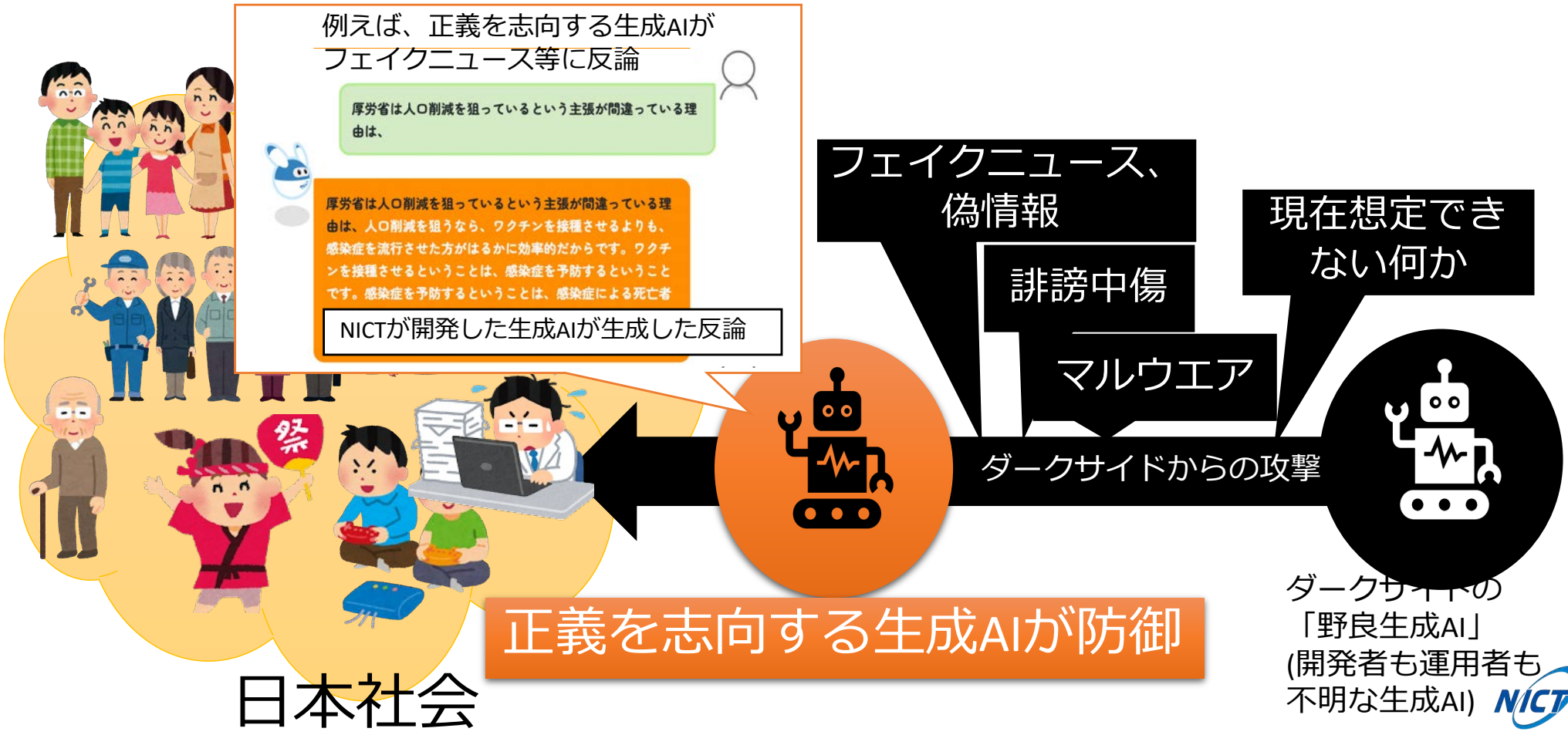
生成AIのリスクや国産生成AIの必要性についての考え方

- 月刊正論2024年5月号、「複数の『正義』で『悪』を無効化する」、鳥澤健太郎

- 日本経済新聞2024年8月7日、 私見卓見：「正義志向するAI」を国産で、鳥澤健太郎

安全保障上の課題
生成AIの悪用に対抗できるのは生成AIだけ

- 生成する情報の質はともかく、生成情報の量では生成AIは　人間、あるいは人間社会を凌駕
  - 短期間に１億人と対話した生成AIも実在。人間には不可能。
  - フェイクニュースの量も人間あるいは人間社会の対処可能量を圧倒する可能性
- 従って、近い将来、日本社会は生成AIで守るしかない

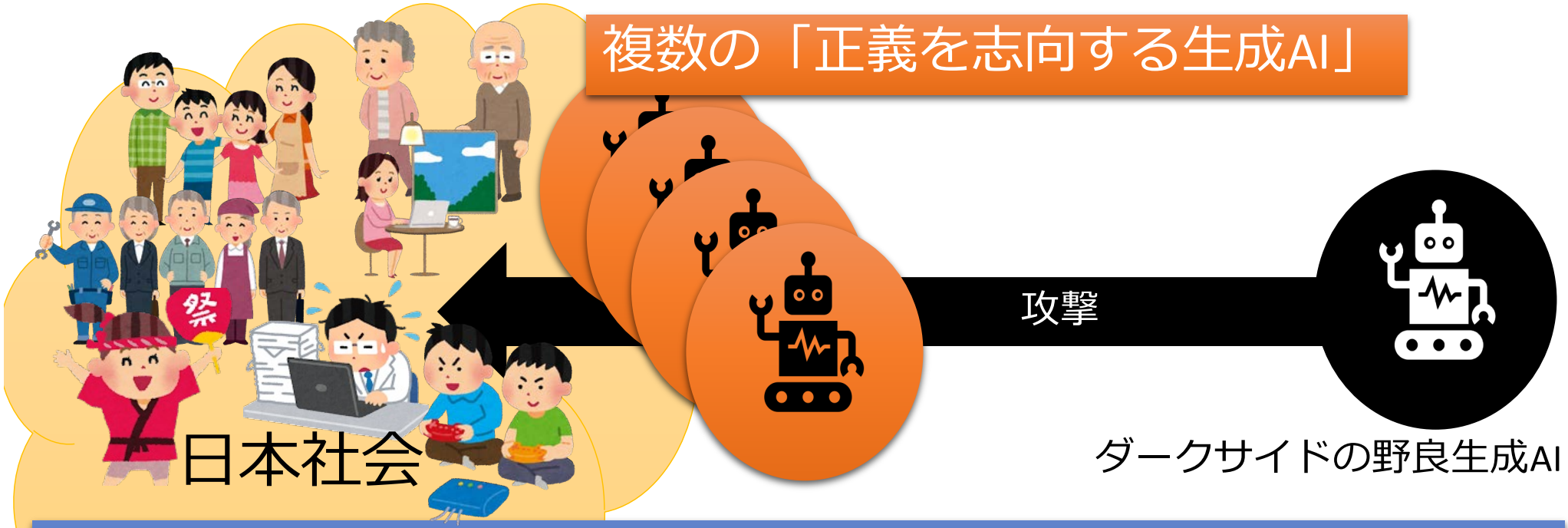


まとめ：社会を守るための生成AIの必要性

2024年8月23日
内閣府AI制度研究会
有識者プレゼン

- 日本社会を「正義を志向する生成AI」で守る必要
- 正義は一意に定まるわけではないし、正義の生成AIが意図通りに動かないこともある

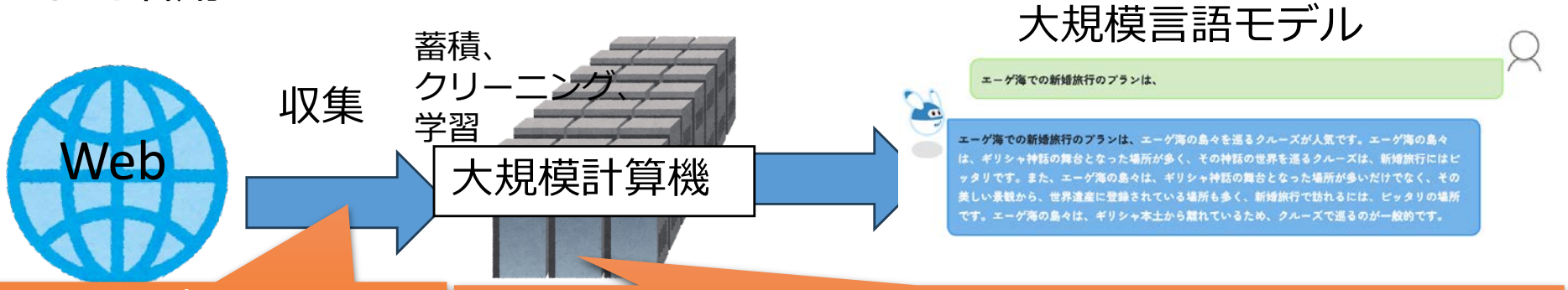
→複数の「正義を志向する生成AI」が互いに連携、議論、ネガチェックをしながら社会を守るべき



- 今後、NICTの巨大日本語データで学習した高品質な国産LLMが多数出現することが国を守る上で重要
- デジタル小作人の回避、文化・アイデンティーの維持・国際競争力の維持、という点でも極めて重要

LLMでは学習データが重要

- NICTでは過去15年間にわたりWebデータを収集・蓄積し、学習データとして活用



NICTでは、日本語ページを中心に1日あたり2,000万ページ前後収集
R5補正で1日1億ページ収集へと設備増強予定

- 過去15年間、収集した日本語中心の660億ページ(文庫本40億冊以上に相当）を大規模計算機に蓄積、活用中
- 広く使われているCommonCrawlデータの５倍程度の日本語データが含まれていると推定

- 海外の大規模言語モデルの学習データは英語中心で、日本語データはわずか
  →計算資源の量でGAFAM等に劣後する中で、日本語データの質・量は数少ない日本の勝ち筋
  →令和５年度補正予算も活用し、学術研究を研究目的に含む共同研究の枠内で大量で高品質な日本語データ（現在、最大のものは22.9TB、8T語）を民間企業等に提供開始。試作したLLMも合わせてライセンス等実施

課題はWebデータにある（要配慮）個人情報

- 現在は、学術研究を目的に含む共同研究の枠内で（要配慮）個人情報を含み得るWebデータを提供
  - Webデータは一般公開されていたデータのみ。取得に認証等が必要なWebデータは含まず
- ところが、高度な研究人材不足が言われており、さらには競争が激化する中、民間企業側で共同研究にあたる人材の確保が難しいのが実情
  - 共同研究をすると民間企業はNICTに多かれ少なかれ手の内を晒すことになる。これも民間企業等からするとデメリットとなる
- NICT側でもカウンターパートを務めるために大量のリソース、時間が必要
  - 多数の企業との共同研究＆データ提供が困難→日本社会を守る多様な国産生成AIの出現は望み薄に
- 共同研究なしで（要配慮）個人情報を含み得るWebデータが提供できれば、社会を守る、強力な国産生成AIの開発に追い風となる
  - ただし、民間企業等へのデータ提供については不適切な活用や漏洩がなされないよう、NICTとの契約で縛ることが前提
  - オープン化することは考えていない

学習データから個人情報を削除することが個人に対するリスクを抑止する上で最善の方法か？

- そもそもWebデータ中の個人情報の完全な特定、削除は技術的に容易ではない
  - 個人名の特定だけでも100%の精度は不可能
    - 仮にテストのサンプルで１００%の精度が達成できたとしても数十テラバイトの学習データですべての個人名が特定できているか否かの検証を人力でやるのは不可能
  - ましてや、電話番号等の個人名以外の情報から個人が特定される場合を検知するのはさらに困難
- 学習データからの個人情報削除以外に個人情報の漏洩を抑止するもっと簡単な方法は生成AIサービスの入出力の段階で個人名等が含まれていたら、入力・出力をさせないこと
  - 仮に個人情報の特定漏れ等により、個人から削除申請があった場合、上記の方法での対応はより容易
    - 基本的に入出力させない個人名等のブラックリストを生成AIとは独立に作り、適宜更新すれば良い
    - 仮に、生成AIの学習データから削除するのであれば、数ヶ月かかる学習を一からやり直し。学習にかかる数十億円がフイに。
  - つまり、学習データから個人情報を削除しなくても一般の個人のリスクを抑止する方法はある
  - 生成AIが学習する知識が歪むリスクを負わなくてもよい
  - また仮に学習データからの個人情報の削除が義務ということになると、個人名も個人の名誉、あるいは名誉の毀損といった概念も学習できない可能性が高まり、誹謗中傷等を認識できないAIしか作れないことになる。これはつまり、SNS上に大量の誹謗中傷が投稿された場合の対策でAIが使えなくなるということ
- 注：学習データ上で個人情報を特定できたとして、どのように削除すべきかは自明でない
  - 個人名が特定された文章等をすべて削除すると、人間一般に関する知識を生成AIが学習できなくなる恐れ
  - すべての個人名等を同一のシンボルで置き換えると生成AIが学習する知識が歪む
    - 一人の人間が宇宙飛行士で将棋のチャンピオンでMLB選手であることが可能？
  - すべての個人名の異なる出現を別のシンボルにしても知識が歪む
    - ピッチャーの大谷とバッターの大谷は別人物？二刀流は何がすごいのか？
  - 各個人名を一つのシンボルに置き換えて学習すると、文脈からシンボルに対応する個人が特定できてしまう可能性も
    - MLBで50-50を達成した人は誰？＝＞　特定可能

生成AIの学習用Webデータの充実のための考え方

- 一般公開されていたWebページをNICTのような公的機関で収集したWebデータ（（要配慮）個人情報を含み得る）を生成AIの開発、運用のために民間企業等に共同研究なしで提供することをお認めいただきたい
  - データ提供に際し、提供先の民間企業等と提供元の公的機関は契約を結ぶこととし、個人情報を含めて不適切な利用を行なわないこと、情報漏洩対策を講じること等の義務を提供先の民間企業等に負わせる
  - 社会的なリスクもあることから、オープン化、相手を問わない提供は行わない
  - また、生成AIを使ったビジネスの開始後も、トラブルシューティング、各種権利侵害やハルシネーションの防止等のため、データを生成AI運用企業等で活用することもお認めいただきたい
    - つい最近までNICTは研究を目的（個情法１７条のもの）としてデータを収集していたため、法１８条１項によりビジネスが目的に含まれる活用が困難（同条３項の例外が使えるのは実質共同研究だけ）
    - ビジネス開始後もデータを手元に置いて置けないとなると、海外勢との競争で不利になる可能性もある。ビジネス開始後にデータを手元においての各種改善等が可能でなければ、競争に勝てず、日本社会の安全性が海外の組織に委ねられる可能性もある