



Google におけるAI 利用における基本方針と 責任ある AI 利用を推進する社会の枠組み

2018年7月

グーグル合同会社
公共政策部
山口奈々子

AI 利用における基本方針 Google は以下のような基本方針を元に AI の利用を決定します。

1. 社会にとって有益である

新しいテクノロジーの普及は、ますます社会全体に影響を与えるようになっていきますAI の進歩は、医療、セキュリティ、エネルギー、交通、製造、エンターテインメントなど、幅広い分野に大きな変化をもたらすでしょうAI 技術の開発・利用の可能性を検討するにあたっては、幅広い社会的・経済的要因を考慮し、期待できる全体的な利点が、予測可能なリスクやマイナスをはるかに上回ると判断した場合に、開発・利用を進めます。

AI は、コンテンツを理解するという処理を大規模に行う能力を強化しますGoogle は製品を提供する各国の文化的、社会的、法的規範を引き続き尊重しながら、AI を活用して高品質かつ正確な情報を人々が容易に入手できるよう努めます。そして、Google のテクノロジーをどのような場合に、非営利目的で提供するかについて、引き続き慎重に検討します。

AI 利用における基本方針 Google は以下のような基本方針を元に AI の利用を決定します。

2. 不公平なバイアスの発生、助長を防ぐ

AI のアルゴリズムやデータセットは、不公平なバイアスを反映したり、強化したり、反対にそれらを緩和することもできます。あるバイアスが公平か、不公平かを判断することは必ずしも容易ではなく、文化や社会的背景によってその判断が分かれます。しかしながら、特に人種、民族、性別、国籍、所得、性的指向、能力、政治的または宗教的信念などといった繊細なトピックにおいて AI が不当な影響を与えることがないよう努めます。

3. 安全性確保を念頭においた開発と試験

害を及ぼす可能性のある予期せぬ事象の発生を避けるために Google は継続的に強固な安全性とセキュリティ対策の実践、強化に努めます。Google は自社の AI システムに適切な慎重性を組み込み、AI セーフティ研究のベストプラクティスに沿って開発を行うことを目指します。しかるべき場合には、AI 技術をコントロールされた環境下でテストし、運用開始後も動作のモニタリングを行います。

AI 利用における基本方針 Google は以下のような基本方針を元に AI の利用を決定します。

4. 人々への説明責任

Google は、フィードバックをしたり、関連する説明を求めたり、異議を唱える機会を提供できるようにAI システムを設計します。Googleの AI 技術は、人間による適切な指示、およびコントロール下に置かれます。

5. プライバシー・デザイン原則の適用

Google は、AI 技術の開発および利用にGoogle のプライバシー原則を適用します。プライバシーに関する通知および同意の機会を提供し、プライバシー保護が組み込まれたアーキテクチャを推奨すると共に、データ利用に適切な透明性およびコントロールを提供します。

AI 利用における基本方針 Google は以下のような基本方針を元に AI の利用を決定します。

6. 科学的卓越性の探求

技術革新は、科学的な手法と、オープンな考察、学問的厳密性、整合性、およびコラボレーションへのコミットメントに根ざしています。AI ツールは、科学の新たな領域を切り拓き、生物学、化学、医学、環境科学などといった分野において、新たな知識を生み出す可能性を秘めています。Google は AI 開発を進めるにあたり、高水準の科学的卓越性を目指します。

また、科学的に厳密で、学際的なアプローチを用い、多様なステークホルダーと協力して、この分野における思慮深いリーダーシップの促進に努めます。加えて、より多くの人々が有用なAI の利用方法を開発できるよう、Google は教育用コンテンツやベストプラクティス、リサーチを通じ、責任ある形でAI の知識や知見を共有します。

AI 利用における基本方針 Google は以下のような基本方針を元に AI の利用を決定します。

7. これらの基本理念に沿った利用への技術提供

多くのテクノロジーには、様々な利用方法があります。Google は、有害もしくは悪質な可能性のある利用を制限するよう努めます。AI 技術の開発や提供にあたって、Google は想定される用途を以下の要件に沿って検証します。

主な目的と用途 テクノロジーやその利用方法の主な目的と想定される用途。(これには、当該ソリューションが有害な用途にどれくらい関連しているか、もしくは有害な利用への転用可能性等の観点を含みます)

- 技術の性格や独自性 Google が提供する技術が独自性のあるものか、または一般に広く提供されているものであるか
- スケール(規模感): 当該テクノロジーの利用が重要な影響を持つかどうか
- Google の関与のあり方: どのような目的にも使える汎用的なツールの提供なのか、特定の顧客のためにツールを組み込むのか、もしくはカスタムソリューションの開発なのか

私たちが追求しないAI 利用

上記の基本方針に加えて、以下のような分野においてはAI の設計及び提供はしません。

1. 総合的にみて有害または有害な可能性があるテクノロジー。重大なリスクが認められる場合、利点が大幅にリスクを上回る場合にのみ関与し、その場合も適切な安全上の制約を組み込みます。
2. 人々に危害を与える、または人々への危害の直接的な助長を主目的とした、武器またはその他の技術。
3. 国際的に認められた規範に反するような監視のために、情報を収集、利用するためのテクノロジー。
4. 広く一般的に認められた国際法の理念や人権に反する用途のための技術

責任ある AI 利用を推進する社会の枠組み

1. 一般市民の AI に対する理解を深め信頼を高める

2. 優先産業分野において AI への理解を促進する

3. AI 実装の障壁に取り組む研究を支援する

4. 責任あるデータ共有を奨励し、AI システムのトレーニングに使用できるデータを充実させる

5. 建設的なガバナンスの枠組みを確立し政府機関の専門性を高める

6. 政府が責任ある AI 採用のロールモデルになる

7. 労働力の移行に備えて準備する