モジュール3 責任あるAIの原則

責任ある AI の原則とプラクティスを採用する

このモジュールは、責任ある AI のプラクティスを採用するのに役立つよう設計されています。 ここでは Microsoft で従っている原則、ガバナンス システム、手順の概要を示していますが、独自の AI 戦略を開発することをお勧めします。

責任ある AI の原則とプラクティスを採用する - Training | Microsoft Learn

お客様事例: 日本郵政様

- 生成 AI 活用ポータルを支えるインフラ選定では、安全安心が最重視されました。**「マイクロソフトは、公平性、信頼性と安全性、プライバシーとセキュリティ、透明性など『責任ある AI の原則』を定めています。また、顧客データを学習に利用しないと明記しています**。
- 日本郵政グループの IT 導入では、セキュリティチェックが非常に厳しく入ります。**今回、Azure OpenAI Service だからこそ、スムーズに進むことができたと思っています**。マイクロソフトも、セキュリティに対する細かい確認事項に対し迅速に応えてくれました。生成 AI 活用ポータルは JP デジタルがテナントの Azure 環境に構築されており、閉じた環境のもとでセキュリティが担保されています」

DX 文化への変革に取り組む日本郵政。Azure OpenAI Service で生成 AI 活用ポータルをリリースし、半年で 70 以上のミニアプリを作成 | Microsoft Customer Stories

責任あるAIの原則

- 「**責任あるAIの原則**」は、マイクロソフトが定義した、AIシステムの利用者が安心してAIを使えるようにするための具体的なガイドラインである

- 「**責任あるAIの原則**」を守ることは、AIシステムを安全で信頼できるものにするために非常に重要である

- **6つの原則**で構成される

責任ある AI の原則とアプローチ | Microsoft AI

AIシステムに求められる原則(1)
公平性

- 考えられる問題点: AIシステムにより、特定のグループが不公平に扱われ、差別が生じる。例：AIによる求人選考で特定の性別や人種が不利になる。

- AIシステムでは「**公平性**」が重要

- AIシステムはすべての人を公平に扱い、同様の状況にあるグループに対して異なる影響を与えないようにするべきである

AIシステムに求められる原則(2)
信頼性と安全性

- 考えられる問題点: AIシステムが、予期しない動作をし、事故や損害が発生する。例：自動運転車が誤作動して事故を起こす。

- AIシステムでは「**信頼性と安全性**」が重要

- AIシステムは信頼性が高く、安全に運用されるべきであり、設計通りに動作し、予期しない状況にも安全に対応できるようにするべきである

AIシステムに求められる原則(3)
透明性

- 考えられる問題点: AIシステムによる判定・決定の理由が不明瞭で、信頼が失われる。例：AIによるローン審査の結果が説明されず、顧客が不満を抱く。

- AIシステムでは「**透明性**」が重要

- AIシステムの決定がどのように行われたか、利用者が理解できるように、説明されるべきである

AIシステムに求められる原則(4)
プライバシーとセキュリティ

- 考えられる問題点: AIシステムから個人情報が漏洩し、プライバシーが侵害される。例：医療データへの不正アクセス

- AIシステムでは「**プライバシーとセキュリティ**」が重要

- 個人情報や企業情報を保護し、データの収集、使用、保存について透明性を持たせ、消費者がデータの使用方法を選択できるようにするべきである

AIシステムに求められる原則(5)
包括性

- 考えられる問題点: 特定のグループが排除され、社会的な不平等が拡大する。例：体が不自由な方がAIサービスを利用できない

- AIシステムでは「**包括性**」が重要

- AIシステムはすべての人々を含め、排除することなく設計されるべきである

AIシステムに求められる原則(6)
責任

- 考えられる問題点: AIシステムの誤動作に対する責任が不明確で、問題が解決されない。例：AIシステムで発生した損害に対する責任が曖昧で、適切な保証がない

- AIシステムでは「**責任**」が重要

- AIシステムの設計者や運用者はシステムの動作に対して責任を持ち、業界標準に基づいた責任の規範を確立するべきである

マイクロソフトの「責任あるAIの原則」まとめ

原則	概要	この原則が考慮されない場合のリスク
公平性 **fairness**	AIシステムはすべての人を公平に扱い、同様の状況にあるグループに対して異なる影響を与えないようにするべきである	特定のグループが不公平に扱われ、差別が生じる。例：求人選考で特定の性別や人種が不利になる。
信頼性と安全性 **reliability and safety**	AIシステムは信頼性が高く、安全に運用されるべきであり、設計通りに動作し、予期しない状況にも安全に対応できるようにするべきである	システムが予期しない動作をし、事故や損害が発生する。例：自動運転車が誤作動して事故を起こす。
透明性 **transparency**	AIシステムの決定がどのように行われたかを利用者が理解できるようにするべきである	決定の理由が不明瞭で、信頼が失われる。例：ローン審査の結果が説明されず、顧客が不満を抱く。
プライバシーとセキュリティ **privacy and security**	個人情報や企業情報を保護し、データの収集、使用、保存について透明性を持たせ、消費者がデータの使用方法を選択できるようにするべきである	個人情報が漏洩し、プライバシーが侵害される。例：医療データへの不正アクセス、個人情報の漏洩。
包括性 **inclusiveness**	AIシステムはすべての人々を含め、排除することなく設計されるべきである	特定のグループが排除され、社会的な不平等が拡大する。例：障害者がサービスを利用できない。
責任 **accountability**	AIシステムの設計者や運用者はシステムの動作に対して責任を持ち、業界標準に基づいた責任の規範を確立するべきである	システムの誤動作に対する責任が不明確で、問題が解決されない。例：AIの誤診に対する責任が曖昧で、患者が適切な治療を受けられない。

まぁ、理解できますけど、それってマイクロソフト自身のAI利用ルールですよね。
AIアプリの開発者である私にはあまり関係ないのでは？

いいえ、人ごとではありません。マイクロソフトのAI技術を利用される開発者の皆様にも関係があります！

マイクロソフトのAI技術の利用者も **責任あるAIの原則**を実践する必要がある

- マイクロソフトは、AzureなどのAIサービスを利用するすべての顧客に対し、「マイクロソフト エンタープライズ AI サービスの行動規範 (Code of Conduct)」への同意と遵守を求めている
- この行動規範には、**責任あるAIの原則**に沿った具体的な実践事項が定められており、顧客はそれらを「善意をもって」実行する必要がある

詳しくはこちら

Code of Conduct for Microsoft AI Services | Microsoft Learn

項目	マイクロソフトの責任	開発者・利用者の責任
モデルの設計・訓練	公平性・透明性・安全性を考慮したAIの開発	利用するモデルの特性を理解し、適切に使う
プラットフォームの提供	セキュアで信頼性の高いAIサービスの提供	APIやツールの利用における倫理的配慮
利用ガイドライン	Responsible AI Standardの策定と公開	ガイドラインに沿ったアプリケーション設計
誤用防止	濫用検知・制限機能の実装	誤用・悪用を防ぐ設計と運用の実施

信頼できるAIシステムの実現に向けて

- **信頼できるAIシステムの実現には、顧客とマイクロソフト双方の取り組みが欠かせない。**（いわゆる「責任共有モデル」）
- ぜひみなさまのAIシステムの開発・運用でも「責任あるAIの原則」を活用し、安心して利用できるAIシステムの開発に役立ててしてください！