AI 開発におけるセキュリティ・プライバシー・倫理・法令に関する開発者の認識と対策

髙瀬 由梨[1,a)]　秋山 満昭[2]　戸田 宇亮[1,5]　若井 琢朗[1]　荒井 ひろみ[5]　大木 哲史[3,5]　森 達哉[1,4,5]

概要：AI 技術の急速な進展に伴い，セキュリティ，プライバシー，倫理，法令に関する要素はますます重要性を増している．このため，AI 開発者がこれらの要素に対して適切かつ深い知識を有することは，産業および社会全体において極めて重要である．本研究は，AI 開発者がこれらの要素をどのように認識し，具体的にどのような対策を実施しているかを包括的に調査し，これらの実践において必要となる課題を明らかにすることを目的としている．本研究では，日常的に AI 開発に従事する開発者を対象に詳細な質問紙調査を実施し，彼らの認識や知識を評価した．調査は，選択肢式および自由記述式の質問を組み合わせることで，定量的および定性的なデータを収集し，多角的な視点から認識と課題を明らかにすることを狙いとした．調査の結果，AI 開発者の多くはセキュリティやプライバシーリスクを認識している一方，実際の対策は不十分であること，透明性や公平性の確保は必ずしも十分でないこと，そして法規制の遵守に関しても課題が残ることが明らかとなった．これらの結果から，今後さらなる教育や支援が必要であることが示唆される．

キーワード：AI 開発，開発者向けユーザスタディ研究

A Comprehensive Survey on AI Developers' Understanding of Security, Privacy, Ethics, and Regulations

Yuri Takase[1,a)]　Mituaki Akiyama[2]　Takaaki Toda[1,5]　Takuro Wakai[1]　Hiromi Arai[5]
Tetsushi Ohki[3,5]　Tatsuya Mori[1,4,5]

Abstract: As AI technology rapidly advances, the importance of addressing security, privacy, ethics, and regulatory compliance is growing. Therefore, it is critical that AI developers have deep and appropriate knowledge in these areas to ensure the well-being of the industry and society as a whole. This study aims to comprehensively investigate how AI developers perceive these critical factors and what specific measures they are implementing. Through a detailed questionnaire survey of developers actively involved in AI development, we assessed their knowledge and awareness. The survey combined multiple-choice and open-ended questions to collect both quantitative and qualitative data, providing a multi-faceted understanding of perceptions and challenges. The results show that while many AI developers are aware of security and privacy risks, their actual implementation of countermeasures is often inadequate. In addition, efforts to ensure transparency and fairness are not always adequate, and compliance with regulatory requirements remains a challenge. These findings suggest the need for increased education and support in these areas going forward.

Keywords: AI development, user study research for developers

[1] 早稲田大学 / Waseda University
[2] NTT 社会情報研究所 / NTT
[3] 静岡大学 / Shizuoka University
[4] 情報通信研究機構 / NICT
[5] 理化学研究所 革新知能統合研究センター / RIKEN AIP
[a)] yuri@nsl.cs.waseda.ac.jp

1. はじめに

AI 技術は近年急速に進化し，医療，交通，製造業など多くの分野で革新をもたらしている．例えば，医療分野で

© 2024 Information Processing Society of Japan

This paper is work in progress and not peer-reviewed.

は AI による診断支援システムが医師の診断精度を向上させ，交通分野では自動運転技術が安全性と効率性を高めている [5], [11]．その一方で，AI 技術の普及と進化に伴い，**セキュリティ**，**プライバシー**，**倫理**，**法令遵守**の問題が重要視されている [2], [8], [9], [10].

AI システムのセキュリティは，外部攻撃からの保護とデータの機密性・完全性の確保に欠かせない．プライバシー保護はユーザーの信頼を得るために不可欠であり，倫理的な観点ではアルゴリズムの公平性や透明性が重要な課題である．法令遵守はこれらの要素を統合的に管理し，AI 技術の適切な運用を保証する．これらの要素は，AI 技術の持続可能な発展と社会的受容を実現するための基盤となる．

本研究の目的は，AI 開発者がこれらの重要な要素に対してどの程度の認識や理解を持っているかを明らかにし，それらの知識がどのように形成され，どのような課題が存在するかを探ることである．この目的を達成するために，本研究では AI 開発者を対象にオンラインアンケートを実施し，選択肢式および自由記述式の質問を組み合わせることで定量的かつ定性的なデータを収集した．これにより，AI 開発者の知識と理解度に関する包括的な分析を行う．本調査で得られた発見のハイライトは以下の通りである．

セキュリティ・プライバシ： 75%の開発者が AI モデルの利用に伴うセキュリティリスクを認識し，58%の開発者がモデルの公開にプライバシーリスクが伴うことを認識している一方，実際にセキュリティ・プライバシーに関する対策を実施している開発者は限られている．

倫理： AI システムの透明性を確保していると回答した開発者は 42%，公平性を確保しているのは 25%と，透明性や公平性の確保に対する取り組みが不十分であることが示された．

法規制： 83%の開発者がデータセットやモデルのライセンスを確認している一方，モデルの訓練に利用したデータの出所を明示している割合は 58%に留まった．

2. 関連研究

以下に，AI 開発におけるセキュリティ，プライバシ，倫理に関する関連文献を示す．

セキュリティ:

Kathikar ら [3] は，Hugging Face プラットフォームにおける AI モデルの脆弱性を評価した．約 110,000 のモデルを対象にメタデータの解析を行い，モデルカードと GitHub リポジトリとの関連性，静的および動的脆弱性スキャナーを用いたセキュリティ評価手法を検討した．結果として，高セキュリティ脆弱性が 35.98% を占めることが示され，AI モデルのセキュリティ管理の重要性が強調された．この研究は，AI リスク管理やセキュリティ対策の基盤を提供しており，本研究におけるモデルのセキュリティ評価に対する視点を強化するものである．

プライバシー:

Lee ら [4] は，AI 製品の開発におけるプライバシー作業のスコープ，動機，実施方法について調査した．35 名の業界実務者へのインタビューを通じて，プライバシーがどのように定義され，動機付けられ，実践されているかを明らかにしている．参加者は，プライバシーを AI 技術が引き起こすリスクと関連付けて捉えており，特に消費者向け AI 製品においては独自のプライバシーリスクが存在することが示された．また，プライバシー作業に対する動機は倫理的な責任感や法的な要求から来る一方で，リソースの制約や他の業務目標との優先順位の違いが障害となっていることが明らかになった．Lee らは，プライバシーに関する意識，動機，能力を同時に向上させるための統合的なアプローチが必要であると提唱し，AI 特有のプライバシー脅威に対処するためのツールやアーティファクトの開発が求められている．この研究は，AI 開発における倫理的なプライバシー実践を促進するための基盤を提供するものであり，原則と実践の間に存在する「ギャップ」を埋めるための具体的な手段を提案している．

倫理:

Pant ら [6] は，AI 実務者の倫理的認識と，AI システムへの倫理の組み込みに関する課題を実証的に調査した．情報技術分野の専門家 100 名を対象に，倫理的認識と実務で直面する具体的な問題を収集した．この研究は，AI 開発における倫理的配慮の重要性を示し，倫理の実装における障害を明らかにした．

Pepe ら [7] は，Hugging Face ハブにホストされている事前学習済みトランスフォーマーモデルの透明性に関する調査を行った．トレーニングデータセットの開示，バイアスの議論，ライセンスの宣言という 3 つの側面に焦点を当て，モデルの透明性向上の状況を分析した．結果として，モデルの 14% がデータセットの開示を行い，61% がトレーニングデータセットを文書化している一方で，バイアスやライセンスに関する開示は依然として不足していることが示された．

Boyd ら [1] は，機械学習エンジニアが倫理的問題を認識し，理解するプロセスに焦点を当て，Datasheets for Datasets というアプローチがどのように役立つかを探求した．実験により，Datasheets がエンジニアに倫理的な問題を早期に認識させ，理解を深める助けとなることが示された．この研究は，データセットの文脈や特性に関する情報を明示することで，倫理的意思決定を支援し，透明性を促進する方法を提案している．

© 2024 Information Processing Society of Japan

3. 調査方法

3.1 質問紙の設計

質問紙は，選択肢式と自由記述式の質問を組み合わせた形式で設計した．質問項目は，参加者の基本情報（年齢，性別，職種，経験年数など）に加え，セキュリティ技術，プライバシー，倫理，法令の領域に関する知識と理解度を評価する内容を含んでいる．質問設計に際して，初期のパイロットテストを実施し，質問項目の明瞭さと理解しやすさを確認し，必要に応じて修正を行った．これにより，調査項目の信頼性と妥当性を高めた．

3.2 データ収集・分析方法

AI 開発者を対象に，前述の質問紙を用いた調査を実施した．対象者は，AI 関連の経験を持つエンジニア，会社員，プログラマー，開発者などとし，多様性を確保するために年齢，性別，業務領域を考慮して選定した．参加者の選定基準としては，AI 開発に関する実務経験があること，アンケート調査への自主的な参加意欲があることとした．これにより，多様なバックグラウンドを持つ参加者からのデータを収集し，研究の信頼性と一般化可能性を高めた．データ収集はオンライン形式で実施し，調査は 8 月 1 日に配布を開始し，8 月 8 日までの期間で回答を募集した．協力者の募集は，AI 関連のコミュニティや SNS を通じて実施した．最終的に，サンプルサイズは 55 人となった．

データ分析には，選択肢形式の質問に対しては統計分析を，自由記述式質問に対してはテーマ分析を採用した．定量データは，選択質問の回答を統計的手法を用いて分析し，参加者の知識や理解度に関する詳細な洞察を得た（4.2,4.3 節）．定性データは，自由記述の回答を内容に基づいてテーマごとに分類し，パターンや傾向を明らかにした（4.4 節）．

3.3 倫理的配慮

調査の実施にあたっては，参加者から事前にインフォームド・コンセント（同意）を取得した．参加者には，調査の目的，方法，データの使用目的，プライバシー保護に関する情報を十分に説明し，同意を得た上でアンケートを開始した．不特定多数を対象としたオンライン調査の場合は，報酬を提供した．想定作業時間が 20 分のため，報酬は 450 円とした．一方，プロフェッショナルを対象とした調査であり，調査の意義を理解していただいた上でボランティアとして参加してもらった場合は，報酬は支払わなかった．また，データ収集プロセスにおいては，氏名は募集せず匿名性を厳守し，プライバシー保護に最大限配慮した．

表 1　参加者の基本統計（性別）

性別	人数	割合
男性	55	100%
女性	0	0%

表 2　参加者の基本統計（年齢，職種）

年齢	人数	割合	職種	人数	割合
20〜25 歳	5	9%	エンジニア	18	34%
26〜30 歳	14	26%	会社員	13	24%
31〜35 歳	13	24%	開発職	6	11%
36〜40 歳	3	5%	技術職	3	5%
41〜45 歳	9	16%	プログラマー	3	5%
46〜50 歳	4	7%	学生	3	5%
50 歳以上	7	13%	その他	9	16%

表 3　参加者の基本統計（AI 開発経験年数）

経験年数	人数	割合
半年未満	12	22%
半年以上 1 年未満	13	24%
1 年以上 2 年未満	24	44%
2 年 5 年未満	6	10%
5 年以上 10 年未満	0	0%

表 4　AI 開発時（モデル選択，API 利用時）に最も重視する項目．(1) 性能・機能面での評価，(2) セキュリティ評価

属性	(1) 割合	(2) 割合
ダウンロード数	26%	30%
いいね数	18%	16%
論文への参照数	12%	19%
モデルサイズ	24%	10%
配布者	15%	20%
その他	5%	5%

4. 調査結果

4.1 参加者の内訳

表 1，2，3 に実験参加者の内訳を示す．全般に，参加者の属性には偏りがあり，参加者全員が男性，年齢分布は 20–30 代の層が大多数，AI 開発年数に関しては 2 年未満の層が大多数であった．これは実際の AI 開発者の分布を反映した結果であると考えられるが，このような偏りが調査に与える影響の評価は今後の課題である．

4.2 AI 開発において重視する項目

はじめに，AI 開発者が AI 開発を進める上で，重視する項目を調査する．ここでの調査対象は，具体的な状況として AI 開発者が AI モデルを選択する，あるいは AI サービスを提供する API を利用する際に (1) 性能や機能面で評価する点，(2) セキュリティ面で評価する点である．表 4 に

© 2024 Information Processing Society of Japan

結果を示す．参加者は選択肢の中から，最も重視する項目を 1 つ選択した．モデルや API の性能，機能面を評価する際は「ダウンロード数」(26 %) と「モデルサイズ」(24 %) が特に重視されており，「いいね数」(18 %) や「論文への参照数」(12 %) も一定の影響を持つことが分かった．一方，セキュリティ，プライバシー，公平性，法的規制に関する言及は少なく，「その他」に含まれる内容（ライセンス，利用可能なライブラリやインターフェース，モデルや API の実際の性能，利用料金，評判等）にセキュリティ対策の有無に関する言及が存在するに留まった．

一方，モデルや API のセキュリティを評価する際には，「ダウンロード数」(30 %) や「配布者」(20 %)，さらに「論文への参照数」(19 %) が重要視されており，実装や基礎技術の提供元に関する信頼性に関連する項目が重視される傾向が見られた．「その他」に含まれた項目として，周辺ソフトや AI システムアーキテクチャ，安全な配布形式，カスタムコードの使用有無，ウイルススキャンの結果，プライバシーやアップデート情報の有無，開発者・組織の連絡先の適切さなどが挙げられた．

4.3 AI 開発者の認識調査（選択式質問）

AI 開発者の認識および対策に関して調査した結果を表 5 に示す．いずれの質問項目も Yes/No の二択の選択式質問であり，認識や実施している対策の有無を定量評価することを狙いとする．主要な結果は以下の通りである．

セキュリティ： AI 開発者は，AI モデルの利用に伴うセキュリティリスクを 75%が認識しており，また，71%がセキュリティ対策の改善が必要であると考えている．一方，実際に自ら開発した AI モデルを公開した経験がある開発者は 22%に留まり，今回の調査対象となった開発者群においては，より高度で責任の伴う AI モデル自体の開発を実行している開発者は少数であることがわかる．

プライバシー： プライバシーに関しては，58%の開発者が公開モデルによる新たなプライバシーリスクの可能性を認識している．その一方で，使用するデータに個人情報やプライバシーに関わる情報が含まれていないことを確認している開発者は 42%に留まる．これらの結果から，開発者はデータのプライバシーに対する一定の意識を持っているものの，すべての開発者がプライバシーリスクを十分に管理できているわけではないことが示唆される．

倫理： AI システムの透明性を確保していると答えた開発者は 42%であった．また，データにバイアスがかからないようにしていると回答した回答者が 42%である一方，「公平性を確保している」と回答した開発者は 25%に留まる．この結果は，一定数の開発者は透明性や公平性の確保に取り組んでいるものの，その対応が十分でないことを示している．

表 5 AI 開発者の認識と対策 (Yes/No における Yes の割合)

項目	質問	Yes
セキュリティ	AI モデルの利用において直面するセキュリティリスクを認識していますか？	75%
	AI 開発に関するセキュリティ対策について，改善が必要だと思いますか？	71%
	自身が開発したモデルを公開したことがありますか？	22%
プライバシー	公開したモデルがプライバシーリスクを生む可能性を認識していますか？	58%
	使用するデータに個人情報などプライバシに関わる情報を使用していないことを確かめていますか？	42%
倫理	AI システムの透明性を確保していますか？	42%
	AI システムの公平性を確保していますか？	25%
	目的に応じて使用するデータにバイアスがかからないようにしていますか？	42%
法的規制	AI 開発において使用するデータセットやモデルのライセンスを読んでいますか？	83%
	AI モデルの訓練に使用するデータの出所を明示していますか？	58%

法的規制： 83%の開発者が使用するデータセットやモデルのライセンスを確認しているが，データの出所を明示しているのは 58%である．このことから，ライセンスの確認は広く行われているものの，データの出所やバイアス対策についての対応が不十分であることが示されている．

これらの結果は，AI 開発者がセキュリティやプライバシーに対する認識を持っているものの，実際の対策や実践には限界があり，セキュリティ対策やデータの取り扱い，AI システムの透明性と公平性に関する教育や支援が重要であることを示唆している．

4.4 AI 開発者の認識調査（自由記述）

以下では，自由記述質問に対する回答をもとに，セキュリティ，プライバシー，倫理，法的規制のそれぞれの項目に対する開発者の認識を集計した結果を報告する．自由記述のコーディングは主著者が 1 名で実施した．

4.4.1 セキュリティリスク・対策に関する認識

AI モデルの利用時におけるセキュリティリスクに関する認識を集計した結果を表 6 に示す．比較的多くの AI 開発者が，AI モデルにおける情報流出を重要なセキュリティ

© 2024 Information Processing Society of Japan

表 6　AI モデル利用時におけるセキュリティリスクに関する認識

カテゴリ	説明	割合
情報流出	AI が学習した機密情報や個人情報が外部に漏洩するリスク	27%
人的要因	人的エラーやリテラシーの低さによるリスク	20%
モデル信頼性	モデルの悪用や不適切なデータ学習に関するリスク	15%
脆弱性	任意コード実行や外部攻撃による脆弱性のリスク	6%
ファイル形式	Pickle 形式などのファイルが悪意のあるコードを含む可能性があるリスク	3%

表 7　AI 開発におけるセキュリティ対策の改善が必要な項目の認識

カテゴリ	説明	割合
データ管理	データの検証・整理を含む AI モデルに使用するデータの取り扱い対策	22%
セキュリティチェック	脆弱性の定期的な確認と対策見直し	18%
モデルの工夫	セキュリティ向上のためのモデル設計や実装の工夫	7%
ユーザー教育	ユーザーに対するセキュリティ意識の向上や教育	7%
知識・情報	セキュリティ対策の具体的な改善方法が分からない，または情報が不足している状況	5%

リスクとして考えており，特に機密情報や個人情報が外部に漏洩する可能性が高いと認識している（27%）．人的要因としては，人的エラーやリテラシーの低さがセキュリティリスクを引き起こすとの意見が多く見られる（20%）．モデルの信頼性に関しては，悪用リスクや不適切なデータ学習によるリスクが指摘されており（15%），フェイクニュースやスパムの生成に悪用される可能性が懸念されている．セキュリティ脆弱性については，任意コード実行や外部からの攻撃リスクが重要視されており（6%），具体的なリスクとして，Pickle 形式などに含まれる悪意のあるコードの可能性が指摘された（3%）．

表 7 に AI 開発に必要なセキュリティ対策に関して，改善が必要だと考えられる項目に関する回答をまとめた結果を示す．比較的多くの AI 開発者が，AI 開発における「データ管理」の重要性を強調しており，データの検証・整理を含む AI モデルに使用するデータの適切な取り扱いが求められている（22%）．また，脆弱性の定期的な確認や対策の見直しが必要であるとの意見が多く，セキュリティチェックの強化が重要視されている（18%）．さらに，セキュリティ向上のためにはモデル設計や実装における技術的工夫が求められており（7%），これにはバックドア対策やリスク評価の実施が含まれる．ユーザーに対するセキュリティ意識の向上や教育の重要性も指摘されており（7%），知識や情報の不足がセキュリティ対策の改善を妨げる要因とされている（5%）．

表 8　公開した AI モデルのプライバシーリスクに関する認識

カテゴリ	説明	割合
個人情報	モデルが個人情報を学習・出力し，プライバシーが侵害されるリスク	50%
プライバシー保護	プライバシーリスクを軽減するための対策	17%
リスクの認識不足	プライバシーリスクの認識が不足している状態	17%
リスクの非認知	プライバシーリスクが存在しないと考える意見	8%

4.4.2　プライバシー

表 8 に公開した AI モデルのプライバシーリスクに関する認識の回答を集計した結果を示す．多くの AI 開発者が，AI モデルによる「個人情報」のリスクを指摘しており，モデルが個人情報を学習・出力し，プライバシーが侵害される可能性が高いと考えられている（50%）．このため，プライバシーリスクを軽減するための対策が必要であるとされているが，実際にこれらの対策が十分に実施されているかどうかについては，さらなる検証が求められている（17%）．また，プライバシーリスクの認識が不足している状態が問題視されており（17%），効果的な対策を講じるためには教育や意識向上が不可欠である．一方で，プライバシーリスクが存在しないと考える意見もあり（8%），リスク評価やその認識についての不十分さが示唆されている．

表 9 に AI が利用するデータにプライバシーに関わる情報が含まれている場合の対応に関する回答を集計した結果を示す．比較的多くの AI 開発者は「データ削除・逆学習・除去」により個人情報を削除し，モデルから逆学習（忘却処理）を実施していることが見てとれる（38%）．また，個人情報対応では，情報のマスキングや削除，匿名化，統計処理などを通じて特定されないようにする取り組みが行われている（23%）．個人情報の有無が不明な場合にそのままにするかどうか，AI の偏りに配慮する判断や配慮が求められているが，この点についても課題が残っている（15%）．データに関する問題や損害が発生した際には，速やかな報告が必要とされており（15%），適切な対応体制の整備が求められていることも明らかになった．プライバシー保護を実現するためのセキュリティ対策としては，データ転送や保存時の暗号化が不十分な場合のリスクが指摘されており，セキュリティソフトや暗号化の実施が行われているものの，より一層の対策強化が必要である（7%）．

© 2024 Information Processing Society of Japan

表 9　AI が利用するデータに対するプライバシ保護対策の認識

カテゴリ	説明	割合
データ削除・逆学習・除去	個人情報を削除し，モデルから逆学習（忘却処理）を実施	38%
個人情報対応	個人情報をマスク，削除，匿名化，統計処理し，特定されないようにする	23%
判断・配慮	個人情報の有無が不明な場合，そのままにするか，AI の偏りに配慮	15%
損害報告	データに関する問題や損害を速やかに関係者に報告	15%
セキュリティ対策	セキュリティソフトや暗号化など，一般的なセキュリティ対策を実施	7%

表 10　AI システムの透明性確保に向けた対策と課題の集計結果

カテゴリ	説明	割合
データ公開	モデルのトレーニングデータやコードを公に公開	41%
公開条件付き	特定条件下でのみデータや情報を公開，または稀に公開	29%
意識と定義の欠如	透明性の重要性を認識していない，または定義が不明で対応が行われていない	21%
情報共有	開発プロセスや結果をチーム内で共有，ただし公開はしない	9%

4.4.3　倫理 (透明性・公平性)

表 10 に AI システムの透明性確保に向けて実施している対策や課題と感じている内容を集計した結果を示す．比較的多くの AI 開発者は，モデルのトレーニングデータやコードを公開することで透明性を確保していると報告している（41%）．しかし，特定の条件下でのみデータや情報を公開するか，稀にしか公開しないケースも多く見られ，公開の頻度や範囲に制約があることが明らかになった（29%）．一方で，透明性の重要性を認識していない，または定義が不明で対応が不十分な状況があり，開発者間での認識の不足が明らかになった（21%）．また，開発プロセスや結果をチーム内で共有するものの，外部への公開は行われていないケースが多く，情報共有のあり方にも課題が残っている（9%）．

表 11 に AI システムにおける公平性確保に向けた対策と課題に関する回答を集計した結果を示す．多くの回答者が，AI システムにおける「意識と定義の欠如」を指摘しており，公平性の概念やその重要性についての認識が不足していることが明らかになった（50%）．また，深層学習のプロセスや成果物を適切なライセンスで公開し，透明性を維持する対策が行われている（17%）．一方，公平性の完全な確保が不可能であるとする意見が多く（17%），技術的な制約や倫理的な難しさが影響していることが示されている．運用中に問題が発見された場合，客観的な意見を基にシステムを改善する意識が見られる（8%）．AI の予測結果に差別や偏見が生じる可能性が認識されているものの，対策が行われていないケースが多い（8%）．

表 11　AI システムの公平性確保に向けた対策と課題の集計結果

カテゴリ	説明	割合
意識と定義の欠如	公平性の重要性や概念が不明で，対応が行われていない	50%
プロセスとライセンスの公開	深層学習のプロセスや成果物を適切なライセンスで公開し，透明性を維持	17%
公平性の限界	公平性の完全な確保が不可能と考えている	17%
運用時の改善	運用中に問題が発見された場合，客観的意見を基にシステムを改善	8%
データとアルゴリズムの偏り	AI の予測結果に差別や偏見が生じる可能性があることを認識しているが，対策が行われていない	8%

表 12　AI モデルの訓練に使用するデータの出自明記に関する対策と課題の集計結果

カテゴリ	説明	割合
出所明示の方法	データ出所の明示方法（例：README 記載，簡易・詳細な方法）	33%
法的・倫理的配慮	法的・倫理的リスクを考慮した出所非公開の理由と炎上リスク対策	25%
ライセンスと公開	ライセンス情報の表示やデータ公開の取り組み	17%
社内方針とリスク管理	社内方針に基づく出所明示の有無やデータ管理，炎上リスクの対策	17%
ガイドライン公開	データの出所に関するガイドラインの作成と公開	8%

4.4.4　法的規制

表 12 に AI モデルの訓練に使用するデータの出所の明示に関して実践している対策や課題について集計した結果を示す．多くの回答者がデータ出所を明示していると回答しており，その方法には README ファイルへの記載や簡易なデータセットの公開などが含まれている（33%）．しかし，標準化された方法は存在せず，一貫した対応が取られていない現状が浮き彫りとなっている．法的および倫理的配慮も大きな影響を及ぼしており，出所非公開の理由と

© 2024 Information Processing Society of Japan

して，著作権やパブリシティ権の侵害リスクに加えて，社会的な反応を考慮した対応が行われている（25%）．ライセンス情報の表示やデータ公開に関する取り組みも進展しており（17%），データ利用の透明性向上に寄与している．しかし，社内方針や炎上リスクがデータ出所の明示に影響を与えているため，全ての組織で統一されたガイドラインの整備が進んでいるわけではなく，標準的なガイドラインの普及が求められている（8%）．

5. 議論

本章では，4章で得られた発見に基づき，得られた知見を整理するとともに，今後の課題を示す．

5.1 AI 開発で重視される項目

AI 開発者がモデル選択や API 利用において重視する項目と，モデルのセキュリティ評価において重視する項目には，共通点と相違点があることが明らかになった．モデル選択や API 利用の際には，ダウンロード数やモデルサイズといった，利用実績やリソース効率性に関連する属性が重視されていることがわかる．これは，開発者が実際に使用するモデルや API が，広く利用されていることや効率的であることを期待していることを示唆している．一方で，セキュリティ評価においては，ダウンロード数に加え，配布者や論文への参照数といった信頼性に関連する属性がより重要視されていることが見て取れる．ここでの「信頼性」とは，モデルや API が安定して動作し，脆弱性がなく，技術的に信頼できること，また提供元が業界内で評価されている権威や実績を持つことを意味する．これらの観点から，セキュリティ評価においては，技術的性能だけでなく，提供元の信頼性やモデルの実績が重要な役割を果たしていることが示唆されており，開発者がセキュリティを評価する際に，多角的かつ実務的な視点を持っていることが明らかになった．また，セキュリティ評価に関連するその他の要素として，安全な配布形式や周辺ソフト，ネットワーク構成など，非常に具体的な要素が言及されていることからも，開発者がセキュリティを評価する際に，多角的かつ実務的な視点を持っていることが明らかになった．

AI 開発者が直面する現実的な課題や要求に応じた，柔軟かつ信頼性の高いモデルや API の提供が，今後の AI 開発においてますます重要になる．これを実現するためには，まずユーザー中心の設計を重視し，開発者の具体的なニーズに応じた機能やサポートを提供することが求められる．また，セキュリティとコンプライアンスの強化も欠かせない．セキュリティ標準や法規制に準拠し，最新の脆弱性診断やデータ管理の確保を行うことで，信頼性を高めることができる．さらに，透明性と説明可能性の確保も重要であり，モデルの内部動作が明確で説明可能であることが求められる．そして，継続的なアップデートとメンテナンス，スケーラビリティと互換性の確保を通じて，開発者の多様な環境やニーズに対応できるようにすることが，柔軟で信頼性の高いモデルや API の提供に繋がるだろう．

5.2 セキュリティの重要性に関する認識と今後の課題

表 5 のセキュリティ質問においては，参加者の大多数が若く経験年数が短かった一方で，少数の年配の開発者や，長期間にわたり業務で AI 開発に従事している開発者は，セキュリティ対策の重要性を認識し，具体的な知識を持っている傾向があった．すなわち，開発経験の差がセキュリティ認識に影響を与えていることが示唆される．より深い分析は今後の課題としたい．一般に，セキュリティリスク管理においては，情報流出リスクに対してデータの匿名化，暗号化，アクセス制御の強化が有効である．また，モデルの信頼性リスクには訓練データの選定と評価手法の見直しが重要であり，セキュリティ脆弱性には脆弱性診断とパッチ適用の継続的な実施が求められる．ファイル形式リスクに対しては，ファイルの検証とセキュアな保存方法が推奨される．さらに，人的要因に対してはセキュリティ教育の強化が必要であり，法的およびプライバシーリスクには各国の法規制に基づくコンプライアンスの遵守が不可欠である．定期的なリスク評価とアップデートがリスク認識を促進し，セキュリティ対策の強化に寄与することが期待される．

5.3 プライバシー保護の重要性に関する認識と今後の課題

調査の結果，多くの回答者が個人情報保護の重要性を認識し，特にデータ削除，逆学習（忘却処理），匿名化などの対策強化が必要であると認識していることが明らかになった．一般に，プライバシー保護を強化するためには，個人情報漏洩リスクを最小限に抑えるための標準的なプロセスの確立と匿名化技術の向上が求められる．また，データ転送や保存時の暗号化，セキュリティソフトの適切な利用，定期的なセキュリティレビューも重要である．さらに，データに関する問題が発生した場合の迅速な報告と対応体制の整備が必要である．教育プログラムの強化や，著作権や倫理に基づくガイドラインの整備も全体的なプライバシー保護の向上に寄与すると考えられる．

5.4 倫理の重要性に関する認識と今後の課題

多くの AI 開発者が倫理的問題を重要視している一方で，具体的な適用方法についての理解や実践はまだ不十分であることが明らかになった．このような状況に対し，AI 倫理に関するガイドラインやフレームワークの開発が必要であり，開発者に対する教育や支援，プロジェクトの初期段階からの倫理的議論の重要性が強調される．また，組織的な

© 2024 Information Processing Society of Japan

サポート体制の整備が不可欠である．これにより，AI システムの公平性や透明性に関する懸念を軽減し，倫理的な問題に対する予防策として機能することが期待される．

5.5 法令および規制に対する認識と今後の課題

AI 開発者は法令および規制に対する認識は高いものの，具体的な遵守方法に関する理解は不十分であることが明らかになった．また，調査の結果，地域ごとに異なる規制への対応が難しいという意見が報告された．そうした状況に対応するためには，法規制の最新情報を把握し，開発プロセスに組み込む体制やツールの導入が必要である．さらに，企業レベルでの法務チームとの連携強化や，規制に関する専門知識を持つ人材の育成も重要であり，包括的な教育プログラムとサポートの強化が求められる．

6. 結論

本研究では，AI 技術の進化と普及に伴い，セキュリティ，プライバシー，倫理，法令遵守といった重要な要素に対する AI 開発者の認識，理解度，実践している対策を明らかにするとともに，どのような課題が存在するかを探った．その結果，AI 開発者の多くがセキュリティリスクやプライバシーリスクを認識している一方で，実際にこれらの対策を実施している開発者は限られていることが明らかになった．また，AI システムの透明性や公平性の確保に対する取り組みが不十分であることも示され，倫理的な側面に対する認識の向上が求められている．さらに，法的規制に関しては，多くの開発者がデータセットやモデルのライセンスを確認しているものの，データの出所を明示している割合は依然として低い状況が確認された．

これらの結果から，AI 技術の持続可能な発展と社会的受容を実現するためには，開発者の認識をさらに深め，実際の開発プロセスにおいてセキュリティ，プライバシー，倫理，法令遵守を徹底するための教育や支援体制の強化が不可欠であることが示唆された．今後は，AI 開発者の知識と理解度を向上させるための包括的な教育プログラムの導入や，ガイドラインの制定，ならびに組織的なサポート体制の整備が求められる．また，セキュリティ，プライバシー，倫理，法令遵守の各要素が AI 開発において統合的に管理されることにより，AI 技術の安全かつ倫理的な運用が保証され，社会的信頼の向上に寄与することが期待される．

謝辞 本研究の実施にあたり，議論頂いた KDDI 総合研究所の披田野清良氏，統計数理研究所の村上隆夫氏に感謝いたします．

参考文献

[1] Karen L. Boyd. Datasheets for datasets help ml engineers notice and understand ethical issues in training data. *Proc. ACM Hum.-Comput. Interact.*, Vol. 5, No. CSCW2, October 2021.

[2] Yogesh K. Dwivedi, Laurie Hughes, Elvira Ismagilova, Gert Aarts, Crispin Coombs, Tom Crick, Yanqing Duan, Rohita Dwivedi, John Edwards, Aled Eirug, Vassilis Galanos, P. Vigneswara Ilavarasan, Marijn Janssen, Paul Jones, Arpan Kumar Kar, Hatice Kizgin, Bianca Kronemann, Banita Lal, Biagio Lucini, Rony Medaglia, Kenneth Le Meunier-FitzHugh, Leslie Caroline Le Meunier-FitzHugh, Santosh Misra, Emmanuel Mogaji, Sujeet Kumar Sharma, Jang Bahadur Singh, Vishnupriya Raghavan, Ramakrishnan Raman, Nripendra P. Rana, Spyridon Samothrakis, Jak Spencer, Kuttimani Tamilmani, Annie Tubadji, Paul Walton, and Michael D. Williams. Artificial Intelligence (AI) : Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, Vol. 57, p. 101994, 2021.

[3] Adhishree Kathikar, Aishwarya Nair, Ben Lazarine, Agrim Sachdeva, and Sagar Samtani. Assessing the vulnerabilities of the open-source artificial intelligence (ai) landscape: A large-scale analysis of the hugging face platform. In *2023 IEEE International Conference on Intelligence and Security Informatics (ISI)*, pp. 1–6, 2023.

[4] Hao-Ping (Hank) Lee, Lan Gao, Stephanie Yang, Jodi Forlizzi, and Sauvik Das. "I Don't Know If We're Doing Good. I Don't Know If We're Doing Bad"': Investigating How Practitioners Scope, Motivate, and Conduct Privacy Work When Developing AI Products. In *Proc. 33rd USENIX Security Symposium (USENIX Security 24)*, August 2024.

[5] NEC ソリューションイノベータ. 自動運転など自動車で活用される ai 技術の事例と今後の課題. `https://www.nec-solutioninnovators.co.jp/ss/mobility/column/07/index.html`, August 2024.

[6] Aastha Pant, Rashina Hoda, Simone V. Spiegler, Chakkrit Tantithamthavorn, and Burak Turhan. Ethics in the Age of AI: An Analysis of AI Practitioners' Awareness and Challenges. *ACM Trans. Softw. Eng. Methodol.*, Vol. 33, No. 3, March 2024.

[7] Federica Pepe, Vittoria Nardone, Antonio Mastropaolo, Gabriele Bavota, Gerardo Canfora, and Massimiliano Di Penta. How do hugging face models document datasets, bias, and licenses? an empirical study. In *Proceedings of the 32nd IEEE/ACM International Conference on Program Comprehension*, ICPC '24, pp. 370––381, 2024.

[8] Rowena Rodrigues. Legal and human rights issues of AI: Gaps, challenges and vulnerabilities. *Journal of Responsible Technology*, Vol. 4, pp. 1–12, 2020.

[9] Tanveer Ahmad and Dongdong Zhang and Chao Huang and Hongcai Zhang and Ningyi Dai and Yonghua Song and Huanxin Chen. Artificial intelligence in sustainable energy industry: Status quo, challenges and opportunities. *Journal of Cleaner Production*, Vol. 289, pp. 1–29, 2021.

[10] NHK NEWS WEB. Chatgpt 公開から 1 年 誤情報拡散などのリスク対応が課題. `https://www3.nhk.or.jp/news/html/20231130/k10014271951000.html#anchor-02`, November 2023.

[11] 国立研究開発法人産業技術総合研究所. 産総研マガジン/医療 ai とは？ `https://www.aist.go.jp/aist_j/magazine/20220525.html`, February 2022.

© 2024 Information Processing Society of Japan