「AI のセキュリティ確保のための技術的対策に係るガイドライン（案）」に関する意見（NACS）2026.1.26

総務省に「AI のセキュリティ確保のための技術的対策に係るガイドライン（案）」に関する意見を提出いたしました。消費者は AI 内部の処理を理解することが容易ではありません。ハルシネーションやデータ漏えい、なりすまし被害などのリスクを自力で回避することが困難であるため、AI の技術的セキュリティ対策は、事業者内部だけでなく事業外利用者（消費者）も影響を受けることから、意見を提出しました。

【意見内容要旨】
AI の安全対策は、影響を受ける一般の利用者（消費者）の安心も守ることを考えるべきだとして、開発者や提供する事業者には、最低限の安全対策を必ず実施し、将来的には義務化することを求めました。消費者は AI の仕組みを詳しく理解できず、誤った回答や情報漏えい、なりすましなどの危険を自分で見抜くことが難しいため、注意喚起や相談窓口、トラブル時の案内など、安心して使える仕組みが必要だと指摘しました。また、AI の暴走を防ぐ仕組みの強化や、やり取りの記録を残すこと、利用者自身が履歴を保存できる機能など、説明責任や救済につながる仕組みを標準設定として整えるべきとしました。さらに、AI の基盤となるシステムに問題が見つかった際の利用者への知らせ方や、開発者と提供者の間で安全に関する情報を共通の形式で共有する仕組みも必要とし、消費者が安心して AI を利用できる環境づくりを求めました。

「AI のセキュリティ確保に向けた技術的対策に係るガイドライン」本編（案）	
該当箇所	御意見
全体	【意見内容】 本ガイドラインの対象となる AI 開発者及び AI 提供者には、最低限の対策を必ず実施することを求めます。将来的には、義務化を検討すべきと考えます。 また、「どの規模・種類の事業者が、どの水準の対策を行うべきか」が不明確です。行政・医療・金融など、社会的影響の大きい重要サービスについては、高度な対策を必須とする旨を明確に示すことが望まれます。 【理由】 業務外利用者（消費者）にとって、AI 開発者及び AI 提供者がガイドラインに沿った対策を講じているかどうかは、「AI が安全で信頼できるか」「AI が悪用され、消費者が被害を受けるリスクがあるか」 を大きく左右します。 AI の内部で行われる処理は利用者からは理解できません。そこで生じる誤りやリスクを利用者が見抜くことはほぼ不可能に近い状況です。事業者の対策が不十分であれば、被害を受けるのは消費者も同じです。 AI の信頼性が確保されなければ、社会全体での普及・発展も難しくなると考えます。P12 で「対策例を実装した場合においても、AI の性質上、脅威を生じさせる要因等を完全に排除することは困難である点について留意が必要である。」としていることでも、対策を講じなければ脅威は拡大していくと考えます。

「対策の位置づけ」	【意見内容】 3.1 で示されている「対策を実装しても脅威を生じさせる要因等を完全に排除することは困難」という整理を踏まえ、業務外利用者（消費者）が「何が保証され、何が保証されないか」を誤解なく理解できるよう、最低限の注意喚起や問い合わせ窓口、インシデント時の案内等の整備について、本文で簡潔に言及してください。併せて、必要に応じて他の関連指針（AI 事業者ガイドライン等）への参照関係を明示してください。 【理由】 技術的対策は重層化しても未知の攻撃等を完全に排除できない以上、業務外利用者（消費者）側が過信しないための情報提供と、トラブル発生時の連絡・救済の導線が重要になります。
「対策の位置づけ」	【意見内容】 不正に操作された事業活動のＡＩ被害は、事業者のチャットボットなどを利用する事業外利用者（消費者）にも及ぶことが予想されます。事業外利用者への配慮も含めた対策を明確にしてください。AI システムのリスクが技術的観点から書かれていますが、事業外利用者である消費者はリスクを自ら回避することが困難です。消費者に分かりやすく、利用時に特に顕在化しやすいリスクを明示してください。事業外利用者（消費者）にも視点を合わせ、消費者の脆弱性を前提とした、ＡＩの信頼性を高める対策を求めます。 【理由】 事業外利用者の被害は自己責任とされてしまう懸念があります。対策として示されている技術の適用は、事業外利用者にとっても有効と思います。しかし、現状では消費者はＡＩを過信してハルシネーションに気付かない、入力データの漏えいに不安を感じる、ディープフェイク・なりすまし被害にあうなどしており、被害の回復も困難な状況です。
「AI 提供者における対策」	【意見内容】 提供者における対策は、ここでは、悪意のある外部ユーザー（攻撃者）に対する対策としてあげられています。ガードレールの役割は大きく、その機能強化をめざす対策は重要と考えます。これらの対策が、外部向けチャットボットなどを利用する事業外利用者（消費者）の安全のために強化すべき点でもあることに言及してください。被害が発生した場合の原因究明や、説明責任、被害救済のためにも、ログ保全が重要となります。これらの技術的対策が、「標準で有効」であり、「消費者に不利にならない初期設定」として推奨されることを求めます。 【理由】 プロンプトインジェクション攻撃により、個人情報の漏えい・詐欺誘導、なりすましの文章・スクリプト作成などのリスクが生じます。間接プロンプトインジェクションによりＡＩの回答が改変・偏向されても、消費者は気付きにくいものです。DoS 攻撃のリスクとしては、行政・医療・金融などのサービスに支障が生じて消費者が利用できず、生活上の安全が脅かされることにつながります。攻撃や不正が生じた場合のそのレベルや、消費者への影響を評価することも必要ではないでしょうか。

「AI 開発者・提供者に係るその他の基本的な対策等」（監査ログの保存によるトレーサビリティの確保）	【意見内容】 監査ログの保存について、サイバー攻撃の痕跡調査等の観点に加え、AI 利用に関する苦情・紛争が生じた際の事実確認および業務外利用者（消費者）への説明の基礎となることを、本文で補足してください。併せて、ログに個人情報等が含まれ得ることを踏まえ、用途・目的や提供条件等に応じて、保存期間、アクセス権限、改ざん防止等の管理方針を定める重要性も明確化してください。 【理由】 本ガイドラインが対象とする脅威は、入力や参照データを介して不正な出力や外部連携の誤動作等を招き得るため、原因（攻撃・不具合・利用者操作）を切り分ける際にログが客観的記録となり得ます。ログは事業者の説明責任の裏付けであると同時に、業務外利用者（消費者）側の相談・紛争解決における事実確認の基礎資料にもなります。
「AI 開発者・提供者に係るその他の基本的な対策等」（監査ログの保存によるトレーサビリティの確保）	【意見内容】 事業者側の監査ログ保存に加え、業務外利用者（消費者）自身が、自己の対話履歴を保存・ダウンロードできる機能（エクスポート機能）を、消費者保護とトレーサビリティ確保の観点から、本ガイドラインまたは AI 事業者ガイドライン等の関連指針において推奨事項として検討してください。 【理由】 トラブル時に事実確認が必要になる場面があります。業務外利用者（消費者）が客観的記録を保持できることは、迅速な相談・解決に資すると考えます。実装に当たっては、個人情報・機微情報の保護、不正利用防止、セキュリティリスクへの配慮が前提となります。
「AI 開発者・提供者に係るその他の基本的な対策等」（システム構成要素の信頼性確認、継続的見直し）	【意見内容】 AI 提供者が基盤モデル等の信頼性確認を行うだけでなく、採用した基盤モデルや外部サービス等に重大な脆弱性・インシデントが判明した場合に、影響が想定される業務外利用者（消費者）へ、必要な範囲で通知・注意喚起を行う手順（運用体制）についても、継続的見直しの運用として触れてください。 【理由】 業務外利用者（消費者）は、背後で利用される基盤モデル等の種類や既知のリスクを把握しにくく、通知がなければ機微情報の入力回避等の自衛策を取りにくいのが実情です。継続的見直しの実効性を高める観点からも、業務外利用者（消費者）への周知を含む運用が重要と考えます。

システムの構成要素のセキュリティに係る信頼性の確認に関して、AI 提供者は、基盤モデルの作成者が開示している情報等を踏まえ、セキュリティに係る信頼性を確認することが重要である。	【意見内容】 「開示している情報等を踏まえ」とありますが、あらかじめセキュリティ関連で開示すべき情報を定めておく必要はないでしょうか。 【理由】 学習データやアルゴリズムはブラックボックスになりがちですが、少なくともセキュリティ関連においては、AI 開発者と AI 提供者の間の受渡情報を化学物質の SDS（安全データシート）のようなもので定型化しておくことが有効ではないかと思います。

「AI のセキュリティ確保のための技術的対策に係るガイドライン」別添（付属資料）（案）	
該当箇所	御意見
「オーケストレータの権限管理」（実行の認可をユーザに都度求める／確認ダイアログの設定）	【意見内容】 外部システムと連携してアクションを実行する AI システムについて、最小権限の原則等の技術的対策に加えて、UI 上の「確認ダイアログ」の実装を、特に外部への接続や情報送信を伴う操作では、より積極的に推奨することを検討してください。なお、将来的に AI システムの高度化・自律化が進み、決済や契約確定等の不可逆的な操作が想定される場合には、そうした高リスク操作において確認ダイアログを原則実装すべき対策として位置づけることが重要と考えます。 【理由】 ガイドラインが示すように、外部参照や外部連携を介した間接的な攻撃等では、業務外利用者（消費者）の注意だけで回避するのが難しい場面があり得ます。高リスク操作において、技術的な権限制限に加えて、最後に人が認可を与える設計とすることは、被害拡大の抑止に資すると考えます。