Ministry of Infrastructure
and Water Management

AI 影響アセスメント

責任ある AI プロジェクトのためのツール

バージョン *2.0*、*2024* 年 *12* 月インフラ・水管理省（*I&W*）*CIO* 室（*CDIB*）、人間環境・交通監察官室（*ILT IDlab* および分析部門）、公共事業・水管理総局（*RWS Datalab*）の同僚との協力による。

ご質問、ご意見は *teamai@minienw.nl* まで。

目次

AI 影響アセスメントは、設計による責任ある AI の確保に役立つ 5

AI プロジェクトのすべてのフェーズで AIIA を活用する 6
この文書の使い方 8
誰が何をするのか？ 9

パート A：アセスメント 10

1 システムの目的と必要性 10

1.1 システムの目的 10
1.2 想定される解決策 10
1.3 組織内での役割 11
1.4 保守と管理 11

2 影響 12

2.1 基本的権利 12
2.2 持続可能性 12
2.3 その他の効果 13

3 AI システムを使うかどうかのアセスメント 14

パート B：AI システムの導入と活用 15

4 技術的堅牢性 15

4.1 バイアス 15
4.2 精度 16
4.3 信頼性 16
4.4 技術的実施 17
4.5 再現性 17
4.6 説明可能性 18

5 データガバナンス 19

5.1 データの品質と完全性 19
5.2 プライバシーと守秘義務 20

6 リスクマネジメント 21

6.1 リスク予防 21
6.2 代替手続き 21
6.3 情報セキュリティリスク 21

7 説明責任 22

7.1 ユーザーに対する透明性 22
7.2 関係者へのコミュニケーション 22
7.3 検証可能性 23
7.4 アーカイビング 24

用語集 25

附属書 1：リスクアセスメント 30

リスクの高い AI システムの定義（AI 法） 30

例外 32

附属書 2：リスクの高いシステム 33

リスクの高い AI システムを使用したい場合の質問.....33

リスクの高い AI システムの開発者（プロバイダ）への質問 35

附属書 3：生成的 AI に関する留意点 37

AI 影響アセスメントは、設計による責任ある AI の確保に役立つ

人工知能（AI）はチャンスをもたらすが、リスクも伴う。予期せぬ悪影響を防ぐためには、AI システムを展開する前に、その影響について明確にしておくことが重要である。これは、AI が提供する機会を最大限に活用できるようにするのに役立つ。AI の責任ある利用を可能にするため、I&W 省の ILT IDlab および分析部門、RWS データラボ、CIO 室は、AI 影響アセスメント（AIIA）を開発した。AIIA は、AI 展開の説明責任、品質、**再現性を**高めるための思考プロセスを支援するツールとして使用されることを意図している。AIIA は、データ収集、AI システム、アルゴリズムにおける障害に着目し、関連法規を考慮する。AIIA が完成すれば、**AI システムを**使用するか否かを決定する際のアセスメントに関して、透明性がもたらされる。

ここでは、AI が特定の文脈において適用されることを想定している。AI システム（またはその一部）を別の目的に使用する場合は、別の AIIA を実施する必要がある。この例としては、船舶用の画像認識モデルを他の車両にも展開することなどが考えられる。

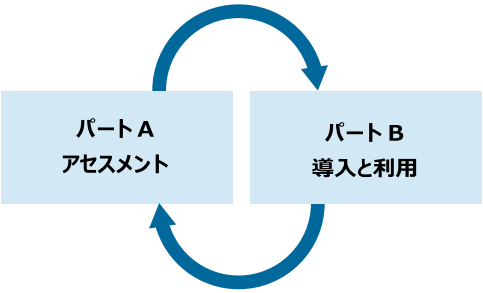
AI 法は、許容できないリスクから最小限のリスクまで、AI システムの一連のリスク領域を概説している。最終的に講じるべき対策は、リスクのレベルによって異なる。 このため、AI 法に概説されているリスクレベルを常に適用すべきである（附属書 1 参照）。例えば、AI システムが低リスクのアプリケーションから高リスクのアプリケーションに移行した場合、AI 法を遵守するために、より厳格な要件が必要となる。

学際的なグループで AIIA を完成させる

AIIA は、組織の学際的なグループによって完成される必要がある。各章を完成させるためには、異なる種類の知識が必要とされるからである。見出しの「誰が何をするのか」は、各章に関連すると思われる役割の概要を示している。データサイエンティストが答えなければならない質問もあれば、法律の専門家が答えなければならない質問もある。

AIIA の各部

AIIA は 2 つのパートに分かれている。パート A では、AI システムを使用する際に考慮すべき要素、すなわち、どのような目的で、どのような効果が期待されるのかについて検討する。この情報は、AI システムの潜在的な適用と必要な措置を評価するために使用される。 これによって、その適用の望ましさをめぐる倫理的議論が実証されることになる。パート B では、AI システムの設計、実装、使用について検討する。



両パートは別々に完成させるものであるが、両パート間には多くの共通点がある。パート A を終了する際、効果的なアセスメントを行うために、パート B の特定のトピックを把握する必要がある場合がある。例えば、アルゴリズムの種類は持続可能性に影響を与えるため、アセスメントに影響を与える。実施中の選択は、AI システムを使用するかどうかの決定に影響を与える。 例えば、パート A での実装に関する仮定が、パート B では正しくないことが判明し（例えば、利用可能なデータや異なる IT インフラの使用により）、その結果、パート A で行われたアセスメントが正しくなくなるような場合である。下表は、AIIA のさまざまな応用可能性を示している。

AI プロジェクトのすべてのフェーズで AIIA を活用する

AIIA は、AI システムの開発・調達のあらゆる段階で活用できる。AIIA は、AI プロジェクトの最初から使用すれば、最も有益である。システムが本番稼動する際には、AIIA を完了していることが必須である（パイロット版の場合も同様）。以下の表は、さまざまな可能性の概要を示している。AIIA は、AI 法に基づくリスクに関係なく、すべての AI システムに義務付けられているため、プレ AIIA は存在しない。

クイックスキャン AIIA	AIIA のクイック・スキャンを使い、AI のアイデアが実現可能か、望ましいかを調査する。その結果、AI システムを調達または開発することが良いアイデアかどうかが明らかになる。選択肢を調査する際には、パート A の青い質問を使用し、必要であればパート B の青い質問を使ってより詳細な分析を行う。
プロジェクト計画をまとめる	AI システムを使用するためのプロジェクト計画をまとめる場合、AIIA を完成させることが必須である。AIIA は、AI システムのアセスメントに関する透明性を提供し、選択した内容を説明するために使用する。また、議論のための効果的なツールであり、関連するすべての側面が考慮されていることを確認するためのものでもある。AI voor opdrachtgevers (AI for commissioning clients)と題されたガイドラインは、AI システムの開発、実施、管理のための実践的なヒントを提供している。
開発中	導入に際しての選択は、AI システムの影響に影響を与える（使用するデータやモデルの種類など）。パート A を使用して、AI システムの影響と使用の選択を判断する。パート B を使用して、実装において関連する側面が考慮されているかどうかをチェックする。
利用中の AI システム	AIIA は、AI システムが本番稼動する時点で義務づけられている。本番稼動後も、AI プロジェクトが要件を満たしているかどうかを評価する必要がある。適用分野が変更されていないか確認する。

影響のあるアルゴリズムのための AIIA

AIIA は AI システム、つまり学習要素を持つ機械ベースのシステム用に設計されている。

もっと単純な**アルゴリズムも**あり、それは基本的に、あらかじめ定義されたルールによる「レシピ」である。この種の**アルゴリズムにも** AIIA を使用できるが、必須ではない。この質問は、アルゴリズムの影響と、責任ある適用を保証する上で関連する実装の選択肢を評価するのに役立つ。

助けが必要か？

AIIA は、あなたが検討している AI プロジェクトや AI システムには適切なツールではないと感じるか？AIIA について他にコメントや質問があるか？AIIA の完成に疑問を感じているか？もしそうなら、I&W の AI チームに連絡を。

バージョン 2.0 の説明責任

AIIA の最初のバージョンは、ILT の ID ラボと分析部門、RWS のデータラボ、I&W の CIO オフィス（CDIB）によって開発された。AIIA は 2022 年 7 月 4 日、I&W 行政審議会によって承認された。このバージョン 2.0（2024 年）は、生成的 AI に関する情報を追加し、ユーザーの経験や AI 法の決定版に関する情報も考慮している。また、基本的権利およびアルゴリズム影響アセ

スメント（IAMA）とのリンクも強化された[1] 。その結果、AIIA は現在、AI をどのように適用できるかという問題だけでなく、AI を展開することが望ましいかどうかという問題にも取り組んでいる。

[1]これは、基本的権利のアセスメントにおいて追加的な支援が必要な場合を除き、AIIA と IAMA の両方を記入する必要がなくなったことを意味する。2.1 章を参照のこと。

この文書の使い方

AIIA を修了するためには、以下の要素が重要である：

- AIIA は、AI システムが本番稼動した時点で義務付けられる。
- AIIA がどの程度完成しているかは、プロジェクト・リーダーの専門知識に依存し、また AI システムの影響を反映する。特に指示がない限り、単純な「はい」「いいえ」では質問への回答としては不十分である。
- 青色の質問は必須である。必ず記入すること。
- 緑色の質問は追加情報を提供するためのものである。該当する場合は記入すること。
- 太字は、附属書に含まれる用語集で定義されているクリック可能な用語である。
- 記入用テンプレートも用意されている。

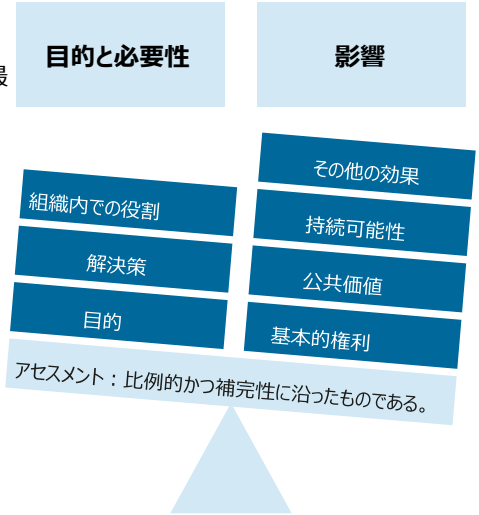
中央政府監査局(*Auditdienst Rijk*)およびオランダ監査裁判所(*Algemene Rekenkamer*)は、AI システムの正確性および安全性をチェックすることがある。AIIA が完全に完了したからといって、必ずしも AI システムが安全でセキュアであるとは限らない。AI 法（規則(EU)2024/1689）の遵守を確実にするため、AI システムが**ハイリスクで**ある場合は、附属書 1 に記入しなければならない。

誰が何をするのか？

AIIA を多職種からなるチームで完成させることは重要である。なぜなら、各セクションには特定の知識や専門性が必要だからである。下の表は、AIIA の特定の章におけるさまざまな役割の関わり方の提案である。もちろん、これらの役割分担が実際に必要なのか、あるいは下表で提案されている以上の役割分担が必要なのかは、プロジェクトの範囲や規模によって異なる。

	第 1 章 目的	第 2 章 影響	第 3 章 アセスメント	第 4 章 技術的堅牢性	第 5 章 データ ガバナンス	第 6 章 リスク マネジメント	第 7 章 説明責任
利害関係者：		X					
CIO：	X	X	X	X	X	X	X
CISO：				**X**	**X**	**X**	**X**
コミュニケーション・ コンサルタント：		X					X
データサイエンティスト	X	X	X	X	X	X	X
データマネージャーまたは ソースデータ所有者：					X		
ドメインの専門家：	**X**	**X**	**X**	**X**	**X**	**X**	
プライバシー プロフェッショナル：	X	X	X	X	X		
法律の専門家：		X	X		X		X
依頼主：	**X**	**X**	**X**	**X**	**X**	**X**	**X**
その他の プロジェクトチームメンバー							
プロジェクトリーダー：	**X**	**X**	**X**	**X**	**X**	**X**	**X**
倫理に関する 戦略的アドバイザー：	X	X	X				

パート A：アセスメント



AI システムの影響は、意図された目的に比例しているか？これは AIIA の最初の部分で重要な質問である。すべての質問に目を通し、目的、意図する解決策、期待される効果に焦点を当てる。AI 展開の比例性は特に重要である。目的を達成するための、より過激でない方法が他にあるだろうか？

異なる利害の衝突や、AI システムが基本的権利を侵害するなどの理由で、これを評価することが困難な場合、倫理的側面について構造化された対話を行うための様々なツールが利用できる。これには、IBDS データ対話、DEDA、*Moreel beraad*（道徳的相談）、その他多くの方法が含まれる。 I&W AI チームは、この対話をサポートするための社内ワークショップを開催している。

1 システムの目的と必要性

これらの質問は、**AI システムの**目的と、組織内での意図された役割に関するものである。

1.1 システムの目的

AI の適用」は、それ自体が目的ではない。AI システムは、仕事をより効率的に、あるいは効果的に行えるようにするなど、組織内のある目標を達成するために展開される。どのような問題を解決する必要があるのか？そのために、AI が役割を果たすプロセス全体を見渡す。これらの質問の答えは、AIIA の残りの部分を完成させる際に使用できる。

1	AI システムの**意図**する目的と意図する**結果を**簡潔に説明する（タイトル、一般的な説明、問題の定義、予想される期間、場所、対象グループ、ドメイン、運用プロセス）。
2	御社の AI システムは、AI 規制のどのリスクレベル（**許容できない、高い、最小限のリスク**）に該当するか。

3	AI システムは組織のどこで（どの工程で）使用されることを想定しているのか？

貴社のシステムに適用される AI 法のリスクレベルを特定するために、附属書 1 のアセスメントを行うことを推奨する。これにより、貴社のシステムの基礎となる法的義務がより明確になる。

もし、あなたのシステムが "**unacceptable** "カテゴリーに属する場合、AI システムは許可されない。その場合、残りの AIIA を完了する必要はない。

AI システムが「**高リスク**」カテゴリーに属する場合は、附属書 1 の残りの質問にも回答する必要がある。これにより、高リスク・システムが満たす必要のある追加義務の概要を知ることができる。

低リスクまたは最小リスクに分類されるその他のシステムについては、AIIA の質問で十分である。

1.2 想定される解決策

このパートでは、適用される AI 技術や使用されるデータなど、先に説明した問題に対する意図された解決策を見る。必要であれば、パート B の質問も参考にして、以下の質問に十分に答えよう。

1	目的とする **AI システムの**簡単な説明（技術、データ、アルゴリズムの種類）を記入すること。

2 なぜこのような AI（生成的 AI、線形回帰、ニューラルネットワークなど）が選ばれたのか？

3 どのような代替案が検討されたか（例：AI なし、より複雑でない AI、異なるタイプのアルゴリズム）？

1.3 組織内での役割

他の IT システムと同様、AI システムにも委託先があり、最終的な責任を負う者がいる。所有権は不可欠である。これらの質問では、システムの開発と使用におけるタスクのディビジョンを決定する。これらの役割は用語集で定義されている。これらの定義に基づいて解答しなさい。

1 AI システム構築におけるディビジョンを記述する（**開発者**、**委託先**、**プロジェクトリーダー**、IT **管理組織**、**最終責任**者など）。外部の者が開発を担当する場合、どのような契約上の取り決めがあるか。

2 AI システムの**ユーザーは**誰なのか、システムを扱う**エンド**ユーザーは誰なのか、AI システムによって影響を受ける関係者は誰なのか。

3 **AI システムの**開発において、どのような利害関係者、人々、および／またはグループと協議したか？

4 異なる背景や経験を持つ代表者たちから、どのようなフィードバックを集めたか？また、そのフィードバックはどのようにフォローアップされたか？

1.4 保守と管理

他の IT システムと同様、AI システムにも保守と管理が必要である。概念実証やパイロットの場合、AI システムをどこでどのように本番稼働させるかを調査するのも良いアイデアだ。こうすることで、事前に適切な選択を行うことができ、例えば、互換性のない IT インフラや AI 技術によって、管理できない AI システムになってしまうことを防ぐことができる。

1 AI システムの管理・保守のディビジョンを記述する（**開発者**、**委託先**、**プロジェクトリーダー**、**管理組織**、**最終責任**者など）。システムの開発に外部の当事者が責任を負う場合、どのような契約上の取り決めがあるか。

2 AI システムの存続期間中に導入または更新される可能性のある新しい法律や規制は、どのように考慮されるのか？

3 AI システムの管理に必要な専門知識は文書化されているか？

4 AI システムのコンテキストの変化はどのように考慮されるのか？

2 影響

このセクションの質問は、特定の文脈における AI システムの適用の影響を特定することを目的としている。ここでは、基本的権利と持続可能性が明示的に挙げられているが、広範な繁栄にプラスまたはマイナスの影響を与える可能性のある他の影響もある。これは、特定のターゲット・グループへの影響や広範な繁栄への影響など、最も広義な意味での影響を指す。

この章は決められた順序で構成されているので、順を追って見ていってほしい。例えば、公共的価値は人権にも関連しており、これらはすでに「基本的権利」の章でカバーされている。また、パート B では影響の多くをカバーしている。

2.1 基本的権利

AI システムの運用に関心を持つすべての人が適切に扱われなければならない。つまり、すべての**関係**者の（基本的）権利が守られなければならない。 質問に答える際には、基本的人権が適用され、これらはオランダ憲法と欧州人権条約に定められている。基本的人権とアルゴリズム・影響アセスメント（IAMA）の附属書には、基本的人権クラスターのリストが掲載されており、以下の質問に答える際に役立つ可能性がある。

AI システムは、基本的権利の軽微、中程度、重度の侵害をもたらす可能性がある。中程度または重度の違反については、特に慎重な検討が必要となり、追加的な措置も必要となる可能性がある。

AIIA のパート B には、個人データ保護、情報へのアクセス権、公正な手続きなど、多くの基本的権利を保護するための措置が含まれている。 また、AI システムの使用によって影響を受ける可能性のある他の基本的権利もあるかもしれない。

助けが必要か？IAMA の第 4 部では、基本的権利に関するステップ・バイ・ステップの計画を提供しており、その中には主題に関する追加情報も含まれている[2] 。

1	**AI システムを利用すること**で、市民の基本的権利にどのような影響を及ぼす可能性があるのか？

2	どのような法的根拠に基づいて AI システムを使用し、AI システムに基づいてどのような意思決定を行おうとしているのか。
3	どの憲法条項が適用されるのか？
4	**AI システムが**不適切に実施された場合、どの憲法条項が違反する可能性があるのか？

2.2 持続可能性

AI システムはその生涯を通じて、エネルギー、水、原材料の消費を通じてエコロジカル・フットプリントを生み出す。例えば、**AI システムを**稼働させるためには、継続的な水の供給が必要となる。**AI システムは**エネルギーも消費する。より具体的には、問題定義から見てエネルギー消費が比例的かどうか、またこのエネルギーがどのように生成されたかを問うことが重要である。

例えば、企業はエネルギー消費量について透明性を持っておらず、この分野の研究はまだ初期段階にある。影響に関する信頼できる数値がない限り、この質問は主に持続可能な選択を検討するための方法として使用すべきである。別の技術、インフラ、モデルを使用することで環境影響を削減できる場合もあり、これにはモデルの高速化という利点もある。

一般論として、「より単純な」モデルの方が、**大規模な言語モデルの**ような大規模なモデルよりも、エネルギー消費量が少ないと言える。さらに、入力と出力のサイズとタイプもエネルギー消費に大きな影響を与える（例えば、テキストは画像よりも少ない計算パワーで済む）。

[2] IAMA 第 4 章の質問 1、2、3 は第 2 章 1.1 の質問と対応し、IAMA の質問 4 は第 1 章 1.1 と対応し、質問 5、6、7 は第 3 章のアセスメントに使用できる。

一方、**AI** システムを使って環境的利益を得ることも可能である。この影響は、例えばシステムを運用するための環境コストと相殺されなければならない。

1	**AI システムの**導入（開発、設置、使用）が環境に与える影響はどのようなもので、それをどのように測定するのか。
2	AI システムが環境に与える（マイナスの）影響を最小限に抑えるために、どのような対策が取られてきたか？

2.3 その他の効果

このセクションでは、前述していない影響や結果について見ていく。AIIA のパート B には多くの影響が含まれているので、すべてが含まれていることを確認するために、これらを見ることが重要である。また、この章では、このプロジェクトにとって特に重要な問題に言及することで、十分な配慮がなされていることを確認することができる。

以下は、潜在的な影響のリストである。これらの点がどの程度関連するかは、プロジェクトの範囲によって異なる。

その他の効果も考慮する必要がある：

- 公共価値：社会として何を価値あるものと認識するかについて、変化する、文脈に関連した考え方。公共価値に関するさらなる説明と洞察は、内務・王国関係省（BZK）の倫理的に責任あるイノベーションのためのツールボックスにある。[3]
- 組織の使命とビジョン：例えば I&W の場合、これはより良く、クリーンで、安全で、持続可能で、利用しやすいオランダを実現するための活動である。
- 短期と長期
- 個人、組織、社会への影響
- AI システムを使うことのプラス面とマイナス面

その際、リスクだけでなく、AI システムを適用することによる機会やプラスの効果も考慮する必要がある。

1	AI システムは組織の使命にどのように貢献するのか？
2	上記の質問の他に、AI 制度がもたらす関連効果（ポジティブ、ネガティブ、リスク、特定のターゲットグループ、さまざまなレベル、幅広い繁栄）で考慮すべきものはあるか。

[3] https://www.digitaleoverheid.nl/overzicht-van-alle-onderwerpen/nieuwe-technologieen-data-en-ethiek/publiekewaarden/toolbox-voor-ethisch-verantwoorde-innovatie/publieke-waarden-centraal/

3 AI システムを使うかどうかのアセスメント

このステップでは、AI システムの目的、解決策、影響を見ることによって、AI システムを展開する意味があるかどうかを判断する。AI システムを適用することは妥当か？意図されたアプリケーションは実際に問題を解決するのか？あるいは、他の技術も適しているか？このアセスメントをどのように行ったかを明確に記述すること。

選択に明確な理由をプロバイダする。すべてを注意深く考慮するよう努める。関係者の異なるグループを考慮する：いかなるグループの人々も、不当に不利益を被ることがあってはならない。もちろん、基本的権利の侵害がある場合や、法律や規則で認められていない場合など、影響が非常に大きく、さらなるアセスメントが必要ない場合もある。

パート B には、AI システムの責任ある適用を確保するための措置が含まれる。プロセス内などにおいて、追加的または具体的な措置が必要となる可能性がある。これらについてここに記載すべきである。

1. その影響は意図された目標に比例しているか、また、これらの目標を達成するためにより急進的でない方法は他にあるか。言い換えれば、明示された目標を達成するために制度を展開することは、**比例的**であり、**補完性に沿って**いるか。
2. 責任を持ってシステムを使用するために、（プロセスの一部など）追加的な対策はあるか？

パート B：AI システムの導入と活用

AIIA のパート B は、AI システムの設計、実装、使用について検討する。より詳細な内容を含み、多くの場合、AI 技術に関するより深い知識を必要とする。この章を完成させ、使用する際には、データ・アナリストなどの専門家を必ず参加させること。

4 技術的堅牢性

AI システムが意図したとおりに機能するかどうかは、その技術的な**堅牢性に**よる。

4.1 バイアス

バイアスとは、物事や人、集団について決めつけることを意味する。これには 2 つの側面がある。一方では、新しい状況でのデータの使用について、事前に結論を出すことが必要である。一般化するとき、私たちは常に仮定をする。一方では、人権や規則に反するような不正や望ましくないバイアスによって引き起こされる不当な歪曲を避けることも重要である。

バイアスはシステムのどこにでも発生する可能性がある。例えば、**入力のバイアス**、**モデルのバイアス**、**出力のバイアスなど**である。AI を設計・使用する際には、**データの**バイアスと**設計のバイアスを**考慮することが重要である。これらのバイアスは、多くの場合、データ中の社会経済的な仮定や、開発者によって作られたものによって引き起こされる。モデルはこれらの仮定を悪化させる可能性がある。バイアスに気づき、それを修正しなければ、AI システムは**関係者全員にとって**効果的に機能しないことになりかねない。

この問題の核心は、意識と誠実さにある。バイアスを完全に排除して仕事をすることは不可能である。バイアスは何十年も続き、認識されるべきであったとしても認識されないままであることが多い。例えば、言語モデル（ChatGPT、Google Bing、Microsoft Co-pilot）を見てみよう。これらはおそらく、インターネット上の情報と、この情報が長年にわたって含んできたすべての仮定を使って訓練されたものだろう。したがって、バイアスのない AI に焦点を当てるのではなく、潜在的な識別的 AI を可能な限り意識することを目指すべきである。

バイアスは、人間の**多様性**、**平等**、**公平**性と密接に関係しているが、思い込みは、自然や生活環境など、人間以外の側面にも関係する可能性があることを認識しておくことが重要である。バイアスの緩和をどのように目指すかを検討する際には、バイアスがもたらしうる**ネガティブな影響**、**ポジティブな影響なし**、**ポジティブな影響を**区別することも関連性があるかもしれない。

1 **入力のバイアス**、**モデルのバイアス**、**AI システムの出力のバイアスなど**、潜在的に望ましくない**バイアスは**どのように考慮されるのか？

入力（データ）のバイアス

2 入力（データ）は、AI システムの意図された目的（1.1 の質問 1）を考慮した上で、適切かつ代表的なものであるか。

3 ランダムサンプリングにおいて、必要であれば下位集団を防御したか？

4 入力変数の選択は、関係者との間で十分に協議され、調整されているか？

モデルのバイアス

5 AI システムにおいて、不公正または不当な**バイアスが**生じたり、悪化したりするのを防ぐために、どのような措置が取られているか？

6 AI システムは、想定される**エンドユーザー**（言い換えれば、年齢、性別、能力などの特性に関係なく）が使用できるか。

出力（データ）のバイアス

7 AI システムの否定的な影響によって社会内の集団が不釣り合いな影響を受けないようにするための停止メカニズム、監督メカニズム、監視メカニズムがあるか？特に ILT について：ここでは、*ondertoezichtstaanden*（監督される当事者（OTS））とそれ以外の社会とを区別する必要がある。

4.2 精度

正確な **AI システムは**、効果的なパフォーマンスを発揮し、そのアセスメントにおいて効果的である。AI システムのパフォーマンスを継続的に測定することが重要である（開発段階と生産段階の両方で）。使用する学習データの質も重要である。どのような AI システムも進行中の作業であり、AI システムを定期的にテストし、再トレーニングする必要がある。システムが誤ったアセスメントを行う可能性を定量化する方法があることが望ましい。

トレーニング）データとシステムの両方について**受け入れ規準を**作成することで、システムの**精度を**事前に決定することができる。規準の例としては、最小限のデータ量や、測定システムの特定のしきい値などがある。精度スコア、精度スコア、リコールまたは F1 スコアなど、**モデルの**品質を定量化するために利用可能な測定システム（データサイエンティストはしばしば「パフォーマンス・メトリクス」と呼ぶ）には、数多くの種類がある。

測定システムと規準が、AI システム（[4]）のデータと意図された目的に適 切に適合していることを確認することが重要である。特に、AI システムを使用する際には、時間の経過とともに新たなリスクが発生する可能性があるため、リスク分析（「リスクマネジメント」を参照）から得られた知見と調整する必要がある。また、システムの品質を継続的に監視することも重要である。再トレーニングやさらなる開発の際には、受入規準や測定システムの選択を再評価する必要がある。

1 システムの継続的な**精度は**どのように測定され、守られるのか？

2 **モデルの入力（データ）**と**出力（データ）**の品質を測定するために、どのような**受け入れ規準が**設定されているか？

3 **受け入れ規準は**データと AI システムの目的に適しているか？

4 **出力（データ）は**どのように無作為に定期的にチェックされ、正しさを継続的に監視されるのか？

5 規準に対する出力（データ）の逸脱をどのように分析し、タイムリーに修正するのか？

6 別の**モデルを使用した場合**、結果はどうなるだろうか？

4.3 信頼性

信頼できる AI システムは、同じようなケースで一貫した結果を出す。信頼性に関して言えば、重要な問題は、同じモデルと同じ**入力（データ）**、同じ設定、同じ**パラメーターを用いて**、個々の**出力（データ）を**再現できるかどうかである。また、システムが、新しい状況においてモデルがどの程度機能するかを示す信頼性の高い指標を提供することも重要である。

1. **AI システムは信頼**できるか？

2 **AI システムの**パフォーマンスに影響を与える最も重要な要因は何か？

3 （サブ）データセットの一部はモデルの学習プロセスから除外され、信頼性を判断するためだけに使用されるのか、あるいはモデルの信頼性はクロスバリデーションによって計算されるのか。

[4]選択された測定システムは、品質測定に使用されるモデルとデータに適していなければならない。例えば、あるテキストの 100 単語中 5 単語をラベル付けしなければならないシステムを考えてみよう。システムがこのテキストで 0 語をラベル付けした場合、モデルの精度は 95%となる。正確さに基づいてシステムの品質を決定する場合、このモデルは非常によく機能しているように見えるが、実際にはリコールは 0 であり、実際にはまったく機能していない。この理由から、精度はこのモデルの性能を決定するのに適していない。

4	.ハイパー・パラメーター・チューニングはどのように立証され、アセスメントされたのか？

4.4 技術的実施

技術的実装では、AI システムが組織の ICT ランドスケープにどのように技術的に統合されるかを説明する。AI システムの具体的なハードウェアとソフトウェアの要件は、システムの展開と管理において考慮できるように文書化される。

さらに、システム・アーキテクチャを見れば、さまざまなソフトウエア・コンポーネントがどのように相互に関連しているかがわかる。慎重に検討されたアーキテクチャは、技術的ソリューションの構築に伴う運用リスクを軽減し、運用要件と技術要件の架け橋となる。多くの場合、、すでに既存の文書が存在するので、これを参照するようにすべきである。 また、インフラストラクチャー・水管理エンタープライズ・アーキテクチャ（IWEA）にも目を通すべきである。

1	AI システムは技術的にどのように実装されたのか？

2	AI システムが既存の技術・システムインフラにどのように適合するかを検討し、その展開のために適切な措置が取られているか（該当する場合）。
3	システム・アーキテクチャ（ソフトウェア・コンポーネントの相互関係）を記述する。
4	特定のハードウェアとソフトウェアの要件は文書化されているか。
5	アプリケーションが外部でホストされている場合、どのような状況でこのようなことが起きているのか？
6	AI システムとそのコンポーネントへのアクセスはどのように設定されているか（一般的な IT 管理対策を考える）。
7	AI システムは他のハードウェアやソフトウェアとどのように相互作用できるか（該当する場合）。
8	ロギングとモニタリングはどのように設定されているか？

4.5 再現性

再現性とは、訓練、妥当性確認、テストのことである。再現性は、使用するデータの登録、モデルの開発、データの変化の記録、同じ**入力（データ）が**一貫した結果を生み出すかどうか、**出力（データ**）に影響を与える可能性のある特定の状況や条件があるかどうかといった問題に関係する。

再現性は**トレーサビリティと**密接な関係がある。トレーサビリティの主な目的は、データセットとプロセスが適切に文書化されていることを保証することである。データのバージョン管理、モデル、モデルのトレーニングは、その重要な側面である。

1.	**AI システムは再現**可能か？それを測定するプロセスは確立されているか？

2	得られた**出力（データ）は**、現在または将来にわたって再構築できるか（すなわち、**モデル**、データセット、条件の旧バージョンは、バージョン管理によって保存されているか）。
3	与えられたパラメータと固定**シードに基づいて**モデルを再構築することは可能か？
4	**AI システムの**大まかなアウトラインは、ドキュメントを使って再現できるか？
5	システムの存続期間中、変更はどのように文書化されるのか？

4.6 説明可能性

技術的**説明可能性**とは、技術的プロセスとそれに関連する人間の意思決定の両方を理解する能力に関するものである。また、どのような異なる設計上の選択がなされたのか、**AI システムを**使用する根拠は何なのかを明確にする必要がある。ユーザーやその他の**関係**者に対する説明可能性、透明性、コミュニケーションの詳細については、「アカウンタビリティ」も参照のこと。

1. **AI システムは、開発**者にとって十分に**説明可能**で、解釈可能なものだろうか？

1. AI システムの開発において、例えばユーザーに対するモデルの説明可能性はどのように考慮されたのか？
2. AI システムが説明可能であることを保証するために、どのような技術が使われ、なぜその技術が選ばれたのか？

5 データガバナンス

データガバナンスとは、アクセス、所有権、使いやすさ、完全性、セキュリティに関するデータに関する手続きのことである。また、使用されるデータの品質も含まれる。

データ・ガバナンスにはプライバシーも含まれる。プライバシーは、AI によって損なわれる可能性のある基本的人権のひとつである。したがって、少なくとも一般データ保護規則（**GDPR**）と I&W のプライバシー・ポリシーに従って、効果的なデータ・ガバナンスと個人データの保護を確保することが重要である。

5.1 データの品質と完全性

データの品質は、**AI システムの**効果的な運用に不可欠である。ガベージ・イン＝ガベージ・アウト」という言葉が存在するのには、それなりの理由がある。さらに、収集されたデータには、社会的に構築された**バイアス**、不正確さ、エラー、間違いなどが含まれている可能性がある（「バイアス」も参照）。これらのデータをさらに利用する前に、これに対処する必要がある。

そのためには、FAIR 原則（Findable、Accessible、Interoperable、Reusable）と、密接に関連する FACT 原則（Fair、Accurate、Confidential、Transparent）をインスピレーションとして使うべきである。これらの原則は GDPR の中で使用されている。データセットと作業手順は、トレーニング、テスト、ロールアウト段階、運用段階といったあらゆる段階でテストされ、文書化される必要がある。これは、社内で構築されたものではなく、買収された AI システムにも適用される。

個人データを使用する場合、データの質は特に重要である。GDPR によれば、個人データが正確であり、必要に応じて更新されることが不可欠である。[5]また、データは分析の目的に必要なものでなければならない。したがって、分析の目的、データの扱い方、データセットや結果の誤りについても記述する必要がある。[6]

法務省・セキュリティ省の「アルゴリズム適用のためのガイドライン」（'*Richtlijnen voor het toepassen van algoritmen*'）[7]は、データ品質を非常に実務的なレベルで明示している。

1. アルゴリズムの入力として、どのトレーニングデータを使用するのか。
2. データの品質はどのように守られるのか？

一般

3 使用されるデータは **AI システムに**必要か？

4 意図しないデータの重複をどのように防いでいるか？

5 状況に応じてトレーニングデータとテストデータを更新することは可能か？AI システムの再教育、一時停止、さらなる開発をいつ決定するのか[8]

6 データは**モデルの**前提条件を満たしているか？

7 AI システムで使用される**インプット（データ）**はどのように収集され、照合されたのか。

8 データはどのようにラベル付けされるのか？

[5]GDPR 第 5 条 1 項(d)。GDPR の第 3 節「データの修正と消去」も参照のこと。

[6]データ分析におけるアルゴリズム使用と公開の基準, JNV, p.20.

[7]https://open.overheid.nl/documenten/ronl-1411e45f-b822-49fa-9895-2d76e663787b/pdf

|AI 影響アセスメント

[8] 第 14.4 条

9 インプット（データ）の質に影響を与える要因は何か（収集方法や保管方法などの制限を考える）？また、それに対して何ができるか？

10 入力（データ）は、トレーニング、テスト、評価の間に起こる変化についてアセスメントされたか？また、長期間にわたってアルゴリズムを使用する間に生じる変化についても評価したか？

出力（データ）

11 出力（データ）が新たな入力として使用される場合、出力（データ）はどのように保存され、正確性と完全性がチェックされるのか？

12 アウトプット（データ）をタイムリーに利用できるようにするにはどうするのか？

5.2 プライバシーと守秘義務

AI システムを設計する際には、プライバシー法を考慮しなければならない。結局のところ、これを後の段階で修復するよりも、事前に適切に行う方が容易である。もちろん、プライバシーは **AI システムの**ライフサイクル全体を通じて保護されなければならない。

個人データを処理する際には、**データ保護影響アセスメント**（DPIA）のプレスキャンを行い、DPIA の全過程を完了する必要があるかどうかを判断しなければならない。

個人データのほかにも、公表してはならない機密情報が使用されることがある。例えば、機密情報や企業秘密などの機密情報の使用がこれに該当する。これらのデータも保護されなければならない。**AI 法は GDPR を**補足するもので、AI システムにおける（個人）データの使用に関する追加規則を含んでいる。機密データは十分に保護され、セキュリティが確保されなければならない（リスクマネジメントを参照）。

1. 個人データや機密データに対して、どのようなアプローチを採用しているか？

個人データについて

2 **AI システムは**個人データを扱うか（GDPR は適用されるか）。該当する場合は、以下の質問にも回答すること。該当しない場合は、「機密データについて」に進む。

3 パーソナルデータの処理は比例的か、補完性に沿っているか（第 3 章のアセスメントを根拠とする）。

4 AI システムのアウトプットは、直接的または間接的に個人を追跡できるか（GDPR は適用されるか）。

5 チーフ・プライバシー・オフィサー、情報セキュリティ・オフィサー、チーフ・インフォメーション・オフィサー、プライバシー・オフィサーなどの職員が関与したか。

6 パーソナルデータの処理の質と必要性はどれくらいの頻度で評価されているか。

機密データ（個人データではないもの）について

7 機密データが使用または保存されるか。

8 この情報のセキュリティはどのように守られるのか？

6 リスクマネジメント

リスクに目を光らせることが重要だ。予期せぬリスクによって、**AI システムが**信頼できない結果を出す可能性がある。その結果、AI システムのパフォーマンスが低下したり、**ハッキング攻撃を受けたりして**損害を被る可能性がある。

6.1 リスク予防

AI システムの開発と**実用**化には危険が伴うが、この AIIA は可能な限りそれに対処することを目指している。しかし、それでも不測の事態は起こりうる。こうした潜在的な危険にどのように対処するかを決めることが重要である。リスクをマネジメントする仕組みがあり、その仕組みがテストされていなければならない。これには、データ・ポイズニングの防止、管理措置の範囲、結果の保管場所のセキュリティなどが含まれる。リスクは AI システム導入後にも発生する可能性がある。少なくとも 3 年に 1 回、あるいは大きな変更があった場合には、定期的にリスクマネジメントをチェックすること。

リスクの高い AI システムについては、リスクマネジメントシステムが必須である。これについては、附属書 1 に追加質問がある。その他のシステムについては、単発のリスク分析で十分である。

1. 適切で的を射たリスクマネジメント対策について、システムはどのようにテストされたか？

6.2 代替手続き

AI システムに問題が発生した場合の計画を立てておくことが望ましい。つまり、代替手段を用意しておくことだ。例えば、機械学習モデルから、より限定的なルールベースの**モデルに**戻すオプションなどが考えられる。さらに一歩進んで、手動でプロセスを実行することもできる。もちろん、システムが一時的に利用できなくなることが許容されるのであれば、まったく計画を立てないことも選択肢となる。

AI システムの使用は、人間の能力を低下させる可能性がある。電卓が私たちの暗算能力に及ぼす影響はその一例である。このため、どのような代替手段も、いきなりこの技能に頼ってはならない。

1. **AI システムの**運用に問題が発生した場合の計画はどうなるのか？

2. システムが故障した場合、どのような影響があるのだろうか？
3. 上記の電卓の例を見てほしい。**AI システムが**実用化された場合、どのような同等の効果が起こりうるか、またそれは望ましいか？

6.3 情報セキュリティリスク

可能な限り、AI システムは**設計段階で**セキュリティを考慮し、安全な**設計で**構築されなければならない。こうすることで、モデルの操作、不正アクセス、**ハッキング攻撃などの**情報セキュリティリスクを管理することができる。組織のリスクマネジメントプロセスを用いて、予見可能なリスクのリストを作成する。これには、CIA の三要素、情報分類レベル、政府情報セキュリ ティベースライン（BIO）対策の実施、セキュリティテスト、BIO のセキュリ ティレベルで十分でない場合は、追加の（技術的な）リスク分析の実施などが含 まれる。

そのためには、適切に設定された権限と、強固な変更プロセスが不可欠である。 さらに、エラーや不正を検知し、技術的に対応できるかどうかをチェックすることも重要である。より実践的な指針は、会計検査院のアルゴリズム・アセスメント枠組み（*Toetsingskader Algoritmes van de* Rekenkamer）[9] や ADR アルゴリズム研究枠組み（*Onderzoekskader*

[9]https://www.rekenkamer.nl/onderwerpen/algoritmes/toetsingskader

algoritmes ADR）[10] にある。AIVD には、AI システムの安全でセキュアな開発に関するガイドがある。[11]さらに、Open Worldwide Application Security Project（OWASP）には、AI の安全でセキュアなアプリケーションに関する多くのリソースがある。

1. 情報セキュリティリスクはどのように識別され、許容レベルまで低減され、（技術的観点から）テストされるか？

2 権限のないサードパーティが AI システムの脆弱性を利用するのをどのように防ぐのか？

3 サードパーティが AI システムのソースコード、データ、結果に不正アクセスした場合、どのような影響があるか？

4 人間の判断の代わりに AI システムが使われているという事実を利用することは可能か？

5 誰が、どれくらいの時間、AI システムを利用したかを記録するシステムはどうなっているのか？

6 標準的な I&W セキュリティ対策に加え、AI システムの安全性を確保するための追加対策が取られているか？

7 説明責任

オランダ中央政府は、組織内、下院、そしてより広い社会に対して、その行動に対する説明責任を果たさなければならない。技術の応用はますます進んでいるが、AI の使用には懸念もある。AI システムの使用と結果に関して説明責任を果たすためには、そのためのプロセスを設定することが不可欠である。

7.1 ユーザーに対する透明性

開発者に求められる説明可能性に加えて、（「技術的堅牢性」を参照されたい）エンドユーザーには、AI システムの操作に関する洞察が与えられなければならない。これは特に、仕事のプロセスの一部として AI を使用する従業員に適用される。従業員は必ずしも AI システムを完全に理解する必要はない。しかし、AI システムの動作とその限界は、大枠において明確でなければならない。

AI の使用について決定を下す必要がある場合、その運用と限界を理解するだけでなく、その決定に対してかなりの程度の影響力を持つことが不可欠である。

1. AI システムの操作や限界について、どのような方法でエンドユーザーに洞察を提供しているのか？また、それらが持続する限り、十分な注意が払われているのか？

2. AI システムのインプットに基づく意思決定において、人間はどのような役割を果たすのか（「ループの中の人間」）、また、どのようにしてこの役割を果たせるようにするのか。

3. どのようにしてシステムを監視し、誰もが理解できるようにするのか（人的監視）。

7.2 関係者へのコミュニケーション

このセクションでは、**エンドユーザーとの** 2 種類のコミュニケーションについて述べる。第一に、エンドユーザーは、（例えば人間ではなく）**AI システムの**結果を扱っていることを認識しなければならない。第二に、エンドユーザーは、**アルゴリズムが**どのように AI システムの結果を決定するかを常に知る権利がある。これはまた、システムの目的と限界が、明確、公正かつ透明性をもってコミュニケーションされなければならないことを意味する。技術的プロセスとそれに関連する人間の判断の両方が理解可能で、検索可能で、必要であれば修正可能である必要がある。これは例えば、AI システムに関する実質的な知識を持つ窓口を任命することで達成できる。

[10]https://www.rijksoverheid.nl/documenten/rapporten/2023/07/11/onderzoekskader-algoritmes-adr-2023

[11]https://www.aivd.nl/documenten/publicaties/2023/02/15/ai-systemen-ontwikkel-ze-veilig

AI は自己学習する性質があるため、システムの動作を完全にトレースバックできるとは限らない。しかし、エンドユーザーに適切な説明を提供することは常に可能でなければならない。

加えて、市民が AI システムに関する情報を取得したり、GDPR に従って権利を行使したりすることが可能でなければならない。AI システムの結果に異議を唱えることも可能でなければならない。これはまた、データと、データが利用可能になる条件が保存されなければならないことを意味する（「アーカイブ」を参照）。

1. AI システムについて、さまざまな関係者グループに対して、どの程度、どのような形で透明性を確保しているのか。
2. エンドユーザーがシステム（データ、技術、対象グループなど）についてコメントできるような仕組みが構築されているか？また、それらはいつ、どのように妥当性確認（分析、フォローアップ）されているか？
3. AI 法に基づき、システムをアルゴリズム登録や（リスクの高いアプリケーションの場合）EU データベースに登録する必要があるか？

4 AI システムの**エンド**ユーザーや関係者は、結果が AI システムによって生成されたものであること、またそれが彼らにとってどのような意味を持つかを知らされているか。

5 マニュアルは作成されているか？

6 AI システムを使用することによる**エンドユーザーの**混乱、嗜好、認知疲労のリスクなど、潜在的な（心理的）副作用は何か？

7 **AI システムの**さまざまな側面について、さまざまな**関係**者グループ（市民、同僚、管理者など）はどのような方法で理解を深めているのか。これには、データの使用、モデル、結果などの分野が含まれる。

8 **エンドユーザーに対する**具体的な説明可能性を実現するために、どのような対策を講じてきたか？

9 システムは、**展開者が**システムの出力（データ）を解釈し、適切に利用できるよう、十分に**透明性が確保されて**いるか 20。

10 必要な場合、エンドユーザーにトレーニングを提供するための措置が取られているか。

11 関係者やエンドユーザーからのコメントを社内でどのように適切に処理しているか？

12 当事者が異議申し立て 21 を行いたい場合、あるいは AI システム 22 について苦情を申し立てたい場合、どのような手順を踏めばよいのか明確になっているか。異議申し立ての場合も同様である。

7.3 検証可能性

検証可能性とは、データ、**モデル**、結果を評価する方法を指す。このプロセスは、内部監査または外部監査によって行うことができる。よりリスクの高い分野で AI システムを使用する場合は、より厳しい要件が適用される。

ソース、システム、結果に対する洞察が不可欠である。これは**システムの所有者の責任で**ある。

1 **AI システムは**誰がどのように検証するのか？

2 AI システムに関する説明責任はどのような形でプロバイダされるのか？

3 AI システムの独立監査は誰が行うのか？また、どのような方法で？

7.4 アーカイビング

アーカイブとは、将来再利用できるように情報を保管することである。このような再利用には、モデルの再構築（「再現性」参照）、新しいスタッフにシステムの仕組みを説明すること（「説明可能性」参照）、**関係**者に説明責任を果たすこと（「説明責任」参照）などが含まれる。アーカイブは、法規制の遵守を確保する上でも重要である。例えば、AI システムにおけるログのリスクの高い用途には最低保存期間が適用され、これらは I&W の**選択リストにも**記載されている。

入力（データ）

入力（データ）はどのように保存されるのか？

インプット（データ）の保存期間は？

モデル

モデルの保存方法は？

バージョン管理はどのように行われているのか？

出力（データ）

アウトプット（データ）の保存期間は？

用語集

本文書では、文献によって定義が異なる様々な用語を使用している。以下のリストは、本文書で使用されている用語を明確に定義したものである。

規準	**AI システムが**満たさなければならない、意図された目的と合意されたデータに基づく条件。これらの条件は、データ量、**出力（データ）**の精度指標、または独立した出力検証メカニズムに関するものである。可能であれば、受入規準は定量化できるようにし、適切な測定システムによって監視できるようにしなければならない。優れた受入規準は、AI システムのすべての側面を効果的に監視できるよう、SMART で十分に区別されたものである。
精度	非常に正確、または細心の注意を払う：正確で緻密なシステムを指す。
正確なアセスメント	式で表される：TP+TN/(TP+TN+FP+FN)で表される。TP=真陽性、TN=真陰性、FP=偽陽性、FN=偽陰性。偽の結果に対する真の結果の数が多ければ多いほど、精度は高くなる。
AI 法	AI システムの開発と使用に関する規則を定めた欧州の法律。
AI システム	さまざまなレベルの自律性で動作するように設計され、展開後に適応性を示す可能性がある機械ベースのシステムであり、明示的または暗黙的な目的のために、物理的または仮想的な環境に影響を与えることができる予測、コンテンツ、推奨、または決定などの出力を生成する方法を、受け取った入力から推測する。
アルゴリズム	問題を解決したり質問に答えたりするために、コンピューターが自動的に計算を行う際に従う規則や命令のセット。
適用地域	AI 法制において、AI システムが使用される文脈を指す用語。例としてはインフラが挙げられる。
人工知能	AI に決まった定義はない。ここでは、オランダ会計検査院が用いた AI の説明を適用する：「外部データを正しく解釈し、そのようなデータから学習し、柔軟な適応を通じて特定の目標やタスクを達成するためにこれらの学習を使用する能力」である。 この文書では使用していないが、欧州委員会の定義にも注目したい：AI とは、環境を分析し、特定の目標を達成するために、ある程度の自律性をもって行動を起こすことによって、知的な振る舞いを示すシステムを指す。
バイアス	偏見。物事や人、集団について、多くの場合実際の測定に基づかない思い込みをすること。
入力のバイアス	データの質、一貫性、完全性は、偏りのない分析のための重要な前提条件である。
出力のバイアス	**アウトプット（データ）の**使われ方は、人々の生活に影響を与える可能性がある。不当な相関関係が因果関係をもたらさないようにすることが重要である。
モデルのバイアス	**モデルは**どの程度正しいのか、データの代表者性における既知の欠陥をどの程度補正しているのか。これは例えば、**AI** システムが何を学習し、何が望ましくない学習効果であると考えられているかに関係するかもしれない。

バイオ	政府情報セキュリティ・ベースライン（*Baseline Informatiebeeiliging* Overheid）[12] は、政府の階層における情報セキュリティの基本基準の枠組みである。
バイ・デザイン	設計プロセスにおいて、AI、プライバシー、セキュリティの関連法規を考慮すること。例として、AI、プライバシー、セキュリティ・バイ・デザインを挙げることができる。
CIO	最高情報責任者
CISO	最高情報セキュリティ責任者
委託先	請負業者に委託する個人または組織部門。 この担当者は、（プロジェクトリーダーとともに）AIIA を完成させる最終責任者でもある。
汚職	システム内のエラーを悪用したり、中立的なシステム特性を悪用すること。25 **意図しない腐敗とは**異なる。
データのバイアス	これは、母集団全体を代表するものではない無作為サンプルを指す。
データパイプライン	データがフィールドからモデルへどのように移動するか、データが通過するプロセスである。
データ対象者	制度の利用または結果に利害関係を有する（または有すると考えられる）自然人または組織。利害関係者」という言葉は、オランダ行政法で使用されている「利害関係者」の定義よりも広いため、意図的に使用しないことにした。その例としては、市民、監督下にある者、またエンドユーザー自身が含まれる。
デザインのバイアス	ハードウェアやソフトウェアといったコンピュータツールの制限を含む、技術設計上の問題。
開発者	**AI システムを**設計、開発、および／または訓練する組織または個人。
多様性	これは、分析においてさまざまなタイプの「被験者」を認識することを意味する。これによって、**AI システムの**開発から不当に排除され、その結果、システムが彼らのニーズに応えられなくなることを防ごうとしている。
ドメインの専門家	**AI システムが**構築されようとしている問題分野に非常に詳しい人物。
DPIA	データ保護影響アセスメント。このツールは、データ処理に伴うプライバシー・リスクを評価するために使用される。
エンドユーザー	エンドユーザーとは、**AI システムを**組織内で実際に使用する人のことである。これは自然人を指す。誰が操作するのか？組織内で AI システムから情報を収集するのは誰か？例えば、検査官や道路交通管制官などである。
事業体	組織の部署内での役職。
平等	これは、同じ種類の被写体が同等の扱いを受けるという考え方を指す。
説明できる	入力変数がアルゴリズムの出力にどのように寄与するかを説明しなければならない。
公平性	すべての科目が平等に扱われない場合、それを説明できなければならない。そのためには、特徴的な対象者の特性を可能な限り包括的に把握することが重要である。これは、、どの特性が実際に役割を果たし（当事者 A のリスクが当事者 B よりも低いとされ）、どの

[12] https://www.digitaleoverheid.nl/overzicht-van-alle-onderwerpen/cybersecurity/bio-en-ensia/ ベースライン・インフォマティベーリング・オーバーハイド/

	特性がそうでないか（その結果、当事者 A と B のリスクが同等であることが立証される）を実証するためである。
GDPR	一般データ保護規則。このプライバシー法は、企業や組織による個人データの管理と保護を規制している。
生成的 AI	AI の一種で、アルゴリズムを使ってコンテンツを生成する。簡単なプロンプトを入力するだけで、ユーザーはテキスト、画像、音声、コンピューターコードを瞬時に生成することができる。最も有名な例は ChatGPT である。
ガバナンス	ガバナンス、行動規範、監督組織を統治する行為や方法。これは、期待を決定し、権力を授け、パフォーマンスを検証する決定に関するものである。これは、独立したプロセス、またはマネジメントやリーダーシップのプロセスの特定の部分から構成される。
ハッキング攻撃	**AI システムへの**侵入。データの汚染、AI システム（またはその操作）に関する望ましくないリーク、ソフトウェアやハードウェアの破損などの結果を伴う。
リスクの高い AI	AI 法は、リスクの高い AI とは何かを定めている。これらは多くの場合、基本的権利および／または製品の安全性に密接に関連する製品である。いずれの場合も、個人データを処理して個人を特定する AI システムである。さらに、以下の分野におけるすべての AI システムも対象となる：禁止されていない生体データ、重要インフラ、教育・職業訓練、雇用、必要不可欠な公共・民間サービスへのアクセス・享受、法執行、移民・亡命・国境管理、司法行政、民主的プロセス。
入力（データ）	所定の目的のために処理されたデータ。**AI システムの**文脈では、これは現実からの観察などの生データを指すことがある。**モデルの**文脈では、これは通常、前処理されたデータを指す。
関心グループ	**多様性を**測定する**利害関係**者のグループ。これは、**エンドユーザーの**グループであったり、システムの影響を受ける人々のグループであったりする。
大型言語モデル	LLM はテキストを生成できる生成的 AI である。非常に大規模なデータセットで学習され、多数のパラメータを含んでいる。
リスク限定 AI	AI 法は、リスクを限定した AI とは何かを定めている。人間と対話したり、感情を認識したり、加工された画像を生成したりすることを目的とした AI である。例としては、スパムフィルター、テキストの要約、航空インシデントのトピックの分類、オフィスの照明を調整する AI システムなどがある。
低リスクの AI システム	操作や欺瞞のリスクがある AI システム。こうした AI システムは透明でなければならず、ユーザーは AI との相互作用について知らされていなければならない。
経営組織	**AI システムの**アプリケーション管理を設定し、最適化する組織。
メタデータ	他のデータの特徴を記述するデータ。例えば、誰のデータなのか、誰が送信したのか、最近変更されたのはいつなのか、などである。
ミニマムリスク AI	禁止されていない、あるいは**高リスク** AI や**限定リスク AI に**分類されていない AI システム。
モデル	情報を処理するために使用される、現実の（単純化された）数学的表現。**AI** システムでは、数学的表現は**アルゴリズムに従って**部分的または完全に学習されることが多く、その

	結果、**開発者であっても**、モデルがどのようにその結果に到達したかを完全に説明することはできない。
道徳相談	物理的環境協議会（2021 年 5 月）、モレル・ベラード
マイナスの影響	例えば、**AI** システムの**バイアスに**基づき識別的に差別されるなど、**AI システムを**適用した結果、否定的な結果を経験した当事者。
ポジティブな影響はない	定義上、AI システムの使用によるマイナスの影響を経験しないが、例えば以前と同じ立場にとどまる関係者。この場合、これらの関係者は、AI システムから、他の関係者が経験するような「プラスの影響」を経験しない危険性がある。
出力（データ）	**AI システムが**生成したデータ。これらはモデルの結果である。
パラメーター	**モデル**内の変数。この変数が変更されると、モデルや計算の結果サイズも変更される。
最終責任者	**AI システムに対して責任を**負う組織内の役割。例えば、AI システムで正しい結果が得られるようにする責任を指す。これは通常、プロセス・オーナーである。
ポジティブな影響	**AI システムを**使用した結果、有利な結果を経験する**当事者**。これは、有利な立場にある少数派である可能性がある。これは、このポジティブなバイアスが過度に楽観的であるため、現実を反映していないという危険を伴うかもしれない。また、この欠点は、他の関係者にとって「マイナスの影響」となる可能性もある。
プロジェクトリーダー	**AI システムが**属するプロジェクトの最終的な責任を負う者。**試運転依頼**者とともに、AIIA を実施する**最終責任者でもある**。
比例	比例性とは、目的と適用される解決策の間に合理的な関係を確保することである。AI の使用は、**アルゴリズムによって**解決しようとする問題に比例しているか。期待される利点は、AI がもたらすリスクを上回らなければならない。
サービス開始	AI システムが「実用化」される瞬間とは、あるプロセスで初めて使用される瞬間を意味する。これに先立ち、外部テストまたはパイロットが実施される。AIIA は、システムが稼動する時点で完了していなければならない。
信頼できる	一貫した行動と一貫した結果という特徴を持つ。
再現可能	記述された手順が実行されるたびに、同じような結果が繰り返し得られる能力。
責任がある	ある**事業**体の行為は、その事業体独自の方法で遡ることができ、その事業体はこれらの行為に対して責任を負う。誰が誰なのか？この AIIA の回答において、どの人物が役割を果たしたかを記入すること。
リスクフリーの AI システム	他のカテゴリーに該当しないすべての AI。これらのシステムは AI 法の対象外である。
堅牢性	予防的アプローチを用いて開発され、事前に予測され説明されたとおりに行動する。許容できないダメージを避ける。
シード	シード」とは、乱数生成器の出発点である。この出発点から、この生成器は常に同じ経路をたどって新しい（擬似）乱数を生成する。シードを文書化することで、（擬似）乱数の生成経路を繰り返すことができる。つまり、**モデルの**再構築を検証するためには、モデルが乱数を使用するたびにシードが必要となる。シード自体も数字である。この数字に特別な要件はなく、「認識できる」ものが選ばれることが多い（たとえば「123456」、「0, 42, 1234」、**開発**者の生年月日など）。

セレクション・リスト	例えばオランダ公文書記録法（*Archiefwet*）に従い、アーカイブ記録をどれくらいの期間保存しなければならないかを記したリスト。
ステークホルダー	意思決定や活動に影響を与えることができる、影響を受けることができる、または影響を受けていると考える個人または組織。stakeholder はまた例えば使用されるデータの所有者である場合もある。
トレーサビリティ	プロセスと結果が検証できる場合
透明	**AI システムの**運用と目的が明確にコミュニケーションされ、AI システムの結果が**説明可能で**ある場合。
アルゴリズムの種類	ニューラルネットワーク、ランダムフォレスト、その他の機械学習など、さまざまなテクノロジーを使って AI を作ることができる。しかし、ビジネス・ルールや決定木など、それほど複雑でない**アルゴリズムを**使用することもできる。
許容できないリスクの AI システム	サブリミナル的、操作的、または欺瞞的なテクニックを適用する、脆弱性を不適切に利用する、バイオメトリックに分類する、ソーシャルスコアリングを適用する、個人の犯罪の可能性を予測する、顔認識用のデータベースを作成する、感情を推測する、リアルタイムで遠隔から感情を認識する。
意図しない腐敗	例えば、誤った入力を入力したり、誤ったボタンを押したりすることで、悪意なく **AI システムの**動作に影響を与えること。意図しない破損は**信頼性に**該当する。意図的な）破損とは区別する。
ユーザー	AI 法によれば「…）認可機関、代理店、またはその他の団体が、その権限に基づいて **AI システムを**使用する。システムを使用するのはユーザーである。これは決して自然人ではない。例としては、ILT や RWS が挙げられる。

附属書 1：リスクアセスメント

AI 法には、AI システムの適用に関するリスク分野がいくつも記述されている。リスクが高ければ高いほど、より多くの対策を講じる必要がある。リスクレベルは、AI システムが展開される応用分野によって決定される。以下の質問は、AI システムのリスクレベルを判断するのに役立つ。

AI 法におけるリスクレベルの分類は絶対的なものである。つまり、プライバシー法に基づくリスクなど、その他のリスクはカウントされない。例えば、ある AI システムはプライバシーの分野で重大な影響を及ぼすかもしれないが、AI 法によればリスクは最小の応用分野であるかもしれない。つまり、リスクの高い AI システムではないということだ。 しかし、AIIA の標準的な質問は、この種のリスクを緩和するのに役立つかもしれない。

リスクの高い AI システムの定義（AI 法）

個人データを処理して個人をプロファイリングする AI システム。さらに、（AI 法で）禁止されていない生体データ、重要インフラ、教育・職業訓練、雇用、必要不可欠な公共・民間サービスへのアクセスと享受、法執行、移民・亡命・学習データ制御管理、司法行政、民主的プロセスといった分野のすべての AI システムも対象となる。

どのリスクレベルが AI システムに適用されるのか、まだわからない？その場合は、以下の質問に答えよう。

A) 容認できないリスクのアセスメント

1. AI システムはサブリミナルメッセージを使用しているのか？それとも意図的に操作したり、誤解を招くようなテクニックを使っているのか？
2. AI システムは、身体的または心理的な損害をもたらす可能性のある、脆弱性のあるグループに対する操作／虐待を利用しているか？
3. AI システムは**ソーシャルスコアリングを**利用しているのか？[13]
4. この AI システムは、医療上の安全保障を考慮する以外に、仕事中の人の感情を推測することを目的としているのだろうか？
5. AI システムは公共の場で生体認証を使うのか？あるいは、特別な個人データに基づいて人々を生体データで分類するのか？

すべての質問に「いいえ」と答えたか？もしそうなら、AI システムはおそらく許容できないリスクのアプリケーションではない。

B) ハイリスクのアセスメント

1. AI システムは個人をプロファイルしているのか？
2. その AI システムは、以下のいずれかの分野の製品、または製品のセキュリティ・コンポーネントであるか：
   - **機械** (指令 2006/42/EC)
   - **玩具**（指令 2009/48/EC）
   - **娯楽用ボートおよび水上バイク**（指令 2013/53/EU）
   - **リフト**（指令 2014/33/EU）
   - **爆発の可能性がある雰囲気での使用を意図した機器および保護システム**（指令 2014/34/EU）
   - **無線設備**（指令 2014/53/EU）

[13]社会的スコアリング：社会的行動、または既知、派生、予測される個人的性格特性に基づいて、自然人または集団の特定期間中の評価または分類を行う AI システム（AI 法 51 条）。

- **圧力機器**（指令 2014/68/EU）
- **索道設備**（規則（EU）2016/424）
- **個人用保護具**（規則（EU）2016/425）
- **ガス燃料を燃焼する機器**（規則（EU）2016/425）
- **医療機器**(規則(EU) 2017/745)
- **体外診断用医療機器（規則（EU）2017/746）**

同規則はまた、高リスクの AI アプリケーションとみなされる製品のリストにも言及している。第 6.1 条、第 102 条から第 109 条、第 122 条を除き、これらの製品には AI 法に基づく高リスク義務はまだ適用されない。しかし、これらの製品に適用される具体的な製品法の策定においては、後の段階で AI 法の要件が活用されることになる。この時期がいつになるかはまだわかっておらず、製品によって異なる。これは以下の製品と法律に適用される：

- **(防御）民間航空**（規則(EC) 300/2008 および規則(EU) 2018/1139）。
- **二輪または三輪自動車および四輪自動車**（規則(EU) 168/2013）
- **農林業用車両**（規則(EU) 167/2013）
- **舶用機器**（指令 2014/90/EU）
- **EU における鉄道システムの相互運用性**（指令（EU）2016/797）
- **自動車とそのトレーラー**（指令（EU）2018/858 および指令（EU）2019/2144）

3. この AI システムは、遠隔生体認証を使用しているのだろうか？
4. AI システムは、重要なデジタル・インフラ（[14]）、道路交通、水、ガス、暖房、電気の供給などの管理・運営におけるセキュリティ・コンポーネントとして使用されるのか？
5. AI システムは、人々の採用や仕事へのアクセスに影響を与えるのだろうか？
6. AI システムは、必要不可欠な民間および公的サービスへのアクセスや利用、支払いに関して使用されるのか？
7. AI システムは法執行で活躍するのだろうか？
8. AI システムは移民、亡命、国境管理に関する分野で使用されるのか？

すべての質問に「いいえ」と答えたか？もしそうなら、AI システムはリスクの高いアプリケーションではないだろう。

C)透明性義務のアセスメント

1. 人と対話する AI システムが関係しているのだろうか？
2. AI システムは人工的なコンテンツ（生成的 AI/GPAI）を生成または操作できるか？
3. AI システムは感情を認識することができるのか？
4. システムはディープフェイク技術を使用しているか？[15]

上記の質問のいずれかに「はい」と答えた場合、必要な透明性対策を講じるべきである。[16]一般的に言えば、これは「ユーザーが AI システムを扱っていることを確実に知ること」を意味する。

[14]重要なデジタルインフラ：インターネットエクスチェンジポイント、DSN サービスプロバイダー、トップレベルドメイン名登録、クラウドコンピューティング、データセンター。

[15]AI によって作成または操作された画像、音声、映像素材やテキストは、人々に誤って本物や真実と見られる可能性がある。

[16]AI 法第 50 条

例外

システムが以下のいずれかの分野にのみ適用される場合は、AIIA の標準的な質問で十分である。例外の対象となる高リスクのシステムの場合は、AI システムを登録する必要がある（アルゴリズム登録で可能）。

注意：これは AI システムが個人をプロファイリングしない場合にのみ適用される。そうでない場合、システムは**常に**リスクの高いアプリケーションとなる。[17]

例外、AI：

1. 軍事目的
2. 法の執行と司法
3. AI の研究
4. 開発段階（市場参入前）のオープンソース AI
5. プロフェッショナル以外の活動
6. AI システムは基本的に意思決定の結果に影響を与えない。例えば、純粋に手続き的なタスクを実行するシステムや、以前の人間の活動の検証や改善（例えば言語の修正）を行うシステムなどが挙げられる。意思決定のパターンを検知するためのシステムもこれに含まれる。[18]

注：AI の応用に関する多くの定義は、まだ発展途上である。そのため、これらの質問を終えても、AI システムがどのカテゴリーに属するのかがまだ完全に明確になっていない可能性がある。この質問と他の質問を、オランダ中央政府の Regulatory AI Sandbox（詳細）に提出することを検討するとよいだろう。

[17]AI 法第 6 条 3 項

[18]AI 法第 6 条 3 項

附属書 2：リスクの高いシステム

AI システムはリスクの高いアプリケーションか（附属書 1 参照）？以下の追加質問を参考にしてほしい。これにより、AI システムを責任を持って使用するための適切なセーフガードを適用することができる。

AI 法では、さまざまな関係者が記載されており、それぞれが AI システムに関して独自の役割や責任を負っている。この附属書には、AI システムの利用者（展開者）と開発者（プロバイダ）を対象とした質問が含まれている。例えばプロバイダは、AI システムを調達し、自分の名前で提供する。あるいは、システムに大幅な変更を加える。[19]

リスクの高い AI システムを使用したい場合の質問

AI 法によれば、リスクの高いシステムはいくつかの要件を満たさなければならない。以下の質問はその助けとなるだろう。

効果

- AI システムは、人々の健康、安全、基本的権利に重大なリスクをもたらすか？20その理由を以下に示すこと。

保守・管理

- システムの監視と運用はどのように保護されるのか？21
- AI システムが作成したログは、少なくとも 6 カ月間はどのように保存されるのか？22
- 必要な能力、訓練、認可を受けた人間による監督を手配する措置が取られているか？

技術的堅牢性

- データ（入力）は、AI システムの意図された目的（1.1 の質問 1）を考慮した上で、適切かつ代表的なものであるか。23
- AI システムの技術文書は提供されているか24
- 人間の監視のために、出力（データ）を検証したり、解釈したり、場合によっては無視したりすることは可能か？25

信頼性

- 正確さのレベルに関する指標は、使用説明書に記載されているか？

データガバナンス

- AI システムの学習データは、以下の品質要件を満たしているか？
- データセットに関連するデザインの選択。
- データの出所に関する透明性をプロバイダに提供する。
- 注釈、ラベリング、クリーニング、更新、強化、集計などの関連する処理活動。

---

[19]AI 法第 25 条。

[20]AI 法第 6 条 2a。

[21]AI 法第 26 条 5 項

[22]AI 法第 26 条 6 項

[23]AI 法第 26 条 4 項。

[24]AI 法附属書 IV。

[25]AI 法第 14 条 4 項。

- データが測定し、代表者でなければならない前提条件を設定すること。
- 必要なデータセットの入手可能性、量、適切性のアセスメント。
- 健康、安全、基本的権利、識別的差別に悪影響を及ぼすデータの潜在的バイアスについてのアセスメント。
- バイアスを検知、防止、制限するための措置。
- 規制遵守を妨げる欠点を識別し、それに対処する。

リスクマネジメント

- AI システムは、使用開始前にどのようにテストされ、意図された目的とリスクマネジメント要件を満たしていることを確認したのか？
- リスクマネジメントシステムが決定され、文書化されているか？これには以下のステップが含まれる：
  - 健康、安全、基本的権利に関する AI システムのリスク分析
  - 起こりうるリスクのアセスメントと評価
  - AI システムのモニタリング・システムに基づく、市場参入後の新たなリスクの評価26
  - リスクマネジメント対策の立案
- 監督当局やその他の所轄当局との協力の可能性をどのように保護するのか。27これは、連絡担当者、アクセシビリティなどの分野を指す。
- 脆弱性グループ（子供など）が AI システムにアクセスする可能性はあるか？その場合、リスクマネジメントシステムは特に厳しくする必要がある。

注：AI オフィスは、これらの義務を果たすための簡略化された方法を展開するプロバイダに提供するため、一部自動化ツールを使用した質問票のテンプレートを開発中である。

コミュニケーション

- リスクの高い AI システムの実用化について、どのようにコミュニケーションするのか？28
- AI システムは EU の高リスクシステムのデータベースに登録されているか？29
- 使用説明書は作成されているか？少なくとも以下の内容が含まれていなければならない：30
  - プロバイダの身元と連絡先；
  - 特徴、能力、限界（目的）；
  - 将来的な変更の可能性
  - 人的監督に関する措置；
  - 必要な計算資源とハードウェア資源、予想される寿命、必要な保守・管理手段（頻度を含む）；
  - AI システムに含まれるメカニズムの説明。

[26]市販後モニタリングシステムは，展開者から提供される，又は他の情報源を通じて収集される，リスクの高い AI システムの性能に関する関連データを，その耐用期間を通じて，積極的かつ体系的に収集，文書化及び分析しなければならず，これにより，プロバイダは，AI システムが第 III 章第 2 節に定める要件に継続的に適合していることを評価することができる。

[27]AI 法第 26 条 12 項

[28]AI 法第 26 条 7 項

[29]AI 法第 26 条 8 項、第 49 条、第 71 条。

[30]AI 法第 13 条 2 項および 3 項。

生成的 AI

- AI システムの可能性と限界を明確にする情報や文書はあるか？
- AI システムによって著作権がどのように保護されるかについての方針はあるのか？
- AI がトレーニングを受けた内容の詳細な要約がプロバイダから提供されているか？

その他

- - 遠隔生体認証を使用している場合、司法当局の許可は得ているか。31

リスクの高い AI システムの開発者（プロバイダ）への質問

リスクの高い AI システムの開発者（AI 法でいう**プロバイダ**[32]）は、以下の質問にも答えなければならない。

- 要求事項への準拠を示す文書が作成されているか。要件については AI 法附属書Ⅳを参照のこと。
- 技術文書はどのようにしてユーザーが見つけることができるのか？たとえシステムの導入に関与していなくても？
- このシステムがハイリスクであることを示すプロバイダは提供されているか、そうであればどのようにか。33
- 品質マネジメントシステムは構築されているか。34また、これを文書化する必要がある。以下のような最低限の情報が必要となる：
  - 規制遵守のための戦略
  - 設計手順
  - 品質管理手順
  - 検査手順
  - 技術的な要求事項の概要と、それを実施するために必要なこと
  - データ管理手順
  - リスクマネジメント・システム
  - モニタリングの手順
  - 重大インシデントの報告手順
  - 所轄当局とのコミュニケーション戦略
  - 関連書類の登録に関するシステムと手順
  - 資源管理および供給安全保障関連対策
  - 説明責任の枠組み
- AI システムのログはどのように保存されるのか？[35]
- 適合性評価手順があり、EU 適合宣言書に記録されているか。36AI システムに対して決定された要求事項が満たされていることをどのように証明するのか。これらは指定団体に提示し、少なくとも 10 年間は保管しなければならない。

---

[31]AI 法第 26 条 10 項

[32]**プロバイダ**」：AI システムもしくは汎用 AI モデルを開発する、または AI システムもしくは汎用 AI モデルを開発させ、有償無償を問わず、自らの名称もしくは商標の下に上市もしくはサービスインする自然人もしくは法人、公的機関、代理店その他の団体。

[33]AI 法第 16 条。

[34]AI 法第 17 条。

[35]AI 法第 19 条。

[36]AI 法第 43 条および第 47 条。

- 適合宣言は関連監督当局によって認可されているか。37これは、リスクの高い AI システムが関連するハイリスク製品（附属書Ｉ）に関連する場合にのみ必要である。
- AI システムが AI 法に準拠することを意図している場合、文書などで CE マークがプロバイダされているか？38
- システムが AI 法に適合しなくなった場合、このシステムの市場からの撤去、使用停止、回収などの是正措置はどのように取られるのか？ユーザーにはどのように通知されるのか。39
- システムは欧州のアクセシビリティ要件に準拠しているか？40

[37]AI 法第 16k 条。

[38]AI 法第 48 条。

[39]AI 法第 20 条。

[40]指令（EU）2016/2102 および（EU）2019/882。

附属書 3：生成的 AI に関する留意点

大規模言語モデル（LLM）のような生成的 AI を使用する場合、AIIA を完了する際に考慮すべき多くの重要なポイントがある。この附属書では、生成的 AI が他の AI システムと異なる最も重要な点に焦点を当てる。

生成的 AI とは何か？

生成的 AI とは、アルゴリズムを使ってコンテンツを生成する AI の一種である。オランダ内閣は、生成的 AI に関する政府全体のビジョン（[41]）において、生成的 AI は人間の幸福と自律性、持続可能性、繁栄、正義、安全・安心の改善を目的としたものでなければならないと述べている。このビジョンによれば、生成的 AI の責任ある応用によって、我々はこの技術が提供する機会をつかむことになる。

生成的 AI は、数十億のパラメータからなるニューラルネットワークを利用する。選択される出力は統計に基づくもので、現実に関する基礎的な論理や知識はない。出力は様々で、再現することも説明することも難しい。このことは、AIIA を完成させる前に考慮すべき最も重要なポイントにすぐにつながる。生成的 AI を使用すべきなのは、結果が説明不可能または検証不可能でも構わない場合だけである。

中央政府組織の暫定見解：原則として認められない

中央政府組織における生成的 AI の使用に関する暫定的な見解は、現在、中央政府における LLMS の使用に厳しい要件を設定している：「ChatGPT、Bard、Midjourney のような非契約の生成的 AI アプリケーションは、一般的に、関連するプライバシーと著作権に関する法律に明確に準拠していない。このため、プロバイダと利用者が関連法規を明白に遵守しない限り、中央政府組織による（あるいは中央政府組織のための）それらの利用は、法律違反のリスクがある場合には原則として許可されない。"

AIIA を修了する際の留意点

以下は、生成的 AI が他の AI システムと異なる最も重要な点についての説明である。これは AIIA を完成させる際に役立つだろう。もし AIIA のいずれかの分野が以下に含まれていない場合、それは関連性がないという意味ではなく、生成的 AI であるという事実に関連する考慮すべき点が特にないという意味である。

目的と必要性

- **公共的価値：**生成的 AI には特に考慮すべき点はない。生成的 AI を使うという選択は、しばしば透明性、説明可能性、持続可能性を損なう可能性がある。
- **基本的権利：**倫理的な方法でモデルが訓練され、微調整されたかをチェックする。これは、反応を改善するために人間のフィードバックが使われる最終段階である。他の多くの製品の場合と同様、これらの人々（「クリックワーカー」）は必ずしも最良の労働条件を持っているわけではない。[42]
- **持続可能性：**より複雑でないシステムやツールを使って、同じ目的を達成することが可能かどうかを調べる。生成的 AI は、学習にもモデルの使用にも膨大なエネルギーを消費する。モデルの小型化、最適化、別のソフトウェア・アーキテクチャなど、省エネ技術を適用できないか調査する。
- **ガバナンス：**生成的 AI を使用するかどうかを決定する際には、中央政府組織における生成的 AI の使用に関する暫定決定を考慮しなければならない（序文を参照）。

技術的堅牢性

- **バイアス：**システムの出力にバイアスがないか常にチェックし、バイアスのリスクは何かをアセスメントする。生成的 AI モデルのトレーニングに使われるデータにはすでにバイアスが含まれていることが多く、モデルの性質（統計的言語モデル）がこのバイアスを悪化させ

[41] https://www.tweedekamer.nl/kamerstukken/brieven_regering/detail?id=2024Z00480&did=2024D01191

[42] AI は、数百万人の有能な若手社員の仕事を引き受けている。そのために何が必要なのか？ - 特派員

る。ソフトウェア・サプライヤーは、人間のフィードバックやガードレールを使って鋭いエッジを取り除くが、バイアスはモデルに残る。さらに、バイアスを減少させたり検知したりするための他の方法も利用できる。[43]

- **正確性、信頼性、再現性：**生成的 AI モデルは、高水準の**正確性**、信頼性、再現性を必要としない目的にのみ使用すべきである。 信頼できない結果がもたらす潜在的リスクのアセスメントを行う。LLM は現実を知らない統計的言語モデルである。出力は常に異なり、必ずしも現実に基づいているわけではない。生成的 AI モデルには「幻覚」を見る可能性もあり、論理的で説得力があるように聞こえるかもしれないが、真実ではない回答を生成する。生成的 AI のもうひとつの難しい特徴は、出力の質を測定するのが難しいことだ。
- **技術的な実装：**外部でホストされる生成的 AI モデルの場合、組織のデータ（入力）の使用について契約上の取り決めを行う必要がある。モデルをさらに訓練するためのデータとして使用したいという要求がしばしばある。適切なプロンプトを利用すれば、他の（権限のない）人がこれらのデータを（使用中に）閲覧できる可能性がある。システムを調達する際、サプライヤーは設定されたすべての要件を満たしていることを証明しなければならない。生成的 AI に関するオランダ内閣の見解は、オープンソースの生成的 AI の使用を優先することを表明している。
- **説明可能性：**結果の説明可能性が要件でない場合にのみ、生成的 AI モデルを使用すべきである。説明可能性の欠如がもたらす潜在的リスクをアセスメントすること。そのシステムで仕事をしなければならないスタッフには、どのような制限があるのかを説明すること。生成的 AI は何十億ものパラメータを持つ非常に複雑なシステムであり、その動作は説明不可能である。現在、この透明性を改善するための取り組みが行われているが、モデルがどのように反応に至るかを示すことは事実上不可能である。[44]

データガバナンス

- **データの品質と完全性：**生成的 AI のトレーニングデータには、個人データや著作権で保護された素材が含まれている可能性が高い。ほとんどの AI モデルでは、どのトレーニングデータ（およびその質）がモデルに使用されたかは不明である。AI 法は透明性を求めている。GDPR は個人データを単に処理することを認めていない。
- **プライバシー：**入力されたデータの処理責任（トレーニングデータとして使用するかどうかなど）についてサプライヤーと合意する。適切な処理契約書を作成する。適切な合意がなされれば、データ侵害の可能性を減らすことができる。
- **情報セキュリティ：**OWASP は、生成的 AI システムの最も重大な脆弱性、ひいては情報セキュリティリスクを 10 個リストアップした[45] 。

説明責任

- 検証可能性：「説明可能性」を参照のこと。

[43]https://www.datacamp.com/blog/understanding-and-mitigating-bias-in-large-language-models-llms (44)

[44]https://arxiv.org/abs/2309.01029

[45]OWASP トップ 10：LLM と生成的 AI のセキュリティリスク

の出版物である：

インフラ・水管理省

私書箱20901

2500EX　ハーグ

2024年12月　　| 73263