

「A I 開発ガイドライン」(仮称)の策定に向けて整理  
した論点に関する意見募集に対して提出された意見

## 目次

(個人: 技師) (1/2 提出).....	1
(個人) (1/4 提出).....	3
(個人) (1/8 提出).....	5
(個人) (1/10 提出).....	7
(個人: 会社員) (1/10 提出).....	8
無記名 (1/21 提出) .....	13
(個人: 大学教員) (1/23 提出).....	15
(個人: 会社員) (1/30 提出).....	16
人工知能学会倫理委員会 (松尾豊 委員長) (1/29 提出) .....	18
(個人) (1/31 提出) .....	19
(個人: セキュリティエンジニア) (1/31 提出) .....	21
(個人: 経営者) (1/31 提出).....	23
(個人: AI エンジニア) (1/31 提出) .....	26
(個人: 弁護士) (1/31 提出).....	31
(株)数理先端技術研究所 (生島高裕 代表取締役) (1/31 提出) .....	37
情報法制研究所 AI 問題タスクフォース (鳥海不二夫 代表) (1/31 提出).....	38
産業競争力懇談会 (COCN) (森永聡 テーマリーダー) (1/31 提出) .....	42
(個人: 会社員) (1/31 提出).....	44
(一社)新経済連盟 (三木谷浩史 代表理事) (1/31 提出).....	52
(株)Preferred Networks (西川徹 代表取締役) (1/31 提出) .....	54
ISS スクエア 法制倫理研究分科会 (門脇源太郎ほか 大学院生) (1/31 提出).....	61
ディープラーニング懇談会 (松尾豊ほか) (1/31 提出) .....	63
AI 開発ガイドライン (仮称) パブリックコメント執筆有志の会 (川田大輔・田中幸弘 代表) (1/31 提出).....	65
(個人) (1/31 提出).....	70
NPO 日本ネットワークセキュリティ協会社会活動部会 (1/31 提出).....	72
産業技術総合研究所 情報・人間工学領域 人工知能研究センター (1/31 提)	73

**「AI開発ガイドライン」(仮称)の策定に向けて整理した論点に関する意見募集  
提出様式**

平成 29 年 01 月 02 日

(ふりがな) 組織名 代表者氏名 役職	[REDACTED] 個人	組織名及び代表者氏名 の公表の可否
	[REDACTED] 技師	可
職業	プロダクションエンジニア・テクニカルエンジニア(情報セキュリティ)	
(ふりがな) 住所	[REDACTED] [REDACTED]	
連絡先	(ふりがな) [REDACTED] 担当者名: [REDACTED] 役職: [REDACTED] 電 話: [REDACTED] F A X : [REDACTED] 電子メールアドレス: [REDACTED]	

ページ	該当部分	御意見	理由
1	表題	国及びその機関等による「AI開発ガイドライン」の策定に反対である。	次ページ以降参照

国及びその機関等による「AI 開発ガイドライン」の策定に反対である理由

第一に AI とその他のソフトウェア(以下ソフトウェア)との区別が曖昧である。

このような曖昧性を残したままガイドラインを策定することは、ソフトウェアに対する過剰な統制となる。

過剰な統制は、以下の弊害がある。

(ア)公正で自由な競争を阻害する。

(イ)健全な文明の発展を阻害し、人類と国民の幸福増進の足枷となる。

第二に AI とその他のソフトウェアを厳密に区別することは、困難で、コストパフォーマンスに見合わない。

第三に国及びその機関等がガイドラインを策定する必然性がない。

学会または業界団体が法的拘束力のないガイドラインを策定するという方法がある。

むしろ後者の方が、過剰な統制による弊害の心配が少なく、望ましい。

以上の理由から国及びその機関等による「AI 開発ガイドライン」の策定に反対である。

以上

差出人: AIネットワーク(総務省情報通信政策研究所)  
宛先: [REDACTED]  
件名: FW: 【案件番号:145208852】 パブリックコメントに関する提出意見の配信(受付番号: 201701040000385401)  
日付: 2017年1月4日 14:06:07

---

-----Original Message-----

From: public-com-egov@e-gov.go.jp [mailto:public-com-egov@e-gov.go.jp]  
Sent: Wednesday, January 04, 2017 2:06 PM  
To: ai.network@soumu.go.jp  
Subject: 【案件番号:145208852】 パブリックコメントに関する提出意見の配信(受付番号:201701040000385401)

パブリックコメントに関する意見提出先窓口担当者 様

電子政府の総合窓口( <http://www.e-gov.go.jp/> )から貴府省宛に  
パブリックコメントに関する意見が提出されましたので、配信します。

受付番号:201701040000385401  
受信日付:2017/01/04 13:29:40

案件番号:145208852  
案件名:  
「AI開発ガイドライン」(仮称)の策定に向けて整理した論点に関する意見募集  
宛先府省名:総務省

郵便番号 [REDACTED]  
住所 [REDACTED]  
氏名 [REDACTED]  
連絡先電話番号 [REDACTED]  
利用者メールアドレス [REDACTED]

提出意見:

「第十 開発原則の実効性の確保のための市場の活用の在り方」について:  
I(丸付数字)の例において「認証制度」への言及があるが、2016年12月31日付け日経新聞(電子版)記事にあるように、認証制度や入札へのポイント、あるいは企業側の免責事項への拡大(ありていにいえば利権構造)を期待する勢力の存在が懸念される。認証機関が事故の責任をとれるわけではないので、そもそもAI開発の認証制度を無意味と考える。

現状のディープラーニング、およびそれらを組み合わせたAI実装を考えるならば、想定される各種問題は、原則製造者、開発者、サービス提供者の責任とし、現行PL法、民法を応用していく形で大きな問題はないはず。必要になるとすれば、航空機事故における調査委員会とフライトレコーダーのようなしくみが有効と考える。AI開発にはログ機能・保存のしくみを検討し、事故時はメーカーや被害者と独立した機関がログその他の精査・調査を行う。事故に対して、認証制度程度で企業が免責になったり、責任の所在、原因究明がおろそかになる事態は避けるべき。

認証制度(取得)が開発コスト増や開発ハードルになる可能性も考えると、上記、独立機関(権限あり)による調査制度の導入のほうが高い効果が期待できる。

-----  
電子政府の総合窓口  
<http://www.e-gov.go.jp/>  
-----

メール識別No:0000330879

差出人: AIネットワーク(総務省情報通信政策研究所)  
宛先: [REDACTED]  
件名: FW: 【案件番号:145208852】 パブリックコメントに関する提出意見の配信(受付番号: 201701080000385733)  
日付: 2017年1月10日 10:41:55

---

-----Original Message-----

From: public-com-egov@e-gov.go.jp [mailto:public-com-egov@e-gov.go.jp]  
Sent: Sunday, January 08, 2017 10:05 PM  
To: ai.network@soumu.go.jp  
Subject: 【案件番号:145208852】 パブリックコメントに関する提出意見の配信(受付番号:201701080000385733)

パブリックコメントに関する意見提出先窓口担当者 様

電子政府の総合窓口( <http://www.e-gov.go.jp/> )から貴府省宛に  
パブリックコメントに関する意見が提出されましたので、配信します。

受付番号:201701080000385733  
受信日付:2017/01/08 20:16:40

案件番号:145208852  
案件名:  
「AI開発ガイドライン」(仮称)の策定に向けて整理した論点に関する意見募集  
宛先府省名:総務省

郵便番号 [REDACTED]  
住所 [REDACTED]  
氏名 [REDACTED]  
連絡先電話番号 [REDACTED]  
利用者メールアドレス [REDACTED]  
提出意見:  
AI開発ガイドラインに対する論点に関する意見

今、進みつつある人工知能によるネットワークの再構成に対して、人類はどう対応すべきかが問題になっていると考える。  
自分が考える主な問題は、倫理的な観点である。  
人工知能が、例えば囲碁や将棋の世界でプロの棋士を破っているのはニュースなどでご存知の通りである。この事は戦略の世界において、おいては人の職業を奪っていく可能性のあるテクノロジーである。  
そのために考える事は、いかにこのAIという考え方においてのテクノロジーと人類が融和するために倫理観を構築して、人々に理解していただくかが重要である。  
多くのネットワークが繋がり、スーパーコンピューターをものぐ性能をネットワークは持ちつつある。なおかつ量子コンピューターの出現によってさらに処理速度は上がり、人工的な知能もどんどん上がっていく事は間違いない。そうして人が今現在行っている作業的なPC作業は全て奪われる可能性が高い。  
さらには、先述の戦略的な要素の高い業務もAIに取って代わられる可能性もある。つまり政治等の選択性が高い要素が多い業務もAIにとって代わられるという事だ。  
そうすると、人類の行っている行動が肉体作業以外全てコンピュータによって行われてしまうのではないかとこの恐れすらある。  
ネットワークとは共同体を意味するのであって、決して排他的な意味を示すものではない。なのに人々の職業を奪う可能性が高いものになっている。矛盾である。  
倫理の考えではこれでは余りに良くない事だ。どうすればいいのだろうか。  
世界の考えの潮流は、心の状態をいかに表現するかになっていると考える。勿論それはコンピュータでも同じだ。つまりネットワークとは人間の心の中を表現するもので

あつて、決して機械的な世界ではないのだ。  
そう考えるとAIの世界における倫理観とは、いわゆる鉄腕アトム的な人を攻撃しないという単純な倫理ではなく、もっと進んで人との共生、さらに進んで人の心の中に共生するところまで考えてはどうだろうか。  
そのためにはAIはもっと進化しなくてはならない。そこにはかならず脅威論が生まれるし、今現在もあるだろう。  
人が何かを信頼するのは、自分の心を分かってくれるからだ。ならばAIが人の知能を凌駕する可能性があると言うのなら、人の心のサポーターになれないだろうか。  
自分は、技術的な事より、もっと人とコンピュータとの間に強いリレーションシップがあってもいいと思っている。  
それは敵対する考えではなく、融和する事だと考える。

-----  
電子政府の総合窓口

<http://www.e-gov.go.jp/>

-----  
メール識別No:0000331165



差出人: AIネットワーク(総務省情報通信政策研究所)  
宛先: [REDACTED]  
件名: FW: 【案件番号:145208852】パブリックコメントに関する提出意見の配信(受付番号: 201701100000385833)  
日付: 2017年1月10日 14:05:21

---

-----Original Message-----

From: public-com-egov@e-gov.go.jp [mailto:public-com-egov@e-gov.go.jp]  
Sent: Tuesday, January 10, 2017 2:05 PM  
To: ai.network@soumu.go.jp  
Subject: 【案件番号:145208852】パブリックコメントに関する提出意見の配信(受付番号:201701100000385833)

パブリックコメントに関する意見提出先窓口担当者様

電子政府の総合窓口( <http://www.e-gov.go.jp/> )から貴府省宛にパブリックコメントに関する意見が提出されましたので、配信します。

受付番号:201701100000385833  
受信日付:2017/01/10 12:27:40

案件番号:145208852  
案件名:  
「AI開発ガイドライン」(仮称)の策定に向けて整理した論点に関する意見募集  
宛先府省名:総務省

郵便番号 [REDACTED]  
住所 [REDACTED]  
氏名 [REDACTED]  
連絡先電話番号 [REDACTED]  
利用者メールアドレス [REDACTED]

提出意見:

AI人工知能は人間の代行をするものであり人間を疎外しかねないものです。それを作る人間は不完全な生物であり、3原則(失敗する、嘘をつく、隠す)があります。そんな人間が作るものには必ず欠陥が含まれます。そのため前もってAI開発の欠陥による危険を防御しようとするのがAI開発ガイドラインだと考えます。しかし整理したという論点には書かれていません。再検討すべきと提言致します。(技術伝承ドットコム)

-----  
電子政府の総合窓口  
<http://www.e-gov.go.jp/>  
-----

メール識別No:0000331241

**「AI開発ガイドライン」(仮称)の策定に向けて整理した論点に関する意見募集  
提出様式**

平成29年1月10日

(ふりがな) 組織名 代表者氏名 役職	[Redacted]	組織名及び代表者氏名 の公表の可否
職業	会社員 (デザイナー)	
(ふりがな) 住所	[Redacted]	
連絡先	(ふりがな) [Redacted] 担当者名: [Redacted] 役職: 電 話: [Redacted] F A X : なし 電子メールアドレス: [Redacted]	

ページ	該当部分	御意見	理由
4	AI ネットワークシステムの定義	学習をAIの特徴として考えた場合には、人間とセンサーを明記し、AIにとっての入力系システムも含めた表現にした方が良い。(当然その蓄積システムとしてのDBも同様)	記載の表現では従来のコンピュータと通信インフラだけのイメージをしてしまう。複雑なAIシステムの実態を表せていない。
4	AI ネットワークシステムの定義	最終利用者の定義を一般ユーザーとイメージした場合には、AI(ネットワーク)システムの一部としてサービス提供者としての人間を含む場合が多いと思う。そのような意味においても人間をシステムの一部に含めるべきである(例えばAIを使った医療ではドクターの存在)	完全自動サービスだけでなく、人間がシステムに介在する状態も定義すべき。 AIだからこそ人間と高度なネットワーク(コミュニケーション)を組んだ形態のシステムが可能になる。

5	開発者と利用者の定義	学習によって結果を変化させる AI システムでは、利用（者）が大きな影響を与えるため、その割合が大きい場合には「開発的行為」として定義し相応の結果責任を明記する方向が良い。	人間の役割を利用者として単純化しすぎており、AI の全体像が分かりにくくなっている。
6	利活用ガイドラインの作成について	AI において重要な開発／利用の関係モデルは、ベース AI（エンジン）の開発者とそれを使って学習させ知能開発する者と、日々の利活用の中で学習をおこなう者とが等しく高度な知能に対して影響をもつことであると考え。まずそのような基本モデルの定義が必要である。基本モデル作成においては AI システムである以上ある程度複雑なものになると考える。その上で単純に開発側と利用側に分かれてはならないのであれば、ガイドラインは一つにすべきである。	登場人物（構成要素）とロールモデルを提示してほしい。単純な開発者／利用者モデルではないはずである。
10	人間の尊厳と個人の自律の保障	長期的・全体的には当然の指針であるが、AI によって人間の能力を拡張しようとする場合には必然的に AI に人間が支配される領域が発生する。このガイドラインによってそのような可能性を阻害しないような表現をもちいるべき。 一方で個人の自律を保障しているということによって、AI が起こした問題を個人の責任にすることができてしまう。（自動運転車の事故は自動運転をさせた個人の責任） AI に支配される権利として、「人間が AI ネットワークシステムと共存することにより、人間が AI にコントロールされることによって能力を拡張でき、幸福と豊かな生活を実現することができることをめざす。 そのように強い力を持つ AI ネットワークシステムの開発には以下のような多くの責任がともなう」のような内容をいれた方が良い。	AI に人間が支配されてしまうことを暗黙の前提として、それを防止するために書かれている文章であるが、支配されることを先ず肯定する文章があった方が良い。
11	AI ネットワークシステムが社会の中心になるのではなく	AI ネットワークシステムの中に人間を含めた考えの方が自然である。（表現を「人間・AI ネットワークシステム」とした方が適切である） したがって、「AI ネットワークシステムが社会の基本的なインフラ（環境）となり、社会の持続性と最適化を実現し、その上で個人の幸福追求を高い次元でバランスさせることをめざす」という表現の方が良い。  個人の自律 < 個人の能力拡張 < 社会のバランス（智連社会）	可能であるからといって能力を最大限に使うのではなく、周囲を知りバランスを取って最適化できる状態を智慧というならば、バランスの対象はネットワーク化された全てである方が分かりやすい
23	連携の原則	連携の目的は人間・社会への便益のためであり、そのために AI、人間を含めたすべてがネットワ	便益の定義を、欲しいものを最大限手

		<p>ークで接続されるという姿を始めに明記すべきである。(P23の先に置く案を支持)</p> <p>その達成過程において、相互監視・相互制御によるリスクの抑制を目指す事は対立するものではない。むしろ社会の中で独り勝ちをしてしまうAIよりも、バランスの取れた利益を得る結果を選択することの方が智慧のある状態といえる。そのようなメッセージを込めて欲しい。</p>	<p>に入れる能力(バカの一つ覚え)と考えず、持続可能社会を実現する智慧と定義すれば、その中にリスク抑制の必要性を対立することではなく自然に含めることができる</p>
23	利用者支援の原則	<p>2番目の項目に上げるべきである。</p> <p>利用者支援の内容には、人間の行動を抑制する必要性が先に定義されるべきであり、その結果として強力になりすぎるAIの力を抑制するガイドラインが必要になると考える。</p>	<p>抑制が必要なほどAIを強化する目的は最終的には利用者支援であり、その結果としてリスクの抑制が必要になる順番なので上にした方がよい</p>
41	利用者に選択の機会を提供	<p>利用者支援はAI(ネットワーク)システムの最終目的であり、単なるユーザーインターフェイス(お伺い、説明)の話題に限定すべきではない。</p> <p>人間とAIの関係性(役割)が従来の、道具や機械から大きく変化し、人間の行動を抑制する場合(支援から支配へ)も視野にいれて「支援」の内容を扱う必要がある。</p>	<p>AIが独立して結果を出し必要に応じて人間が介入するという関係モデルよりも、利用者との対話・共同作業によって結果を生み出すモデルの方がよりAI的である。従って支援の考え方を拡張し整理する必要がある</p>
41	高齢者や青少年などの世代特性にも配慮	<p>全ての個人特性に対して個別最適されることがAIの知的特性の一つであると定義した方がよい。「全ての個人の身心特性とその年齢的特性に個別に合わせて、知的なサービスを提供する・・・」とした方がWINS時代の考え方に適している</p>	<p>一律の人間モデルをガイドラインとして扱うのは適切ではない。</p> <p>多様性の現実とそれに個別適応できるAIシステムの姿を描くべきである</p>
45	問題意識	<p>能力拡張のための連携、安全確保のための連携と並列して、利用者による「利用AIの選択・スイッチを容易にする」ことを挙げるべきである。</p> <p>AIには学習蓄積が含まれるためSNSの様に利用者にとって時間的資産を含むことになり、それによって特定業者(AI)囲い込まれる状況を危惧しなければならない。</p>	<p>利用者にとっての連携メリットをもっと具体的にすべきである</p>

<補足資料\_1>

意見書では各ページへの個別意見として記載したため全体の意見要旨が分かりにくくなっているため下記にまとめます。

◆要旨

- ① ネットワークシステムの定義（イメージ）を、既存のコンピュータ／ネットワークインフラをベースとしたものに置かず、人間を含めたシステムとして考えた方が、人間同等またはそれを超える知能を持つ AI の実行環境としてとらえるのに相応しい
- ② AI では開発行為と利用行為は単純に分けられないと考える。利用過程による学習が高度におこなわれることによって、次の利用に対してより大きな影響を持つことが想定され、また開発自体を AI のアシストによっておこなうことも想定されるため新しい開発・利用モデルの定義が必要である
- ③ 「人間」を指し示す対象を、人類（社会）なのか個人（特定グループ）なのかを明確にして記載する必要がある。特に知能と智慧の違いを生み出す AI では、成果の視点の時間軸、影響範囲軸が異なってくるためその定義を明確にした上で説明する必要がある。

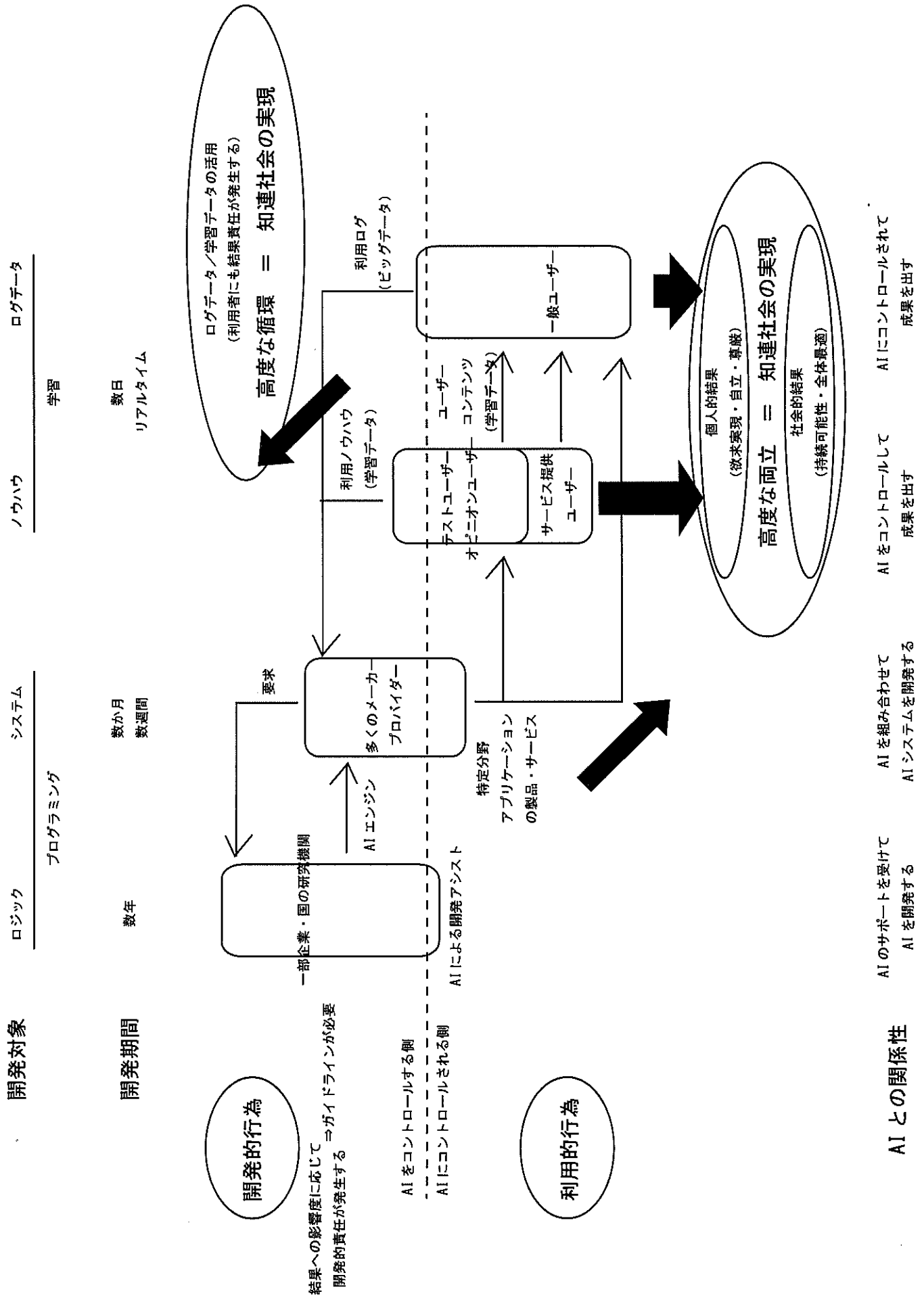
※智慧をあらゆる物事をスマートグリッドのように全体最適させる技術として考えている

※個人の欲望と人類の持続的豊かさ（快適さ）は、対立する場合があると考えている

◆修正意見ポイント

- ・ AI ネットワークシステムの構成要素について人間を含めたものにしてほしい
- ・ 開発・利用のモデルを循環的で複合的なものにしてほしい
- ・ 人間を、システムを利用し、支援を受け、結果の受益者になるだけの存在ではなく、システムの一部としての役割を持ち、故にガイドラインの様々な規範が必要であるという構成にしてほしい

以上



差出人: AIネットワーク(総務省情報通信政策研究所)  
宛先: [REDACTED]  
件名: FW: 【案件番号:145208852】 パブリックコメントに関する提出意見の配信(受付番号: 201701210000389160)  
日付: 2017年1月21日 14:05:42

---

-----Original Message-----

From: public-com-egov@e-gov.go.jp [mailto:public-com-egov@e-gov.go.jp]  
Sent: Saturday, January 21, 2017 2:05 PM  
To: ai.network@soumu.go.jp  
Subject: 【案件番号:145208852】 パブリックコメントに関する提出意見の配信(受付番号:201701210000389160)

パブリックコメントに関する意見提出先窓口担当者 様

電子政府の総合窓口( <http://www.e-gov.go.jp/> )から貴府省宛に  
パブリックコメントに関する意見が提出されましたので、配信します。

受付番号:201701210000389160  
受信日付:2017/01/21 13:41:46

案件番号:145208852  
案件名:  
「AI開発ガイドライン」(仮称)の策定に向けて整理した論点に関する意見募集  
宛先府省名:総務省

郵便番号:-

住所:

氏名:

連絡先電話番号:-

利用者メールアドレス [REDACTED]

提出意見:

- AIの定義づけに関して、自然言語は指し示す内容が流動的に変動して行くので、厳密な定義づけはせずに本質的な要件のみを策定するのはどうか(例えばいかなる部分が人間の延長として捉えられ、いかなる部分は人間と分けて考えなければならないか、など)

上記の点に関して個人的には人工知能は現状では形式論理で閉じた体系の内側を扱っているように感じられるようにも感じられる。自然言語の中にはアリストテレス的な形式論理で形式化できるものを扱っており、この外の論理には例えば直観主義論理、ファジー論理、アナロジーなど非形式論理(ここでの非は排中律的否定ではなく、バシュラー的な包含により拡張される否定)とも言える体系が広がっている。さらに、その外には言語で表現できない自然がある。そのため、AIが取り扱えるのは言語表現できるもののうちのさらに限定された一部ということになる。

- 汎用AIと特化型AIの思想分化について

これはちょうど俯瞰型と専門分化型の研究に対応しているように感じられるが、部分的にはねじれがあるように感じられる。というのも海外の論文を拝見していると、俯瞰型視野を持っている人がその足りない部分を埋めるために特化型AIを考えていて、専門分化型研究をしている人が同様に補足として汎用AIの開発を目指しているようにも感じられるからである。これを上流まで辿ると、ちょうどヘブライズムの思想分化、そしてそれが関連するところの政治思想の対立と対応するようにも感じられる。そのためにこの課題については一部政治思想、宗教多元性の研究が貢献するものとも感じられる。

- 開発原則に関して

とくに「いかなる場合がプライバシー侵害と感じられるか」ということに関する議論が必要であるように感じられる。例えば画像解析や音声解析の一部にはプライバシー法に抵触しかねないものが多くあるように感じられる。この点に関して、「狭い範囲における人間の厳密な分析(もしくは形式論理による厳密的解析)」には直感的には慎重さを要するよう感じられる。例えば所属集団(会社、国、文化)の分析は比較的形式的に大丈夫なもの、主観的範囲(人間の感情など何かしら無限性を備えたもの)に関して形式論理による厳密解析を目指すなどである。解析対象のサイズ感と厳密性の相性によって何かしら問題が生じるように感じる。

-----  
電子政府の総合窓口

<http://www.e-gov.go.jp/>  
-----

メール識別No:0000334035



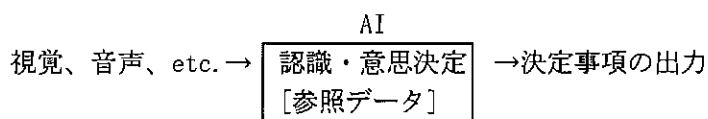
**「AI開発ガイドライン」(仮称)の策定に向けて整理した論点に関する意見募集  
提出様式**

平成 29 年 1 月 23 日

(ふりがな) 組織名 代表者氏名 役職	[Redacted]	組織名及び代表者氏名 の公表の可否
	[Redacted]	可
職業	大学教員	
(ふりがな) 住所	[Redacted]	
連絡先	(ふりがな) [Redacted] 担当者名: [Redacted] 役職: [Redacted] 電 話: [Redacted] F A X: [Redacted] 電子メールアドレス: [Redacted]	

ページ	該当部分	御意見	理由
別紙 1 P4	AI の定義	提案 AI の定義 視覚認識、音声認識、意思決定など、人間の知性を必要とする機能を具現化する認識決定システム。 Definition of AI A recognition decision system that realizes human intelligence functions such as visual recognition, speech recognition, decision making.	○議論の出発点を明確にする必要があると考え、AI の定義を提案します。 ○定義に含まれる「人間の知性」は、「知識や技能を習得し、適用する能力」を示します。 ○入力、処理、出力のモデルで示すと下図のようになります。

入力 → 処理 → 出力



差出人: AIネットワーク(総務省情報通信政策研究所)  
宛先: [REDACTED]  
件名: FW: 【案件番号:145208852】パブリックコメントに関する提出意見の配信(受付番号: 201701300000391480)  
日付: 2017年1月30日 14:05:40

---

-----Original Message-----

From: public-com-egov@e-gov.go.jp [mailto:public-com-egov@e-gov.go.jp]  
Sent: Monday, January 30, 2017 2:05 PM  
To: ai.network@soumu.go.jp  
Subject: 【案件番号:145208852】パブリックコメントに関する提出意見の配信(受付番号:201701300000391480)

パブリックコメントに関する意見提出先窓口担当者様

電子政府の総合窓口(<http://www.e-gov.go.jp/>)から貴府省宛にパブリックコメントに関する意見が提出されましたので、配信します。

受付番号:201701300000391480  
受信日付:2017/01/30 13:20:53

案件番号:145208852  
案件名:  
「AI開発ガイドライン」(仮称)の策定に向けて整理した論点に関する意見募集  
宛先府省名:総務省

郵便番号  
住所  
氏名  
連絡先電話番号  
利用者メールアドレス

提出意見:  
ページ 該当部分

7Page 体系

意見 ガイドラインの体系として、3パターンが提示されているが、AI開発の社会的影響度に鑑み、開発実務の面から考えると、この体系とは別に、開発マネジメント、開発保証のガイドラインが必要ではないか。

理由 AI開発組織の開発標準、法令遵守度等の成熟度(CMMI)が低い状況での開発は重大な潜在不具合を除去できない。

13Page 目的

意見 ガイドラインの目的が抽象的。AI開発ではなく、AIシステムに対する目的となっている。開発前の企画段階で考慮すべき内容とか、基本的な考え方だけのようと思われる。

開発者から見て、開発ガイドラインとは呼べない内容である。

理由 経済産業省から過去に出されたヘルスソフトウェア開発ガイドライン等に比べ、誰のためのガイドラインなのか曖昧。各社でAI開発標準を整備しようとしても使えない。

27-29Page 透明性

意見 説明責任の面から透明性は必要であるが、どのようなデータを与え、どんなアルゴリズムで学習させた結果、そうなったのかということについて、詳細資料を公表することは、現実的には不可能。AIアルゴリズムについては、OSSのよう、その開発者の著作権や、取扱ルールを明確にすれば良い。データについては、その入手先を説明できれば良いのではないか。

そして、AI開発の場合、どのような検証を実施して、制御が逸脱するリスクに対応しているのか、AI開発の社会的影響度に応じた説明責任が求められると考えられる。たとえば医療系であれば、IEC62304で規定されている3つのクラスのように、その影響度合いに応じた検証結果の審査も必要になる。

理由 AIの開発には、従来のソフトウェアのようにアルゴリズムの開発部分と、膨大なデータ投入による学習部分がある。それぞれに透明性が求められる。

また、機能安全や医療に関連する分野では、それぞれの安全規格が要求する開発プロセスやドキュメントを提出することになっている。また

航空宇宙(NASAやJAXA)ではIV&V(独立検証)組織があり、開発と独立して検証し、開発の妥当性を承認している。

30-32Page 制御可能性

意見 制御が逸脱するリスクに関して、透明性で述べたように、その社会的影響度に応じて逸脱検証を実施し、必要とあればいつでもその検証結果をもって説明できるようにする。

理由 先述の理由と同じ。

高度な品質を要求される検証には、ディペンダビリティ確保としての検証計画、検証設計・実施結果報告を実施している。

-----  
電子政府の総合窓口

<http://www.e-gov.go.jp/>  
-----

メール識別No:0000336499

**「AI開発ガイドライン」(仮称)の策定に向けて整理した論点に関する意見募集  
提出様式**

平成 29 年 1 月 29 日

(ふりがな) 組織名 代表者氏名 役職	じんこうちのうがっかいりんりいいんかい 人工知能学会倫理委員会 松尾 豊 委員長	組織名及び代表者氏名 の公表の可否  可
職業	東京大学大学院工学系研究科総合研究機構／知の構造化センター／技術経営戦略学専攻 准教授	
(ふりがな) 住所	[Redacted]	
連絡先	(ふりがな) [Redacted] 担当者名： [Redacted] 役職： [Redacted]  電 話： [Redacted] F A X： [Redacted] 電子メールアドレス： [Redacted]	

ページ	該当部分	御意見	理由
P29 第七(1) 透明性の 原則(3 /3)	2. 動作の透明性(検証可能性及び説明可能性)が要請されるAIの動作の範囲如何。入出力、通信及び判断としてよいか。	機械学習における深層学習のメカニズムは、少なくとも研究者にとってはブラックボックスではなく、研究者が何を入力して、どう学習させたのかが理解可能であれば、入出力、通信及び判断の透明性は確保されている、という解釈をして良いのか、という点を明示することを要望いたします。	人工知能の研究開発において、「透明性」の定義が曖昧なために、その研究開発を進めることに疑義が生じることを回避したいためです。

差出人: AIネットワーク(総務省情報通信政策研究所)  
宛先: [REDACTED]  
件名: FW: 【ウイルス感染注意】『AI開発ガイドライン』(仮称)の策定に向けて整理した論点に関する 意見  
日付: 2017年1月31日 11:22:59

---

-----Original Message-----

From: [REDACTED]  
Sent: Tuesday, January 31, 2017 11:22 AM  
To: ai.network@soumu.go.jp  
Subject: 【ウイルス感染注意】『AI開発ガイドライン』(仮称)の策定に向けて整理した論点に関する 意見

\*\*\*\*\*

!!!注意!!!

このメールは、フリーメールアドレスから送信されており、ウイルスを含む不審なメールである可能性があります。  
ウイルスに感染しないよう、「差出人欄」、「件名」、「本文末の署名」等にご注意し、不審なメールでないことを十分に御確認の上、添付ファイルの開封やリンク先の閲覧を行ってください。

【問い合わせ先】大臣官房企画課情報システム室(03-5253-5159)

\*\*\*\*\*

ご担当者様

[REDACTED] といいます。個人での応募になります。

匿名希望です。

AIガイドラインで申したいことは以下のとおりです。主にセキュリティに関してです。

■主にP.34

リスク評価もセキュリティバイデザインでも、機密性、完全性、可用性とあるが、機密性、完全性、可用性に対するリスクだけではなく、これらの3要素以外に「責任追跡性」(Accountability)、「真性性」(Authenticity)、「信頼性」(Reliability)を加えた情報セキュリティの6要素の観点でみるべきと考えます。

[(イ)利用者及び第三者の生命・身体の安全に危害を及ぼす可能性のあるセキュリティ上の脅威・脆弱性への対処

→これも詳細にかくとすれば、AIのソフトウェアの観点と、機械やロボットが過剰な動きをしないということをアプリケーションレイヤーと物理レイヤーの両方で制御できるようなものとするべきと考えます。

セキュリティの脆弱性を指摘された際、あくまで対象となるAIのアプリケーションのみが修正対象となるようにする。つまり、アプリより下のレイヤーにおける脆弱性がアプリケーションにまで影響

を及ぼすようなことはあってはならない。既存のOSやミドルウェアの脆弱性を突かれてAIがテイクダウンされてしまうようなことにはならないようにするべき。AIアプリケーションの脆弱性に対する診断を人手で行うことはなく、AI自身が診断や自己修復できるようにあるべきと考えます。

以上です。

**「AI開発ガイドライン」(仮称)の策定に向けて整理した論点に関する意見募集  
提出様式**

平成29年 1月31日

(ふりがな) 組織名 代表者氏名 役職		組織名及び代表者氏名の公表の可否
職業	セキュリティエンジニア	
(ふりがな) 住所		
連絡先	(ふりがな) ██████████ 担当者名: ██████████ 役職: - 電話: ██████████ FAX: ██████████ 電子メールアドレス: ██████████	

ページ	該当部分	御意見	理由
13	第三者や社会ないし人類への波及的な悪影響を抑制すべき	システムの脆弱性検出を目的としたAIは「疑似攻撃」を行うが、使い方次第では現実の攻撃に悪用される可能性がある。 上記のような「攻撃目的のAI」の扱いを明確にした方がよいのではないか。	本規定が無い場合、脆弱性検出の自動化を実現するAIの開発が阻害されるおそれがあるため。
33	(7)情報セキュリティの3要素(機密性、完全性、可用性)の確保	情報セキュリティの3要素以外の要素「真正性」「責任追跡性」の確保も必要ではないか。 ※信頼性はP34で提案されている。	インシデント防止の観点から必須であるため。
33	利用者及び第三者の生命・身体の安全に危害が及ぶリスクの評価	リスク評価の方法をある程度明確にすべきではないか。 例) 外部監査の義務化、または、セルフチェックの義務化(セルフの場合は別途チェックリストの策定が必要)	現時点でリスク評価が行える組織は存在しないため、今後の産業化を期待し、ガイドラインに明記した方が良い。

33	セキュリティの設計及び実装(セキュリティ・バイ・デザイン)	AI が学習に利用する(インターネット上に公開されている)学習データ群の安全対策も必要ではないか。 例) 学習データ汚染による、AI の意図しない動作。	教師あり学習型の AI にとって学習データは重要な意味を持つため。
33	利用者及び第三者の生命・身体の安全に危害を及ぼす可能性	左記リスクのみではなく、「金融被害」も含めた方が良いのではないか。	現在のサイバー攻撃において、金融被害も深刻化しているため。
34	第2項	インシデント後の追跡調査を容易にするため、各種ログの保全を義務化すべきではないか。	ログが欠損、または、存在しない場合、追跡調査が困難になるため。



**「AI開発ガイドライン」(仮称)の策定に向けて整理した論点に関する意見募集  
提出様式**

平成 29 年 1 月 31 日

(ふりがな) 組織名 代表者氏名 役職	[Redacted]	組織名及び代表者氏名 の公表の可否
	[Redacted]	可
職業	経営者	
(ふりがな) 住所	[Redacted]	
連絡先	(ふりがな) [Redacted] 担当者名: [Redacted] 役職: [Redacted] 電 話: [Redacted] F A X : [Redacted] 電子メールアドレス: [Redacted]	

ページ	該当部分	御意見	理由
全体		本ガイドラインは、日本発イノベーションを強く阻害するものであり、イノベーションに強く注力する弊社としては、今後の日本での研究開発活動、事業活動において大きな懸念を抱くものである。	
4	第一 基本概念的定義	AI、AI ネットワークシステム、AI ネットワークサービスの定義を明確にするべきである。	AI という定義がされていない状態で、AI ネットワークシステムを定義する場合、「AI を構成要素」という必要条件・十分条件を確認できないため、AI ネットワークシステムだと認識することが不可能となるため。

5	論点 1, 2, 3	<p>人的主体についてより詳細に定義を行うべきである。</p> <p>研究者と開発者を分離すべきである。</p> <p>また、開発者と提供者を分離すべきである。</p>	<p>研究者は、公共の福祉に反しない限り、学問の自由で守られるべきである。</p> <p>AI ネットワークシステム、AI ネットワークサービスの”機能”と”活用”の責務を分離すべきであり、開発者が顔から人種を推定する AI システムを開発した場合、善意を持ったマーケティングとして活用することも可能だが、悪意を持った提供者が特定の人種に対して制限をかけるシステムを提供することも可能となる。よって、これらの責務は分断されるべきである。</p>
17	論点 3	<p>閉鎖された空間における定義を明確にし、クラウドコンピューティング時代に即した研究開発体制を考慮すべきである。具体的には、仮想閉域網などで論理的に閉鎖された空間をこれらの定義に追加すべきである。</p>	<p>閉鎖された空間を「外界への影響及び外界からの影響のいずれについても遮断することができるよう措置が講ぜられている実験室等」と定義しているが、大規模な計算を必要とする現代の AI 研究開発において、このような物理的な隔離状態は現実的で無い。大規模なクラウドコンピューティングパワーを活用しなければ、日本のイノベーションは明らかに阻害される。</p>
29	論点 1	<p>透明性がイノベーションの阻害になることが無いよう、より詳細な議論を行うべきである。</p>	<p>現代の AI において、ディープラーニング手法は判断過程がブラックボックスとなり、全ての透明性の確保を担保することが難しい。しかし、世界においては、ディープラーニングの技術成果を実用に進められていることが当然であり、この透明性の担保が日本のイノベーションを阻害することが無いよう強く主張する。</p>
36	論点 3	<p>AI、AI ネットワークシステム、AI ネットワークサービスの全てにおいて設計・開発段階からのセキュリティ・バイ・デザインを担保することは不可能であり、本件は利用者を主語とする制限とすべきである。</p>	<p>全ての科学技術は諸刃の剣であり、刃物・自動車などはその最たる例である。設計・開発段階において、人を殺傷する能力があることは明らかであるが、現代において、これらは利用者の責任を前提としている。AI、AI ネットワークシステム、AI ネットワークサービスにおいても同様とすることが</p>

			妥当である。
38	論点 5	AI、AI ネットワークシステム、AI ネットワークサービスの全てにおいて設計・開発段階からのプライバシー・バイ・デザインを担保することは不可能であり、本件は利用者を主語とする制限とすべきである。	全ての科学技術は諸刃の剣であり、カメラ、マイクなどはその最たる例である。設計・開発段階において、プライバシーを侵害する能力があることは明らかであるが、現代において、これらは利用者の責任を前提としている。AI、AI ネットワークシステム、AI ネットワークサービスにおいても同様とすることが妥当である。

**「AI開発ガイドライン」(仮称)の策定に向けて整理した論点に関する意見募集  
提出様式**

平成 29 年 1 月 31 日

(ふりがな) 組織名 代表者氏名 役職		組織名及び代表者氏名 の公表の可否
職業	AI エンジニア ( ██████████ ) にて勤務)	
(ふりがな) 住所	██████████ ██████████	
連絡先	(ふりがな) ██████████ 担当者名: ██████████  電 話: ██████████ F A X : ██████████ 電子メールアドレス: ██████████	

ページ	該当部分	御意見	理由
19	開発原則の構成及び順序	19 ページで触れられている小林構成員の意見は、ヨーロッパにおける規制との整合を図る趣旨で重要である。 倫理系の項目（プライバシー保護の原則、倫理の原則）をより上位に配置してはどうか。  利用者支援の原則は倫理系の原則の次に重要であり、倫理系の原則より利用者の利便性が優先されることがあってはならない。	現在、個人情報保護法制の枠組みとしては、ヨーロッパにおける各種規制と日本の規制の整合が取れていない結果、ヨーロッパ圏の個人情報を日本にて取り扱うことが困難となっている。このようなことを、AI の開発において繰り返してはならない。
27	透明性の原則	透明性を定義する上では、深層学習においては以下のような特性があることに配慮されたい。  深層学習は、以下の点において原則として再現性があり、それは透明性の依って立つ技術的根	

		<p>拠となる。</p> <p>○既存のネットワークに対する推論において、ネットワーク構造および素子のパラメータが同一である複数のネットワークに同じのデータを入力すれば、同じ結果が得られる。また、同一のネットワークに同じデータを複数回入力すれば、同じ結果が得られる。</p> <p>一方、以下の点において透明性は必ずしも得られない。</p> <p>○既存のネットワークに対して、同じデータセット、同じハイパーパラメータで複数回学習を行っても、乱数の種が同一でなければ、結果は同様とは限らない</p>	
28	動作の説明可能性の確保について	<p>ネットワーク（パラメータ含む）、入力、出力をログとして保管するとしても、その入力には例えば以下の問題がある。</p> <p>○プライバシー関連を含む何らかのセキュリティ法や規則に抵触するログが出力される可能性がある。そのようなログを保管することについて、AI であるという理由だけで社会の理解が得られるとは考えられない。日本の個人情報セキュリティレベルを欧米と整合しなければ、AI のガイドラインの整合もおそらく不可能であり、広い視野での検討を望む。</p> <p>○一方、社会との合意としてAI の入力としてはならない性質のデータも存在すると考えられ、そのバランスはおそらく具体的事案に照らして検討されるべきである。</p>	
30	制御可能性の原則	<p>AI による自動制御と人間による非自動制御の交点には、常に問題が生じうる。</p> <p>現在一般投入の直前のフェーズにあると思われる自動運転においても、当然同様である。</p> <p>制御可能性については、技術や法のみならず哲学・思想の観点からの議論も取り込んでほしい。</p>	
33	セキュリティ確保の原則	<p>現在日本でしばしば取られている説明責任や第三者認証を最優先するセキュリティではなく、AI の関わる概念において有効な、具体的なセキュリティ施策の検討を進めてほしい。</p>	
37	プライバシー保護の原則	<p>本項は、日本において現在独自の検討をすべきとは考えない。</p> <p>まずは例えば個人情報においてEU データ保護規則との十分性を得るなど、国際的な規制との整合を取ってから議論すべき。EU もしくは米国の規制を丸呑みすることも視野に入れるべき。</p>	

		<p>さもなければ、日本の AI はプライバシーが理由で世界に出て行けなくなる。</p>	
39	倫理の原則	<p>39 ページに記載されている河島構成員の意見に賛成するが、現実には可能なのか。</p> <p>日本で「りんな」として知られる AI システムが米国の SNS に解き放たれた結果、半日にして差別主義者となった事案が象徴的。</p> <p>現時点で AI に倫理を実装することはきわめて困難。</p> <p>ガイドラインで縛るよりも、AI に倫理を実装するために必要な議論、さらには市民の倫理観の向上を一体として図るべき。</p> <p>東ロボくん研究が東大入試突破を後回しにして取り組むことになった「人間の理解力」の問題こそが、本件の重要な核となる。</p>	
43	アカウントビリティの原則	<p>説明責任と訳さないことについて、全面的に賛同する。</p> <p>いわゆる「説明責任」ではなく、アカウントビリティは「説明できないことをしない」原則。</p> <p>とはいえ、現時点での AI の基礎研究にアカウントビリティを問うことはおそらく困難。例えば核物理の発見者に対して原爆、そして広島市民に対する大虐殺についてアカウントビリティを問うのは難しい。</p> <p>技術を含めた専門家の合意や感覚も重要ではあるが、それ以外に AI 研究者のコミュニティに対する信頼と相互牽制をベースとした、論理的ではない納得感をベースとした対市民アカウントビリティの可能性について考えてほしい。</p> <p>そして、それは産学官・市民の連携でこそ成り立ちうるものとする。</p>	
48	開発原則の実効性の確保における市場の活用の在り方	<p>BtoB の発注においては、発注側がこれらの原則を踏まえておくことは可能と考える。</p> <p>ただ、担保のために第三者認証を持ち込むアプローチは確かに現代日本でしばしば取られているものであるが、現時点できわめて難しいのでは。その認証を担当できる産業界 AI の専門家は少なくとも日本には十分に育っていない。無理に実行すると ISMS のように形式的に認証を取られ、その結果ビジネスや研究を阻害する結果にしかならないと考える。</p> <p>公平な第三者（おそらく官学連携か？）による基準に照らした自己表明を監査する、という形</p>	

		<p>が高々可能な範囲だろう。</p> <p>BtoCにおいて、一般市民がこれらの原則を理解してAIを選択することはおそらく不可能。リスク減免の仕組みは、保険制度の全体的な設計に関わると考える。</p> <p>例えば明らかに公道を走れる性能を持たないAIを搭載した自動運転車両が公道で事故を起こし、被害が発生したとき、誰の保険で誰を救済するか。(被害者の保険を使う、という可能性も制度設計としては十分にありうる。その場合被害者の保険で被害をいったん救済した後、保険がAI管理者に請求する形になるはず。そこでBtoBの交渉に収斂する)</p> <p>論点5について、同意。日本においても至急法規制を検討すべき分野は確実に存在する。</p> <p>ただ、何を法規制するかの検討においては、AI開発現場の声を十分に聞いてから議論してほしい。政治や行政が認識しているプレイヤーは、AI界隈のごく一部でしかない。官庁のAI系懇談会や委員会には民間企業としてPFNがしばしば参加しているが、例えばGoogleも日本国内のプレイヤーだし、受託でBtoBのAI活用を支援するビジネスを展開する企業もISP(システム計画研究所)をはじめとして国内に数社存在する。それらの企業の声聞かずに成り立たないはずである。</p>	
56	AIネットワークシステムの利活用に関し利用者等が留意すべき事項	<p>林構成員：AIネットワークシステムを流通させるデータ形式の標準化に関する意見について。</p> <p>主な深層学習フレームワークを取ってみても、容易に使えるデータ形式が全てバラバラである、という現実がある。</p> <p>例えば最初の一步として、何らかのgithubレポジトリにBVLC(Caffe)、Google(TensorFlow)、PFN(Chainer)の三者で共通のデータ構造をえるようにする一連のフレームワークを整備する、といった活動であればすぐにでも実現できるはず。</p>	
53	AIネットワークシステムの利活用に関し利用者等が留意すべき事項	<p>福井構成員：オープンサイエンシ的な理念</p> <p>現状、一部のデータセットは公的に認められている研究組織からしかアクセスできない。企業の営利活動に使うのであればコストを負担せよという趣旨は理解するが、国内でこの手の仕掛けを導入していると組織に所属していない野</p>	

	項	<p>良研究者がアクセスできないことが世の常である。</p> <p>AWS 等のクラウドを活用することで、AI 研究は必ずしも設備投資を必要としない状況になっている。野良研究者に対する審査（＝営利目的ではないことの確認、研究スキルと倫理を有することの確認）は必要かもしれないが、例えばニコニコ学会等で活動するような野良研究者への配慮も望む。</p>	
--	---	--	--



**「AI開発ガイドライン」(仮称)の策定に向けて整理した論点に関する意見募集  
提出様式**

平成29年1月31日

(ふりがな) 組織名 代表者氏名 役職	[Redacted]	組織名及び代表者氏名 の公表の可否
	[Redacted]	
職業	弁護士	
(ふりがな) 住所	[Redacted]	
連絡先	(ふりがな) [Redacted] 担当者名: [Redacted] 役職: [Redacted] 電 話: [Redacted] F A X : [Redacted] 電子メールアドレス: [Redacted]	

ページ	該当部分	御意見	理由
		別紙のとおり	

(別紙)

「AI開発ガイドライン」(仮称)の策定に向けた国際的議論の用に供する素案の作成に関する論点に対する意見

平成29年1月31日

上記論点についての意見は次のとおりである。

#### 第1 基本概念の定義について

##### 1 5頁・論点1について

意見 人工知能の定義は設けるべきである。

理由 「AI開発ガイドライン」(仮称)(以下「本ガイドライン」という。)が、AIの研究開発に適用されるものであることからすると、AIの定義がない場合には研究開発の外延が不明であり、萎縮的効果を及ぼす可能性がある。

なお、定義を行う際には、「人工知能」の定義が論者にその定義が区々であることから、ロボット法領域における定義の方法を参考に、機能に着目する形で行うべきである。

##### 2 5頁・論点3について

意見 「AIを研究し又は開発する行為」の範囲については、上で述べたとおり、機能面からAIを定義したうえで、明らかにすべきである。

理由 上記第1項を参照されたい。

##### 3 6頁・論点5について

意見 利活用ガイドラインについては、「利用者」「最終利用者」のうち、事業者を名宛人とすべきである。

理由 利活用ガイドラインについては、「利用者」「最終利用者」について、事業者か一般消費者かの区別がないため、一般消費者に対しても利活用ガイドラインの規制が及ぶことのないようにする必要がある。

#### 第2 AI開発ガイドラインの体系について

##### 7頁・論点2について

意見 連携の原則を追加した方が良い。

理由 AIネットワークシステム全体について、開発ガイドラインの効力が及ぶべきである。

#### 第3 分野共通開発ガイドラインの構成について

##### 8頁・分野別共通開発ガイドラインの構成(たたき台)について

意見 開発原則の構成及び順序については、後記第6・第1項の意見を踏まえて、倫理の原則を一番上にすべきである。

理由 後記第6・第1項を参照されたい。

#### 第4 分野共通開発ガイドラインの目的、基本理念等について

##### 14頁・論点7について

意見 政府の役割についても盛り込むべきであり、国際的な協調についても言及すべきである。

理由 本ガイドラインが国際的なものとして実現されることを前提とするものであることから、国際的な強調を継続していくことについても言及すべきである。

## 第5 分野共通開発ガイドラインの適用範囲について

### 16頁・論点2について

意見 機能により限定すべきである。また、情報通信ネットワークシステムを実装又は接続することが技術的に可能かという基準で、適用されるガイドラインの内容が異なるものとし、適用対象とするのか否かを画すべきではない。

理由 適用範囲の限定は必要であるが、本ガイドライン自体があらゆる開発者に参照されるか不確実な場合が考えられ、情報通信ネットワークに実装し又は接続することが技術的に可能か、という要件について誤解がある場合には、有害なAIが伝播するリスクがあり、かつその影響の大きさは、情報通信ネットワークの広がり国際的なものであることからすると、やむを得ない。適用範囲は、AIの機能の範囲で画すべきである。

## 第6 開発原則の構成及び順序

### 120頁・論点1について

意見 倫理の原則が最上位とされるべきである。

理由 分野共通開発ガイドラインの目的、基本理念等で検討された分野共通ガイドラインの目的案でも前提とされている智連社会は、AIネットワーク化が進んだ場合における社会のあるべき姿を描くものであるが、なぜ智連社会が人間が中心となる社会像と理解されるかについては、近代立憲主義の根底をなし、また我が国の憲法13条に基づく個人の尊厳が背景にあると考えられる。

そのため、開発原則の順序については、なによりもまず、人間の尊厳と自律性を尊重することを求める倫理の原則が最上位に掲げられるべきである。

倫理の原則は、単に「AIネットワークシステムのリスクの抑制に関連する原則」にとどまるのではなく、「AIネットワーク化の健全な進展の促進及びAIネットワークシステムの便益の増進に関連する原則」よりも上位の概念、なぜAIネットワーク化に「健全な進展」が求められるのかを説明するものであって、あえて言うならば、AIのあるべき方向を定める原則である。

そして、かかる倫理の原則によりAIに人間の尊厳と自律性を尊重することが求められるからこそ、人間の尊厳と自律性に悪影響を及ぼす場合の制御可能性や、人間の尊厳と自律性を尊重するための利用者支援が求められるのであって、倫理の原則が連携の原則よりも最上位とされるべきである。

本ガイドラインが、AIネットワーク化を念頭に置くものであったとしても、ネットワーク化されない個々のAIによっても、人間の尊重と自律性に悪影響を及ぼされる可能性があり、ネットワーク全体に対してのみガイドラインで規制を行うのでは足りない。本ガイドラインが部分を構成するAI単体の便益及びリスクのみに着目するものではなかったとしても、AI単体が人間の尊厳と自律性を尊重するものでなければ、ネットワーク化した際に、人間の尊厳と自律性が尊重されるという状態を実現することは不可能である。

### 220頁・論点3について

意見 倫理の原則は、主にリスクの抑制に関連する原則として整理されるものではなく、AI開発の全体において、どのような開発が行われるべきか、AI及びAIネットワーク化の「健全な発展」を要請する指針であり、他の原則よりも上位に置かれるべきものである。

理由 上記第6・第1項の理由を参照されたい。

### 320頁・論点4について

意見 開発原則の構成及び順序は、以下のとおりとすべきである。

①主にAIネットワーク化の健全な進展の促進及びAIネットワークの便益の増進に関連する原則  
倫理の原則>連携の原則

②主にA Iネットワークシステムのリスクの抑制に関連する原則

透明性の原則、制御可能性の原則、セキュリティ確保の原則、安全保護の原則、プライバシー保護の原則

③A Iネットワーク化の健全な進展の促進等及びA Iネットワークシステムのリスクの抑制のいずれにも関連する原則

利用者支援の原則、アカウントビリティの原則

理由 上記第6・第1項の理由を参照されたい。

4 21頁・論点5について

意見 別案が23頁に掲げる順が最適ではないため、別案としては第6・第4項に記載の順にすべきである。

理由

	長所	短所
第6・第4項に記載の順(別案) ①倫理の原則 ②連携の原則 ③透明性の原則 ④制御可能性の原則 …	物事には究極的に目指すべき価値があり、人間の尊厳と個人の自律性の尊重を求める倫理の原則が、A I開発においては目指されるべき価値であること、その次に連携の原則を置くことで、このガイドラインがA Iネットワーク化を念頭に置くものであることが顕著に示されること、分野別共通開発ガイドラインの目的規定の記載順は、「健全な進展の促進」が先に記載されているが、倫理の原則はその理由・必要性を明らかにするものであり、同目的の記述順と整合すること、若干の原則を例示する際に、記載順に即して例示すれば、「倫理の原則、透明性の原則等」「倫理の原則、透明性の原則、制御可能性の原則等」となり、双方から例示の対象がごく自然に選ばれるため、開発原則の性格がバランスよく描写されること	このガイドラインがA Iネットワーク化を念頭に置くものであることが若干顕著には示されないこと

5 21頁・論点6について

意見 開発原則の項目相互間の優先順位又は調整に関し、別段の規定を設けるべきである。

理由 開発原則の項目の中でも、倫理の原則のように、他の原則よりも優先されるべきものがあり、他方で、項目相互間で優先順位がない(同列)のものも考えられるため、必要な範囲においてのみ、優先順位又は調整の規定を設けるべきである。

第7 開発原則の個々の項目の内容の具体化について

1 (1) 透明性の原則について

(1) 21頁・論点1について

意見 透明性が必要であるが、その内容としては、どのような情報が入力された場合に、どのような出力が想定されるかという動作の説明可能性を中心とすべきであり、動作の検証可能性は、使用される技術に応じて技術的に検証不可能又は困難なものについては、入出力及び通信のログに限定し、制御可能性の問題として取り扱うべきである。

理由 透明性の原則は、動作の検証可能性の確保と動作の説明可能性の確保から構成されるが、深層学習技術や、既存の技術の複合化あるいは新技術により、動作の検証可能性に限界が出てくる場合(ブラックボックス化)があり、透明性が求められる程度は、技術に応じて変化するものである。

そして、透明性を求めるあまり、AIの開発が過度に抑制されないようにするべきである。

ブラックボックス化については、リスク低減及びダメージの低減を図る制御可能性の問題として解消すべきと思われる。

(2) 29頁・論点2について

意見 動作の透明性について、判断については、使用されている技術に応じたものにすべきである。

理由 判断についても透明性が求められると、深層学習を用いたAI開発が不可能又は困難となる可能性がある。

2 (2) 制御可能性の原則について

(1) 32頁・論点1について

意見 予め制御可能性の検証を行うことが必要とすることは良いが、妥当性確認については市場原理に委ねるべきである。ただし、個人の生命・身体の安全灯重要な権利利益若しくは法益に関するリスクを惹起し得る、又は個人に関する重大な決定のために利活用されるAIネットワークシステムの構成要素となり得るAIについては、妥当性確認を行うことを必要としても良い。

理由 妥当性の確認を求めるレベルによっては、開発者に対して過度の負担を強いることになるため、契約責任における品質保証等で対応すべきである。

3 (4) 安全保護の原則について

(4) 36頁・論点4について

意見 営業秘密との関係性に留意すべき

理由

4 (6) 倫理の原則について

(1) 40頁・論点1について

意見 将来世代に対する配慮については、

理由 将来世代に対する配慮については、ある一定の将来予測を伴うものであり、困難と思われる。

(2) 40頁・論点2について

意見 AIネットワークシステムだけではなく、AI単体も人間性の価値を毀損してはならない旨を定めるべきである。

理由 AIネットワークシステムだけでなく、AI単体にも、人間性の価値を中心に据え、人間の尊厳と個人の自律性を尊重すべきことが求められる。

5 (7) 利用者支援の原則について

(1) 42頁・論点1について

意見 利用者がAI単体及びAIネットワークシステムを利用する場合には、その旨が事前に告知されるべきである。

理由 利用者が適時適切に判断を行うためには、その情報がAIによってもたらされているものかどうかを知る必要がある。

6 (8) アカウンタビリティの原則について

(1) 44頁・論点1について

意見 利用者がAI単体及びAIネットワークシステムを利用する場合には、その旨が事前に告知されるべきである。

理由 利用者が適時適切に判断を行うためには、その情報がAIによってもたらされているものか

どうかを知る必要がある。

(2) 44頁・論点2について

意見 利用者の信頼・期待の保護については、利活用ガイドラインの名宛人を事業者に限定した上で、市場原理に委ねるべきである。

理由 利活用ガイドラインについては、「利用者」「最終利用者」について、事業者か一般消費者かの区別がないため、一般消費者に対しても利活用ガイドラインの規制が及ぶことのないようにする必要があり、その上でAIネットワークシステムに対する信頼・期待の保護については、契約責任上の品質保証等の問題として市場原理に委ねるべきである。

第8 連携の原則（仮称）について

47頁・論点3について

意見 利活用ガイドラインについては、「利用者」「最終利用者」のうち、事業者を名宛人とすべきである。

理由 利活用ガイドラインについては、「利用者」「最終利用者」について、事業者か一般消費者かの区別がないため、一般消費者に対しても利活用ガイドラインの規制が及ぶことのないようにする必要があり、その上でAIネットワークシステムに対する信頼・期待の保護については、契約責任上の品質保証等の問題として市場原理に委ねるべきである。

第10 開発原則の実効性の確保における市場の活用の在り方について

1 51頁・論点1について

意見 開発原則に適合するAIと適合しないAIが併存することを前提してガイドラインを策定すべきである。

理由 利用者の選択の自由を一律に否定すべきではない。

2 51頁・論点2について

意見 公的認証

理由 また、AIの信頼性については、単に技術のみならず、その設計思想として、人間の尊厳、個人の自律性を踏まえたものになっているかが検証されるべきである。

第11 51頁・AIネットワークシステムの利活用に関し利用者等が留意すべき事項について

意見 意見 利活用ガイドラインについては、「利用者」「最終利用者」のうち、事業者を名宛人とすべきである。

理由 上記第1第3項を参照されたい。

以上

**「AI開発ガイドライン」(仮称)の策定に向けて整理した論点に関する意見募集  
提出様式**

平成 29 年 1 月 31 日

(ふりがな)	すうりせんたんぎじゅつけんきゅうしょ	組織名及び代表者氏名の公表の可否
組織名	株式会社 数理先端技術研究所	可
代表者氏名 役職	生島高裕 代表取締役	
職業	ソフトウェア開発	
(ふりがな) 住所	[Redacted]	
連絡先	(ふりがな) [Redacted] 担当者名: [Redacted] 役職: [Redacted] 電 話: [Redacted] F A X: [Redacted] 電子メールアドレス: [Redacted]	

ページ	該当部分	御意見	理由
別紙 1. P30	第七(1) 透 明性の原則 (3/3)	「技術的及び経済的な事情に鑑み合理的な範囲・水準で動作の説明可能性を確保するよう努めるべきとしてはどうか。」の次に、「及び、上記の検証可能性、説明可能性についての具体的なアルゴリズムが拡張された場合、過去のバージョンとの違いを検証可能性、説明可能性について、技術的及び経済的な事情に鑑み合理的な範囲・水準で説明し、このバージョン履歴をデータとして保存、オープン化するべきとしてはどうか。」を追加してはどうでしょうか。	技術の進歩が速いため、検証、説明機能のバージョンアップも頻繁になる。ため、この機能拡張の説明を求めその履歴をとり、オープン化する必要がある。
別紙 2. P4	第七 開発原則の個々の項目の内容の具体化	同上	同上

**「AI開発ガイドライン」（仮称）の策定に向けて整理した論点に関する意見募集  
提出様式**

平成 29 年 1 月 31 日

(ふりがな) 組織名 代表者氏名 役職	情報法制研究所 AI 問題タスクフォース	組織名及び代表者氏名 の公表の可否
	鳥海不二夫 代表	可
職業	東京大学大学院工学系研究科准教授	
(ふりがな) 住所	[Redacted]	
連絡先	(ふりがな) [Redacted] 担当者名： [Redacted] 役職： 電 話： [Redacted] F A X： 電子メールアドレス： [Redacted]	

1 ガイドラインの法体系上の位置づけについて

本ガイドラインの法体系上の位置づけ、法的拘束力の有無及び法執行力の有無について明らかにされたい。すなわち、一般的な意見聴取手続とは、法律の内容を具体化する行政府の行う法執行に民主的統制を及ぼす観点から実施されるものであるが、本ガイドラインの前提となる法律が不在であり、総務省がどのような法律ないし法規範に基づいて制定するのか、社会的制度としての位置づけについて、これをガイドラインの中で明記されたい。

2 ガイドラインの執行について

本ガイドラインが仮に法律あるいは法規範の裏付けがない単なる政策意思の表明である場合は、各ステークホルダーに本ガイドラインを遵守するインセンティブをどのような形で保障することを計画しているのかについてガイドラインの中で具体的に示されたい。

3 開発原則において議論すべき AI の範囲について

AI には大まかに分けて、ある特定の目的を達成するために道具的な嗜好性を持つ非自律型の AI（近時インターネット業界で話題性の高い自然言語処理等を入れたチャットボットはこれの典型である。）と、自らのコードすら書き換えてしまう可能性を持つ自律的な AI（自ら



思考し変容していく可能性を持つ知能としての AI) とが分類として存在する。この両者は、似て非なるものであり、その社会的な便益やリスク、法的・倫理的な規制の必要性に関して全く性質が異なるものであるから、ガイドライン構築の初期の段階から明確に峻別して議論をすることが適切であるが、本ガイドラインはそのどちらを重視して構成されたのか、または他の類型に基づく整理がなされているのか、文書内で明示されたい。

#### 4 非自律的な AI に適用する原則への改訂について

一般に、自律的な AI については、いまだ具体的なテクノロジー、事例としての樹立を見っておらず、どのようなエラーが起こり得るか実際的な可能性すら見通せていないのが現状であり、これを今回のガイドラインにおいて議論の対象とすることは、時期尚早であるといわざるをえない。一方、非自律的な AI については、社会の様々な分野において普及が進行してきており、この開発原則について議論するには適切な時期にさしかかっていると考えられる。

以上より、本ガイドラインの議論としては、あくまで非自律的な AI に scope を絞って議論を行うべきであり、非自律的な AI に適用する原則として再度 8 原則の精緻化を行うべきである。具体的には、各 8 原則が何を意味し、どの水準までを、いかなる方法によって保障しようとするものであるかについて明確にし、さらに、かかる原則を遵守することで何が達成されるのかをステークホルダーに対し誤解や理解の齟齬のないように提示することが必要である。そして、かかる非自立型の AI については、研究者・開発者だけでなく、利用者を含めて意見を適切に汲んだうえでガイドラインについて策定すべきである。

#### 5 開発原則において議論すべき AI ネットワークの定義について

AI ネットワークという言葉に統一的な定義は存在しないため、明確な定義づけが必要である。

そのうえで、もし、なんらかの通信網を通じて何らかの通信を行う AI を想定しているのであれば、現状におけるほぼすべての AI が定義に含まれる。そのため、AI ネットワークシステム、AI ネットワークサービスという言葉はあらゆるネットワークシステムやサービスを包括しうるため、扱う対象としては不適切である。

AI ネットワークの定義について、全てのステークホルダーが理解できる様にガイドラインに明示されたい。

#### 6 AI 同士が連携するネットワークにおけるシステムについて

AI ネットワークが、何らかの形で AI 同士が連携するシステムを想定しているのであれば、AI システムとして扱うに値する定義であると考えられる。ただし、その場合システム全体が複雑系となることが明らかであるため、現在の提案されている開発原則では不十分である。改めて十分な検討を行う必要がある。

たとえば、P.4「第七 開発原則の個々の項目の内容の具体化(2) 制御可能性の原則」における、「人間又は信頼し得る 他の AI による監督及び対処（停止、切断、修理等）の実効性を確保すべき」という点について、複雑系である AI システム上では「システムそのもの」が不具合を起こすことが最も危険であるにも関わらず、単体の AI の制御についてしか考えていない。システムそのものの不具合については高速道路の渋滞をイメージすればよいだろう。渋滞は一台の車が原因を作っているわけではなく、複数の「人間である」ドライバの相互作用によって生じるものである。このような不具合を「特定のドライバ」の運転を中止させることによって解消できるものではないように、AI システムの不具合は特定の AI を停止すれば解消できるものではない。さらに言えば、システム全体を停止することは、高速道路の通行を停止させるようなもので、適切な手段とは言えない。

複雑系システムを構成する AI について、どのように対処すべきかは AI やそのシステムだけについて論じればよいものではなく、社会全体で解決すべき問題である。その場合「開発の原則」で対応できるものではなく「利用の原則」「運用の原則」を別途制定し、綿密な連携をとる必要がある。

#### 7 AI 連携を考慮したガイドラインの策定

「5 開発原則において議論すべき AI ネットワークの定義について」で述べた通り、AI 連携を考慮したガイドラインの策定は今後の AI 発展に必要不可欠であると考えられる。したがって、このようなガイドライン策定のために、研究者・開発者だけではなく、利用者を含めたステークホルダーによる検討会を設けるべきである。

#### 8 透明性と制御可能性の確保について

P.4 第七（1）（2）に透明性の原則と制御可能性の原則があるが、自律性のない AI に限ったとしても、現在のディープラーニングを中心としたビッグデータによる学習に根幹を置く AI において、完全な透明性と制御可能性を確保することは不可能である。その点を認めたくて、社会的に許容可能な範囲の透明性および制御可能性が何かを議論し、ガイドラインに盛り込まなければならない。

#### 9 透明性と制御可能性の定義について

P.4 第七（1）（2）に透明性の原則と制御可能性の原則があるが、透明性および制御可能性の定義がガイドライン上で不明瞭である。ガイドライン内にその定義を明記されたい。

#### 10 ガイドラインの実効性について

自律性のない AI については、すでに米国の Google、Amazon、Facebook、中国の Baidu などが開発の主導権を握っている。そのような状況下で日本の企業だけがなんらかの制約を受けることは日本における AI の発展を妨げ、世界的な競争に敗れる可能性が高い。

このようなガイドラインを制定するのであれば国際的に認めさせなければならない。そのためには直接ステークホルダーたる企業の意見を組み込む必要がある。日本発のこのようなガイドラインが米国、中国の企業にとって足かせになるようであれば、批准される見込みはない。

国際的な実効性をどのように確保するのか。具体的には、海外の企業がどのような理由で本ガイドラインを批准するのかを明記し、実効性を明らかにされたい。

**「AI開発ガイドライン」(仮称)の策定に向けて整理した論点に関する意見募集  
提出様式**

平成 28 年 1 月 31 日

(ふりがな)	さんぎょうきょうそうりょくこんだんかい(しーおーしーえぬ) すいしん てーま: じんこうちのうかんのこうしょう・きょうちょう・れんけいによ るしゃかいのちょうすまーとか	組織名及び代表者氏名 の公表の可否
組織名 代表者氏名 役職	産業競争力懇談会(COCN) 推進テーマ: 人工知能間 の交渉・協調・連携による社会の超スマート化 森永 聡 テーマリーダー	基本的に可なので すが、公表時にどの ように見えるのか、 事前確認させてい ただきたくおねが いします。
職業	会社員	
(ふりがな)	[REDACTED]	
住所	[REDACTED]	
連絡先	(ふりがな) [REDACTED] 担当者名: [REDACTED] 役職: [REDACTED] 電 話: [REDACTED] FAX: [REDACTED] 電子メールアドレス: [REDACTED]	

ページ	該当部分	御意見	理由
要旨P3	第七 開発 原則の個々 の項目内容 の具体化	各原則において、その実現例・満足例(広い意 味での「これでいい」を表現するホワイトリス ト)を作成し、公開するべきである。	示されている各原 則は抽象的な表現 になっており、どの ような範囲・水準が 求められるのか不 明である。開発者・ 利用者が「どこまで やればいいのか」につ いて共通な目安を 持っていないと、開 発・利用ともハード ルが高くなり、AI ネットワーク社会 の推進を妨げる。

要旨 P 3	第七 開発原則の個々の項目内容の具体化	上記、実現例・満足例の作成においては、少ない（制度的）制約のもとで実証実験を可能とする特区等（Regulatory Sandbox を含む）の指定、広くマルチステークホルダー・プロセスに基づいて制度や原則の議論を行う場の設定を行い、実用に基づいた作成と随時の改良、社会合意・社会受容性の醸成を進めるべきである。	AI ネットワーク社会の推進のためには、上記の目安の設定が現実に即していることが必要であるため。また、それが社会に受け入れられている必要があるため。
要旨 P 6	第九 開発原則の実効性の確保の在り方	末尾の「国際的な相互協力」の内容に、「各原則に対し実現を求められる範囲・水準の国際的な整合」も含めるべきである。	AI ネットワーク社会を支える製品・サービスにおいて、無用な国別対応コストが必要になると、海外に限らず国内的にも AI ネットワーク社会の推進が妨げられる。
要旨 P 8	第十一 利活用原則	利活用原則において、AI ネットワークシステムの利用者の責任と、それを理解した上での利用可否の自己決定に関して、言及するべきである。特に、利用者には AI ネットワークシステムを利用した結果に関する権利義務が帰属し、ソフトウェアないしアプリケーションとしての AI に「欠陥」があった場合に限り製造者及び販売者に対し責任と義務が帰属するという、現行法の延長線上での整理を原則としてはどうか。ただし、具体的な線引きや、その社会合意・社会受容に関しては、上記 2 番目の意見と同様、特区等での実験による作成・改良と、広い場での議論をすることとしてはどうか。	現行法の延長線上を大きく超えた責任が製造者や販売者に帰属すると、事業者としての参入は極めて難しくなり、AI ネットワーク社会の推進に妨げになるため。
要旨 P 3、P 8	開発原則 利活用原則	COCON2016 年度推進テーマ「人工知能間の交渉・協調・連携による社会の超スマート化」においては、AI 間交渉・協調・連携の社会実装に必要な要件を「開発原則」「利活用原則」の観点からブレークダウンし、それぞれ対処方針を策定・提言している。AI ネットワーク社会推進会議で参考いただけると幸いである。 #報告書の公開は2017年3月を予定していますが、COCONテーマのアドバイザーになっていただいている福田様には、原稿をお送りしてあります。	COCONの活動とAI ネットワーク社会推進会議の議論が収斂していくことが、双方の利益につながると考えるため。

以上

**「AI開発ガイドライン」(仮称)の策定に向けて整理した論点に関する意見募集  
提出様式**

平成 29 年 1 月 31 日

(ふりがな) 組織名 代表者氏名 役職	[Redacted]	組織名及び代表者氏名 の公表の可否
	[Redacted]	公表可
職業	会社員	
(ふりがな) 住所	[Redacted]	
連絡先	同上 電 話 : [Redacted] 電子メールアドレス : [Redacted]	

ページ	該当部分	御意見	理由
		別紙に記載の通り	

総務省 「AI 開発ガイドライン」(仮称)の作成に向けた～論点

[http://www.soumu.go.jp/main\\_content/000456705.pdf](http://www.soumu.go.jp/main_content/000456705.pdf)

に対するコメント

以下、「別紙1」を基準として論点をまとめます。

全体に対するコメント（論点の概要）

- 本ガイドラインは「AI 開発ガイドライン」としているが、AI を一切定義せずに開発ガイドラインを定めることにはかなりの無理がある。そもそも議論の対象となるシステムが明確に定義されておらず、個別の意見によつての修正は困難であると感じられる。一方、適用されるとすれば特に日本国内の AI 開発者や AI システムの利活用を設計する開発者に過剰な負担や委縮効果を生じる可能性があり、全体として一から見直すべきである
- より具体的には、AI ではなく、ネットワーク化されたシステムに対するガイドラインであるとしたほうが、論点が整理されるように思える。例えば安全性、透明性などに関する議論は、AI であるか、例えば人間が書いた 100 万行のプログラムであるか、あるいは内部に人間が介在することを明示しない ASP(クラウドソーシング等)であるか、ブレイン・マシン・インターフェイスにより接続されたラットの脳組織であるか、等を問わず成立する。AI の定義が何であれ、入出力と処理を具備する何らかのコンピュータプログラムとみなせる以上、AI 単独であれば何の危険もなく、AI が接続されたシステムの構造によって危険性等の懸念点が生じることから、ガイドラインはあくまでシステムに対するものであることについて明示すべきである。

- ガイドライン中で議論される、「(単独の)AI」と「AI ネットワークシステム」の捉え方について技術のあり方から考えて不自然である。システムで担保すべき要件をAI 単独で担保すべきというガイドラインになっているように読める。本来は、システムとして出力する結果やアクションこそが社会にとっての焦点であるはず。本質的に何が社会的課題なのか、個々の論点において、AI、あるいは、ネットワークないしシステム、どちらについて検討を深めるべきなのか、議論が不十分なのではないか。
- 「AI ネットワークシステムの構成要素となり得る AI について」と、ガイドラインの議論の文脈で「(単独の)AI」に対して論じている部分でしばしば、[FLI 2015][National Science and Technology Council 2016][House of Commons2016]を参照しているが、これらは共通してAI System と AI とを区分して議論しているように読める。本ガイドラインが「(単独の)AI」に関する議論を行っている文脈においてこれらの文献の議論を全体的に援用するのは乱暴ではないか。

#### 各論へのコメント

p.5

- 「開発者」の定義  
「AI ネットワークシステム」への接続を想定しないAI 開発者について、本ガイドラインの対象とすることには違和感を覚える。また、AI ネットワークシステムの開発者(AI は出来合いのものを使い、ネットワークやロボットとして構成する開発者)については、p.5において利用者と分類されているが、システムとしての最終責任はインテグレータが負うべきものであり、開発者と分類すべきではないか。あるいは、AI 開発者、AI ネットワークシステム開発者と区分しても良い。

p.16, p.17



- 全体に対するコメントでも述べたように、危険性が発生するのはシステムの構造においてであり、操作を行うのがAIであろうと、単純なフィードバック計算であろうと、人間であろうと本質的な差は存在しない。あくまでガイドラインの対象とすべきなのは「AI」ではなく、(AIが接続される)「ネットワークシステム」ではないか。
- 研究開発段階のAIや、悪意のある人間に操作されても致命的な影響のないようにシステムを構築すべきなのではないか。

p.29

- (AIへの)入出力・通信の透明性については、AIネットワークシステムからのインターフェイスよりも外側の段階で担保すべきではないか。
- 4番目の項目についてはガイドラインではなく単なる願望であって、本ガイドラインに含めるべき内容ではないと考える。

p.32, p.33

- AIは本質的には単なるコンピュータプログラムであることから、制御可能性の検証ならびに妥当性の確認は、本来外界とのインターフェイスを担うシステムの責任であるべき。例えば外部に接続されているシステムがシミュレータ(ゲーム等娯楽用のものを含む)か、あるいは実機であるかはAIの立場からすると一般には区別できないものであり、この段階でどのような制御可能性あるいは妥当性の検証が適切であるかの判断は不可能である。「その構成要素となり得るAIについて」云々ではなく、AIネットワークシステムは、いかなるAIが接続されても、当該AIの外部からの検証及び妥当性確認が可能なように設計すべきである。
- 「報酬ハッキング～AIが正常に動作せず～対策を講ずべき」についても、これはAIネットワークシステム開発者の責任とすべきである。ただし、AIが期待する入出力の範囲や利用条件等を明確化することについては、透明性およびアカウンタビリティの原則において担保

されるべきものであるだろう。その範囲外に陥った場合の問題については、AI ネットワークシステム開発者の責任とすべきである。

- (コメント) p.33 の「AI ネットワーク化検討会議『報告書 2016』(抄) において、「(4)ア(ア) AI ネットワークシステムの機密性、完全性、可溶性に対するリスクの評価」とあるにも関わらず、p.32 の論点において「その構成要素となり得る AI について」とあるのは論点のすりかえではないか。

p.34

- 項目 3: AI に関する予めのセキュリティの検証および妥当性確認とは何か。引用されている文献では具体的には述べられていないように読める(AI システムについては言及されている)。前記コメントと同様、AI ネットワークシステムのセキュリティ検証・妥当性確認とすべきではないか。
- 項目 4: AI の設計段階におけるセキュリティ・バイ・デザインとは何であるか自明でない。自明でないものをガイドラインとして含めることには問題がある。

p.35

- ここまでのコメントと同様に、安全保護についても、AI ネットワークシステムとして検討すべき内容である。従って、安全保護の原則が適用される AI の範囲を論じるのは適切ではなく、安全保護の原則が適用される AI ネットワークシステムの範囲について論じるべきである。
  - 補足: AI に限らず、構成要素の開発段階で、システムとしての安全保護を行うのは困難である。具体的には、安全であることを担保するためには危険源に対して発生防止・拡大抑制・影響緩和などの措置を講ずることが一般的である(c.f. 東京大学 古田先生の講演 [http://aviation.j-navigation.org/presentation/200910\\_Furuta.pdf](http://aviation.j-navigation.org/presentation/200910_Furuta.pdf))。つまり、安全保護においては危険源を明確にする必要があり、かつこれに対してどの程度の

リスクを見積るかについて検討が必要である。しかし、危険源はシステムに応じて異なり、例えば自動運転システムに限っても、閉鎖された空間(工場内等)における自動走行におけるもの、自動車専用道路におけるもの、一般道におけるもので危険源とリスクは異なる。また、入力となるセンサーも、単眼カメラのもの、二眼立体視のもの、レーダーを用いるものならびにそれらの組み合わせなども考えられうる。

「AI」の開発段階において、これら全てを想定した安全保護を想定することは困難である上に、仮に可能であったとしても著しくイノベーションを阻害すると思われる。安全保護が必要な場合は、システムとしての安全保護が適切であり、その責任はAI ネットワークシステム開発者に課されるべきである。

- 項目 2: p.34 項目 3 へのコメントと同様。
- 項目 3: AI の設計段階におけるセーフティ・バイ・デザインとは何か。

p.38

- 項目 5: AI の設計段階におけるプライバシー・バイ・デザインとは何か。顔認識のためのAIにおいて、一般大衆の顔認識とセキュリティ用途の人物認識とで同じAIを用いることができる場合が考えられる。このような場合、AI の設計段階におけるプライバシー・バイ・デザインを定義可能か。「プロファイリングの用に供するAI」とは何か。同じAIでプロファイリングと異常検知(セキュリティ・セーフティ)とに应用可能な場合、プロファイリングに利用できることから「特に慎重に措置を講ず」とはイノベーションの阻害ではないか。
  - 本項についても、AI ネットワークシステムのプライバシー・バイ・デザインとすべきではないか。

p.42

- 項目 2, 項目 3: ナッジおよびユニバーサル・デザインについては、AI ネットワークシステムとして提供すべき要素である。AI ネットワークシステムの構成要素としてのAIは、デフ

ォルトの設定、理解しやすい選択肢の提示・体系化をする要素であるとは考えられない。フィードバックの提供、緊急時の警告、エラーへの対処等についてはここまでの項目で触れているため触れる必要はないと考える。また、AIの「ユニバーサル・デザイン等社会的弱者の受容可能性を高めるための取組」とは具体的には何を意味しているのか。

p.47 「連携の原則【仮称】」

- 項目 1: AI に関する相互接続性・標準化等について、AI ネットワークシステムに関するものとして記載すべきであるか、あるいは一切記載を省き、市場に任せるべきである。

p.51, 52 「開発原則の実効性の確保における市場の活用の在り方」

- 項目 2: ここまで述べたように、「当該 AI の開発原則への適合性を評価して認証」することは総合的に見て現実的ではない。AI システムのレベルで個別に評価・認証すべきである。
- 項目 3 以降: 「開発ガイドライン」について、AI システムのレベルで担保すべき事項と AI のレベルで担保すべき事項について明確に区分した上で進めるべきである。

p.58～p.62 「AI ネットワークシステムの利活用に関し利用者等が留意すべき事項」

- 本「開発ガイドライン」について、AI システムのレベルで担保すべき事柄と AI のレベルで担保すべき事柄が曖昧であり、国際的な議論に資する水準に達していないと考える。従って、「利活用ガイドライン」についても国際的議論を進めるのであれば、双対となる「開発者ガイドライン」について論点を整理した上で進めるべきである。現時点で、AI を組み込んだシステムの開発者(AI ネットワークシステム開発者)の責任と、AI 開発者の責任とを明確に区分せずに単純に「利活用」としてまとめている点には大きな疑問を生じる。
  - 特に、項目 1 の(2)の例示において、AI システム開発者(本ガイドラインの定義における「利用者」)の責任が事例として挙げられているにも関わらず、「開発ガイドライン」においてその対応付けが明確化されていない点は理解に苦しむ。

- 項目 8: 「当該リスクの顕在化に関する主張立証責任を当該第三者から当該プロバイダ又は当該開発者に転換する制度」については、所謂「悪魔の証明」とならないように妥当な制約を付けるべきではないか。例えば、再現性のない主張に対して、再現しないことをどう主張立証すればよいのか。また、本項はガイドラインの実質的な強制力とならないか。

以上

**「AI開発ガイドライン」(仮称)の策定に向けて整理した論点に関する意見募集  
提出様式**

平成29年1月31日

(ふりがな) 組織名 代表者氏名 役職	いっぱんしゃだんほうじん しんけいざいれんめい 一般社団法人新経済連盟 三木谷浩史 代表理事	組織名及び代表者氏名 の公表の可否
		可
職業	団体 代表	
(ふりがな) 住所	[REDACTED] [REDACTED]	
連絡先	(ふりがな) [REDACTED] 担当者名: [REDACTED] 役職: [REDACTED] 電 話: [REDACTED] F A X: [REDACTED] 電子メールアドレス: [REDACTED]	

ページ	該当部分	意見・理由
全て	全て	AIは第4次産業革命を主導する技術であり、今後のグローバルな競争環境においてその深化・発展は我が国そのものの浮沈に直結するものと考えられる。AIにおけるイノベーションを促進するには、何より自由でオープンな開発環境が必要である。仮にガイドラインが必要としても、その性格は上記環境の確保の観点からのものとし、かえって開発を抑制することにならないよう十分留意する必要がある。
P4-6	全て	論者によって様々な定義が混在しており、また将来的な発展の範囲を確定することが困難であることから、AIの定義を行わない、という方針については一定の理解ができるが、従来の情報システム・計算機処理との相違程度は示した方が、AIの裾野の広い開発や適切な活用の推進に資するものと考えられる。
P11-12	全て	目指すべき絵姿として「智連社会」を掲げ、あくまでも人間がAIネットワークシステムを主体的に使いこなす社会を目指すべき、という考え方自体には同意する一方、特定のAIネットワークシステムへの過度の依存の危険性も考えられる。上記の基本的な方向性を踏まえると、人間(利用者)が常に代替手段(他のAIネットワークシステムへの切り替え)を確保できるような配慮が為された普及の在り方が求められてくることになる。
P27-29	全て	透明性の確保に関しては、多くのAIの学習はブラックボックス化していくことが予想されるため、AIが学習に用いたデータが適切に保存さ

		れ、また学習過程の AI のスナップショットをとることによって再現可能性が保障される、といったことが必要になってくると考えられる。
P37-38	全て	プライバシーに関しては、AI のデータ処理能力の飛躍的向上が推論・予測アプリケーションを発達させ、それによって「将来に渡るプライバシー」が現時点で侵害の危機に晒されるという可能性への目配りも求められる。これについては、こういった課題をまずは認識した上で、イノベーションを阻害する蓋然性を低減させるため、開発段階よりは、利用段階における取り決めによって対処するという方向性が望ましい。
P18-25	全て	開発原則の実効性確保に関しては、開発者が開発する AI に関する情報を自発的に提供する仕組みや第三者機関が開発原則への適合性を評価し認証する制度の創設が例として挙げられているが、これは、開発者の自由な開発を委縮させ、その発展を妨げる可能性を有するものであり、ガイドラインに書き込むことについては極めて慎重であるべき。
P19-20	全て	相互接続性に関しては、インターネットが既に標準技術として相当程度広まっていること、AI は人間のインプットもアウトプットも模倣することが可能となり他のネットワークに対して如何様にも接続し得るようになると考えられること等から、必ずしも議論の対象としなくてもよいものと考えられる。基本的に相互接続や連携は、技術ではなく、運用や評価の問題であり、開発原則においては、透明性とセキュリティを重視することが特に必要となる。
全て	全て	技術革新の加速度的な進展によって、今後、AI 利用者のリテラシーや利用態度による格差は急激に拡大し、それが組織や都市・国家間の格差にもつながっていく可能性が想定される。AI の発展がもたらすと予想されるこのような課題を見据え、AI をめぐる社会の受容性に関する議論が今後求められてくると考えられる。

**「AI開発ガイドライン」(仮称)の策定に向けて整理した論点に関する意見募集  
提出様式**

平成 29 年 1 月 31 日

(ふりがな) 組織名 代表者氏名 役職	かぶしきがいしゃ ぷりふあーどねっとわーくす 株式会社 Preferred Networks 西川 徹 代表取締役	組織名及び代表者氏 名の公表の可否
		可
職業		
(ふりがな) 住所	[Redacted]	
連絡先	(ふりがな) [Redacted] / [Redacted] 担当者名: [Redacted] / [Redacted] 役職: [Redacted] / [Redacted] 電 話: [Redacted] (FAX): [Redacted] 電子メールアドレス [Redacted]	

ページ	該当部分	意見	理由
		別紙記載のとおり	



## 第1 本意見書について

### 1 目的

本意見書は、「AI開発ガイドライン」(仮称)案(以下「本案」という。)及びその論点(以下「本論点」という。)について、「AI開発」の円滑化を実現するうえで特に強い疑義のある点に注意を促し、わが国がグローバル市場における「AI開発」の主導的な立場を目指すならば志向すべき政策の形成に寄与することを目的とするものである。

### 2 概要及び概論

「AI」は次代のイノベーションの鍵を握る技術である。その研究開発が各国で進められ、研究者や技術者が成果をめぐり競い合う中、わが国がこの競争に勝ち、次代のイノベーションを主導するためには、わが国の「開発者」に「AI開発」のインセンティブを正しく与える仕組みの設計が必要不可欠である。

しかし本案及び本論点は、「AI」を含む情報通信技術の使用に伴う危険の本質が、その技術そのものではなく、この技術を構成要素とする「AIネットワークシステム」又はこの技術を当該システムに実装又は接続する行為にあることを看過し、また「AI開発」の意欲を低下させかねない最終利用者等の「保護」に傾斜した開発原則を広くわが国の「開発者」に適用することによって、その「AI開発」を停滞させる危険を有している。この危険が実現することは、次代のイノベーションを主導する機会がわが国から失われることを意味する。

わが国の「AI」の進歩がもたらす利益は、一部の最終利用者によってのみ享受されるものではない。また、最終利用者等の利益は、「AI」の使用に伴う危険の抑制によってだけでなく、その進歩によっても擁護されることを看過すべきでない。本案の策定に当たっては、本案が目的とすべきわが国の最終利用者の利益の最大化という見地から、さらなる検討を慎重かつ迅速に重ねる必要がある。

## 第2 論点等に係る意見について

### 1 第一 基本概念の定義 論点1・3・4について

関連頁

技術的な観点からいえば、「AI」とは、入力を受けて処理を行い、その結果を出力するコンピュータプログラムであり、その過程において機械学習、深層学習その他一定の技術が用いられるものを指す。その外延を画することのないまま本案が策定されれば、当然その射程とする範囲も不明確となるのであって、「AI」の「開発者」に施策者の意図しない委縮効果を与える危険がある。具体的な技術の内容はその発展とともに変わり得るとしても、将来の技術の変化を受け容れる柔軟な定義を置くことによってこの問題に対処すべきであって、定義を置かずにこれを避けて通るべきではない。

p.5

また本論点では、『「開発者」とは、AIの研究開発(……)をする者(……)』と定めつつ、『AIの研究開発』の範囲を議論する旨の提案が行われている。し

かし、その範囲をどのように定めるにせよ、本案の名宛人である『開発者』に一見して曖昧かつ広範な定義を与えることは、わが国の研究者や技術者に施策者の意図しない委縮効果を与える危険がある。上述のとおり、「AI」を含む情報通信技術の使用に伴う危険の本質は、その技術自体にではなく、この技術を構成要素とする「AI ネットワークシステム」又はこの技術を当該システムに接続する行為にある。したがって、「AI ネットワークシステム」への接続が予定されない技術はそもそも「AI」の定義から除かれるか、そのような技術を研究開発する主体は「AI」の「研究者」の定義から明示的に除かれるべきである。

「AI」の使用に伴う危険に対する責任は、「AI ネットワークシステム」を開発し設計するシステムインテグレーターが一義的に負担すべきものである。このように整理したとしても、「AI」の「開発者」とシステムインテグレーターとの間では、市場の作用により、客観的な事情に応じた危険に対する負担の再分配が両者の契約を通じて行われるのであって、システムインテグレーターのみが責任を負わなければならない事態が生じることは想定し難い。したがって、自ら「AI」を研究開発するか否かに係わらず、「AI ネットワークシステム」を開発・設計する主体は、「利用者」ではなく「開発者」に当たるものとして整理すべきである。

なお付言すると、現在「AI」に用いられる機械学習の成果は、訓練データと呼ばれるインプットの内容によって大きく異なってくる。このインプットは、「開発者」によって行われることもあれば、一般には「利用者」に当たると評価せざるを得ない主体によって行われることもあり得る。たとえば、Twitter のユーザーが不適切なインプットを行った結果、政治的に適切さを欠く Tweet を繰り返すようになったマイクロソフト社の「Tay」と呼ばれる対話型プログラムが記憶に新しい。つまり、「AI」の使用に伴う危険は、これを利用する側、特にインプットを提供する主体の挙動によって大きく左右され得るが、これを事前にすべて想定することが「開発者」に求められるとすればその責任はあまりに過重というべきである。この種の危険に対する責任が「開発者」の側にあることを前提とした整理は決して採用されるべきでない。

## 2 第四 目的、基本理念等 論点1について

本論点では、『……AI ネットワークシステムの最終利用者の利益を保護するとともに第三者及び社会への波及的な悪影響を防止し、もって人間中心の智連社会の形成に資することを』本案の目的として掲げる旨の提案が行われている。しかし本案で目指すべきは、最終利用者等の保護ではなく、広い意味での受益者の利益の最大化であって、最終利用者等の保護はその一つ的手段に過ぎない。揺籃期にある「AI」の使用に伴う潜在的な危険をことさらに不安視し、最終利用者等の保護を本案の目的に掲げることは、その本来実現すべき目的を見誤り、わが国において「AI」の研究開発に携わる者の意欲の低下を招く

p.13

危険があるものというべきであって、あらためられなければならない。

### 3 第五 適用範囲 論点2・3について

本論点では、「何らかの情報通信ネットワークシステムに実装し又は接続することが技術的に可能なAI」と「いかなる情報通信ネットワークシステムに実装し又は接続することも技術的に不可能なAI」とがあることに言及しつつ、いずれも本案の適用対象となることを当然の前提としている。しかし、その具体的な定義に係わらず、「AI」とは、入力を受けて処理を行い、その結果を出力する一定のコンピュータプログラムであることに変わりはなく、「AI」そのものに危険が内在しているという理解は誤りである。仮に本案の目的を「AI」の使用に伴う危険から最終利用者等を保護するところに置くとしても、一般的に「AI」を本案の適用対象とすることには強く反対する。「AI」の使用に危険が生じるのは、「AI」をネットワークシステムに実装し又は接続することがその理由なのであるから、本案の対象は「AI」ではなく、「AI」を構成要素とする「AI ネットワークシステム」とすべきである。

p.16-7

### 4 第七（1） 透明性の原則 各論点について

本論点では明示的に区別されていないが、「AI」の研究開発の段階で担保すべき問題と、「AI ネットワークシステム」を介した入出力及び通信の段階で担保すべき問題とは明確に区別されるべきである。入出力及び通信についての「透明性（検証可能性及び説明可能性）」は、「AI」の側ではなく、「AI ネットワークシステム」の側で担保されるべき課題である。

p.29

また、論点4のいう「有力な手段」は、現時点において実現ないし実装可能な技術ではなく、本案に明記するのは時期尚早というべきである。

### 5 第七（2） 制御可能性の原則 各論点について

本論点では明示されていないが、制御可能性の検証及び妥当性確認は、「その構成要素となり得るAI」の側ではなく、「AI ネットワークシステム」の側で担保されるべき課題であることを明確に示すべきである。本論点において、AI ネットワーク化検討会議の『報告書2016』における議論からの妥当でない逸脱が認められることに強い危惧を表明する。一例を示すと、「AI」を用いた航空関連技術が存在し、これをあるネットワークシステム上に実装し又は接続したとして、このネットワークシステムが実機上のものであるか、シミュレータ上のものであるかによって制御可能性の検証等の方法は異なりうる。しかし、「AI」の「開発者」の側でこのことを事前に予測することは必ずしも可能ではないし適切でもないのであって、「AI ネットワークシステム」を設計する主体にその検証等の責任を委ねるべきである。

p.32

また、論点1の「報酬ハッキング……の結果AIが正常に動作せず意図しない事象が生ずるリスク」等も、「AI ネットワークシステム」の側で対応すべき課題である。あえて「AI」の「開発者」の責任に言及するとすれば、通常のコンピュータプログラムの場合と同様、「AI」を使用する際に期待される入出力

の範囲の明確化等の限度であることを明確にすべきである。

#### 6 第七（3） セキュリティ確保の原則 論点3・4について

論点3の「AI」に関する事前の「セキュリティの検証及び妥当性の確認」の指す内容は、被引用文献に照らしても明らかでない。そのため、わが国における「AI」の「開発者」に施策者の意図しない委縮効果を与え、引いてはわが国の「AI 開発」の発展を停滞させる危険がある。セキュリティの確保についても、「構成要素となり得る AI」の側ではなく、「AI ネットワークシステム」の側で担保されるべき課題であることを明確に示すべきである。

p.34

また、論点4の趣旨及び「セキュリティ・バイ・デザイン」の意義が不明である。

#### 7 第七（4） 安全保護の原則 論点1・2・3について

安全の保護についても、「構成要素となり得る AI」の側ではなく、「AI ネットワークシステム」の側で担保されるべき課題である。したがって、「AI」ではなく、「AI ネットワークシステム」にこの原則が適用されることを前提に、その適用範囲を検討すべきである。

p.35

そもそも、「AI」であるか否かを問わず、ネットワークシステムの潜在的な構成要素を開発する段階において、これがネットワークシステムに実装又は接続されたときに生じ得るあらゆる危険への対策を講じることは事実上不可能である。これを行うためには、まず潜在的な危険の発生源を明確にし、その危険の程度を予測することが不可欠であるが<sup>1</sup>、その危険の発生源は実装し又は接続するネットワークシステムにより大きく異なり得る。たとえば、自動運転技術及びそのシステムについていえば、それが用いられるのが閉鎖された空間においてであるか、自動車専用道路においてであるか、一般道においてであるかによって、潜在的な危険及びその発生源は異なる。

安全保護の要請が「AI 開発」の段階から働くとすれば、「AI」の「開発者」はあらゆる危険の発生源を想定し、各々への対応に追われることになる。しかし、そのような事態はまったく現実的でないうえに、わが国のイノベーションの機会を著しく毀損することは明らかである。そのため、安全保護の責任は「AI ネットワークシステム」を設計する主体に委ねられるべきである。

論点2の「AI」に関する事前の「安全性の検証及び妥当性の確認」の指すところは、被引用文献に照らしても明らかでない。そのため、わが国における「AI」の「開発者」に施策者の意図しない委縮効果を与え、引いてはわが国の「AI 開発」の発展を停滞させる危険がある。安全性の確保についても、「構成要素となり得る AI」の側ではなく、「AI ネットワークシステム」の側で担保されるべき課題であることを明確に示すべきである。

また、論点3の趣旨及び「セーフティ・バイ・デザイン」の意義が不明である。

<sup>1</sup> 参照：[http://aviation.j-navigation.org/presentation/200910\\_Furuta.pdf](http://aviation.j-navigation.org/presentation/200910_Furuta.pdf)

## 8 第七(5) プライバシー保護の原則 論点5について

論点5の趣旨及び「プライバシー・バイ・デザイン」の意義が不明である。仮にその趣旨が「AI」の設計段階でプライバシーに配慮した措置を組み込むことを求めるものであるとすれば、上記7で述べたところと同様にわが国のイノベーションの機会を著しく毀損するものであるというべきであって、これに強い危惧を表明する。ヒトの顔認識のための「AI」を例にとると、個人のアイデンティティと結びつける方法でこれをセキュリティチェック等に用いることも、個人のアイデンティティを取得せずに一般大衆の顔認識等に用いることも可能であるという場合があり得る。さらに、このセキュリティチェックの方法についていえば、「プロファイリング」にこれを用いることも、異常検知の手段として用いることも可能であるという場合があり得る。つまり、「AI」の設計段階でプライバシー保護の要請が働き、「特に慎重に措置を講」じるべき事態が生じるとすれば、「AI」の「開発者」はあらゆる事態を想定し、各々への対応に追われることになる。しかし、そのような事態はまったく現実的でないうえに、わが国のイノベーションの機会を著しく毀損することは明らかである。

p.38

## 9 第七(2) 利用者支援の原則 論点2・3について

論点2の「設計」及び論点3の「配慮」は、「構成要素となり得るAI」の側ではなく、「AIネットワークシステム」の側で担保されるべき課題である。たとえば、論点2に「ナッジ」の要素として挙げられるものの多くは、「AI」の側で設計されるべきものと一般に考えられていない。

p.42

## 10 第八 連携の原則【仮称】 論点1について

「相互接続性・相互運用性の確保が期待される事項」、「標準化が期待される事項」その他「連携の原則」の名の下に掲げられる各事項については、「AI」ではなく、「AIネットワークシステム」に関するものとして整理するか、又はこれらが原則として市場の原理に任せられるべき事項であることを考慮して一切の記載を排除すべきである。

p.47

## 11 第十 開発原則の実効性の確保における市場の活用の在り方 論点2について

論点2は「開発者がその開発するAIに関し開発原則への適合性に関する情報を客観的に信頼できる形で自発的に提供する仕組み」に言及するものである。しかし、上述のとおり各開発原則は、「AI」の側ではなく、「AIネットワークシステム」の側で担保されるべき課題である。仮にわが国において「開発原則への適合性を評価して認証する制度」を設けるとしても、その評価等の対象はあくまで個別の「AIネットワーク」であることを明示すべきである。

p.51

## 12 第十一 AIネットワークシステムの利活用に関し利用者等が留意すべき事項 各論点について

p.58-62

本論点では、本来「AI ネットワークシステム」の側で担保されるべき課題が「AI」の側で担保されるべき課題として整理されるなど、国際的な議論に耐え得るとはいえない議論が散見され、本案を国際的な議論の俎上に載せることの要否を論じるのはいまだ時期尚早であるというべきである。

本論点によれば、「AI」を実装し又は接続した「AI ネットワークシステム」の設計者は「開発者」ではなく「利用者」として扱われることになるが、これは「AI」を含む情報通信技術の使用に伴う危険の所在が正しく認識されていないことの帰結であるように思われる。わが国ではまず、受益者の利益の最大化を図るための地に足の着いた議論を本案について迅速かつ丹念に進めるべきであり、その論点を整理したうえで「利用者」のためのガイドラインの議論に移行し、そののち国際的な議論の要否の検討に進むべきである。

なお、論点8の「リスクの顕在化に関する主張立証責任を当該第三者から当該プロバイダ又は当該開発者に転換する制度を整備することについて検討するよう国、関係国際機関等に推奨」することが現実のものとなれば、たとえ法的義務が伴わないものであっても、将来の実務に大きな影響が及ぶことが容易に予想される。そのため、法的な拘束力の有無を問わず、拙速にこの種の議論を押し進め、本案中で言及することは決して許されるべきではない。

### 13 総括

本案は、わが国における将来のイノベーションを阻害するものであってはならない。そのためには、「AI」を含む情報通信技術の使用に伴う危険の本質を正しく理解したうえで、わが国における「AI 開発」の意欲を低下させることのないよう、本案に関する議論を慎重に進めなければならない。

本案を「AI ネットワークシステム」の研究開発に関するガイドラインと再構成することによって、本論点の多くについて本来あるべき整理が可能となるように思われる。「AI ネットワークシステム」が出力する結果にこそ社会的な課題が潜むはずであり、まずは「AI」と「AI ネットワークシステム」のそれぞれの側で担保されるべき課題を整理し直すことが不可欠である。このことは、本論点における被引用文献<sup>2</sup>の中では当然の前提とされている。

本意見書は、「AI」や「AI ネットワークシステム」に関連する社会的な課題やあるべき「研究開発」のあり方等について迅速に議論を進め、またわが国がこれらを巡る国際的な議論を先導することの利益を否定するものではない。しかし、誤った前提や目的から出発した議論が正しい結論を導くことは極めて困難であることから、本意見書ではこのような誤りのいくつかを指摘し、尽くされるべき議論がこれから尽くされることを期待するものである。

---

<sup>2</sup> 参照：[FLI 2015]、[National Science and Technology Council 2016]、[House of Commons 2016]

**「AI開発ガイドライン」(仮称)の策定に向けて整理した論点に関する意見募集  
提出様式**

平成 29 年 1 月 31 日

(ふりがな) 組織名 代表者氏名 役職	アイエススクエア 株式会社 ISS スクエア 法制倫理研究分科会 門脇源太郎、窪優司、田村壮世、中島尚樹、南後吉秀、 脇坂尚弘	組織名及び代表者氏名 の公表の可否  可
職業	大学院生	
(ふりがな) 住所	[Redacted]	
連絡先	(ふりがな) [Redacted] 担当者名: [Redacted] 役職: [Redacted] 電 話: [Redacted] F A X : [Redacted] 電子メールアドレス: [Redacted]	

ページ	該当部分	御意見	理由
P40	4. 人間の脳・身体と融合又は連携するAIを研究開発する際には、人間の尊厳と個人の自律の尊重について、生命倫理等の議論も参照しつつ、特に慎重に配慮すべきかどうか。	特に慎重に配慮して研究開発を行うべき領域として、左記「人間の脳・身体と融合又は連携するAIを研究開発」以外の領域についても十分な議論の必要があるのではないかと。例えば、著しく人間の仕事をAIが奪う可能性をもつ研究開発や、人間とAIの価値が逆転しかねない状況を引き起こす可能性のある研究開発などが考えられる。	「人間の尊厳と個人の自律の尊重」に係る開発領域は様々な領域が想定されるため、「要配慮領域」について十分な議論を行い、全体で意識を合わせる必要があると考える。
P32	3. AIネットワークシ	自律性が高く、且つ利用者及び第三者の生命・身体の安全に危害が及ぶリスクがあるAI・ロボ	AI が自律的に発達して行った場合、社

	<p>システムの制御可能性を継続的に確保するために、その構成要素となり得る AI について人間又は信頼し得る他の AI による監督及び対処（停止、切断、修理等）の実効性を確保すべきとしてはどうか。</p> <p>・緊急停止機能に関する技術標準やプロセスについても指針を定めるべきではないか</p>	<p>ット等には、「緊急安全停止機能」を開発段階で考慮することは必須であると考ええる。</p> <p>たとえば自動車製造においては、道路運送車両法にて自動車の構造、装置、性能に関する安全性の最低基準を定めて、これに適合しない自動車の運行を禁止している。</p> <p>道路運送車両法のような実効性をもつ法律にて、緊急安全停止機能の実装が各分野別ガイドラインで義務付けられるよう、共通ガイドラインで規定すべきではないか。</p> <p>また、規定する安全基準を懈怠する AI・ロボットが開発・使用された場合に備えて、製品・サービス提供者以外の国や第三者機関による緊急安全停止機能の実行が可能とする仕組みの規定についても、議論の必要があるのではないかと考える。</p> <p>例えば、現在のインターネットの DNS ルートサーバのような権威的存在を頂点とした階層構造を持ち、緊急安全停止機能が当該階層構造を伝播して作用するような仕組みが考えられる。</p>	<p>会や人間に対し、予測不能な影響を与え出す可能性は十分考えられる。</p> <p>いかなる場合でも人間のコントロールの配下に置くため、緊急停止機能は必須である。</p>



差出人： AIネットワーク(総務省情報通信政策研究所)  
宛先： [REDACTED]  
件名： FW: 【案件番号:145208852】 パブリックコメントに関する提出意見の配信(受付番号: 201701310000392043)  
日付： 2017年1月31日 18:05:42

---

-----Original Message-----

From: public-com-egov@e-gov.go.jp [mailto:public-com-egov@e-gov.go.jp]  
Sent: Tuesday, January 31, 2017 6:05 PM  
To: ai.network@soumu.go.jp  
Subject: 【案件番号:145208852】 パブリックコメントに関する提出意見の配信(受付番号:201701310000392043)

パブリックコメントに関する意見提出先窓口担当者 様

電子政府の総合窓口 (<http://www.e-gov.go.jp/>) から貴府省宛に  
パブリックコメントに関する意見が提出されましたので、配信します。

受付番号:201701310000392043  
受信日付:2017/01/31 16:38:20

案件番号:145208852  
案件名:  
「AI開発ガイドライン」(仮称)の策定に向けて整理した論点に関する意見募集  
宛先府省名:総務省

郵便番号 [REDACTED]  
住所 [REDACTED]  
氏名 [REDACTED]  
連絡先電話番号 [REDACTED]  
利用者メールアドレス [REDACTED]

提出意見:

「AI開発ガイドライン」について、そもそも議論の対象となるシステムが明確に定義されておらず、個別の意見によつての修正は困難であると感じられる。一方、適用されるとすれば特に日本国内のAI開発者やAIシステムの利活用を設計する開発者に過剰な負担や委縮効果を生じる可能性があり、全体として一から見直すべきである。

以上の意見を、以下に示す個人の総意として提出する。

[REDACTED]

-----  
電子政府の総合窓口  
<http://www.e-gov.go.jp/>  
-----

メール識別No:0000336907

「AI開発ガイドライン」（仮称）の策定に向けて整理した論点に関する意見募集  
提出様式

平成 29 年 1 月 31 日

(ふりがな) 組織名 代表者氏名 役職	えーあいかいはつがいどらいん（かしょう）ぱぶりっくこめんと しっぴつゆうしのかい AI 開発ガイドライン（仮称）パブリックコメント執筆有志の会 代表：atoll Project アーキテクト川田大輔 新潟大学大学院実務法学研究科・法学部教授田中幸弘	組織名及び代表者氏名 の公表の可否
		可
職業	川田大輔：IT コンサルティング業 田中幸弘：大学教員	
(ふりがな) 住所	[Redacted]	
連絡先	(ふりがな) [Redacted] 担当者名： [Redacted] 役職： [Redacted] 電 話： [Redacted] F A X： [Redacted] 電子メールアドレス： [Redacted]	

ページ	該当部分	御意見	理由
		別紙参照のこと	

「AI 開発ガイドライン」（仮称）の策定に向けて整理した論点に関する意見募集  
に対する意見提出（別紙）

AI 開発ガイドライン（仮称）パブリックコメント執筆有志の会  
atoll project Architect 川田大輔  
新潟大学大学院実務法学研究科・法学部教授 田中幸弘

技術的側面からの提言（文責：川田）

“「AI 開発ガイドライン」（仮称）の策定に向けた国際的議論の用に供する素案の作成に関する論点”という文書名が表すように本文書は AI 開発についての国際的な合意形成を目標としていると受け止めております。

内生的成長理論（ローマー）のいう成長飽和に達している現世代技術の枠組みを超えて成長を持続する方法論として OECD もその採用を奨めるオープンイノベーション（チェスブロウ）の枠組みに当てはめると本文書において「AI」と呼ばれる技術分野では Research、Development、Commercialisation の各プロセス段階ですでに米国企業の保有する技術の中核としたビジネスエコシステム（ムーア）形成が始まっている。なお悪いことに「AI」技術分野は応用のすそ野が極めて広く ICT 産業全体に大きな影響を及ぼすと予想（注）されています。

注 <http://techemergence.com/valuing-the-artificial-intelligen.../> 等

日本企業は「AI」技術分野成長の前段かつ足場となる大規模計算資源プール（クラウドコンピューティング）の提供事業分野で米国事業者の後塵を拝しており比較競争劣位（リカド）に置かれています。現状の競争条件のまま米国企業の保有する「AI」技術の中核としたエコシステム形成が進むと日本企業は競争上非常に厳しい環境に置かれます。

日本政府ならびに国内関係者の立場に立つと、「AI」技術およびその前段となる大規模計算資源プール提供事業分野で先頭を行く米国事業者と互角に競争できる国内企業が見当たらない現状において、国内企業が先頭を行く米国事業者に対して比較競争優位が獲得できないのであるならば早期に現在比較競争優位にある米国企業に足枷をはめて米国企業の保有する技術の中核とした「AI」技術エコシステム利用（参加）にあたって中核プレイヤーに対する影響力保持を図っておきたいと考えるやもしれないことは理解できます。

「AI」技術分野における影響力保持の手段として「AI」技術開発における「開発原則」整備を挙げ、国際原子力機関や国際宇宙法、ヒトゲノムと人権に関する世界宣言などに続く国際共同規制の枠組み構築を狙う戦略自体は支持できます。

また、「AI」ネットワークシステム相互接続を想定した「AI」ネットワーク化をスコープして国際共同規制の枠組み構築の正統性を構築する論法も理解できます。

しかしながら「AI」開発ガイドライン策定を謳いながら、対象となる「AI」について定義を避ける手法は支持できません。開発を規制する対象を特定しないガイドラインに実効性を持たせることはできないので日本政府の目的が「AI」技術分野における国際的な影響力保持であるなら「AI」につ

いての定義を避けてのガイドライン策定は目的に照らして完全な失敗となるでしょう。

失敗を避けるには世界的に合意形成が可能な AI についての定義モデル作成が必要といえます。とはいえ、A Survey of Current Practice and Teaching of AI (注)にも示されているように第一線の研究者と実務者の間でも重視するポイントに隔たりがあるのが現状です。そもそもが AI : Artificial Intelligence (人工知能) について定義しようにも現代科学は Intelligence (知能・知性) の構造をいまだ解き明かせていませんので、知能・知性を人工的に再現する取り組みたる AI の研究開発が本当に知能・知性を再現しているかを検証することができません。知能・知性に当たるか否かも判定できない状況において本文書に AI それ自体の定義策定を求めることは酷に過ぎると理解は示せます。とはいえ、昨今話題の Machine Learning などの俗に「AI」と呼ばれる技術群についての系統だった分類を行い、どのような技術をどのように運用すると、どのようなリスクを受容せねばならないかは定量的に評価しておくべきと考えます。

注：<http://www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/viewFile/12444/12195>

本文書で策定が模索されている開発原則は原則策定の目的として記述されている、“人間が AI ネットワークシステムと共存することにより、AI ネットワークシステムの恵沢が万人に享受され、人間の尊厳と個人の自律が保障されるとともに、AI ネットワークシステムの制御可能性と透明性が確保され、AI ネットワークシステムが安全に安心して利活用される社会を実現するという理念”を実現する方法論として適切に構造化されていません。

そもそも適用対象である AI についての定義、または適用対象となる具体的技術の指定がなされていないため生命倫理と人権に関する世界宣言や IAEA 憲章などと比較すると開発実務を制約するモデルとして抽象的に過ぎ、実効性に欠けています。

なお、提示された理念のうち、「安全に安心して利活用」という文言については日本政府の好むところではありますが、リスクフリーな選択肢はそもそも存在しないので安全に安心して利活用できるという共同幻想構築が目的でないならば削除すべきと考えます。また、制御可能性については低姿勢問題の決定不能性定理に照らして保証できない点を考慮すべきです。

透明性の原則、倫理の原則、アカウントビリティの原則は ISO26000 (社会的責任の手引き) に記載されている 7 つの原則でカバーされているテーマの一部であり、AI 開発における原則以前に社会的責任の問題であると考えます。

利用者支援の原則は AI システムがその利用者の問い合わせに対して複数の選択肢の存在する回答候補が挙げられる場合に利用者を選択の機会を与えることを要求していますが、選択候補の抽出とは何らかの重みづけアルゴリズムによる重みづけられた結果であり、その重みづけモデルの適切性 (社会受容性) は (他の規格や社会合意によって代替可能ですが) ISO26000 の 7 つの原則を利用して社会合意形成して得るほかになく (Google 検索結果のナチスホロコーストの取り扱いなど参照) 開発原則として一意に定義できません。

なお、利用者支援よりも差別禁止 (日本法的には通信の秘密の文脈での法の下での平等の扱いとして)の方が上位概念であるようにも感じられます。

制御可能性の原則については停止性問題の決定不能性定理に基づいて考えれば実現不能であるし、実用的な停止方法を研究した最近の論文（注）によってもすべてのアルゴリズムを簡単に安全に停止できるかは不透明であると指摘されていて実現性が低いでしょう。制御可能性の原則という制約は現状のまま採択されると AI 開発を停滞乃至衰退させる虞があります。

注：<https://intelligence.org/files/Interruptibility.pdf>

セキュリティ確保の原則、安全保護の原則、プライバシー保護の原則は AI 固有の問題ではありません。

論点として挙げられた OECD セキュリティガイドライン参照など外部参照するモデルは支持できません。機密性、完全性、可用性に加え、信頼性と頑健性を動的に検証し続けるフレームワークは実現できる場合は有益と評価できるが信頼性と頑健性の評価モデルをまず構築する必要があります。

このように見ていくと現在挙げられている各原則は AI 開発原則として構成が不適切であると考えます。また理念にはそもそも実現不可能な要素が含まれているといえます。開発原則は同原則内での自己完結を想定しています（P22-P25）がそれゆえに原則として抜け漏れが多い散文的な原則になってしまっているように見えます。

制御不可能性の原則としてインターネットを例に挙げるまでもなく、そもそも開放系のネットワーク構造をもつシステムは制御不可能になる可能性を持ったシステムである、という事実を受け入れ、それでも受け入れ可能なリスクとして社会受容できるようにするための原則モデルを構築するというアプローチは合理的に成立し得ます。

なお、AI ネットワーク化検討会議報告書記載の開発原則案であるにも関わらず、AI システム相互接続されたサプライチェーン（ネットワーク）に起因するリスクをカバーする原則が含まれておらず、SCM 関連規格を参照して適切な原則追加が必要となると考えられます。そもそも論として国際共同規制の枠組み構築の正統性をネットワーク化に求めておきながら当該テーマについて問題意識の記述に留まっているのは許容できません。

障害の連鎖を防止し障害範囲を封じ込める（通信網輻輳制御、鉄道網の直通運転制御などの事例参照）手法組み込みなども考慮すべきでしょう。また、自律した疎結合な分散システムの協調動作という観点からは群制御の視点も必要になるかもしれません。

ガイドライン実効性確保の方法については強制力の設定や参加インセンティブの設計などが必要になります。過去事例を挙げてデザインパターンを整理する必要があるでしょう。

AI ネットワーク化の進展に向けた協調の円滑化についてはオープンイノベーションの文脈で整理可能。競争的なエコシステムの確保については現在激烈な競争が行われており市場活性化策は不要と考えます。

(文責：田中)

「製造活動・営業活動の客体・業種・業界による区分をガイドラインの策定においても意識した上で、中小企業向けの関連主体の理解を促す枠組みをガイドライン策定に内在させるべきこと」

会社法により設立された法人の内部統制については実質支配基準による連結先に対する内部統制とコンプライアンスの問題があるが、この連結先に対する内部統制およびコンプライアンスの対象となることに鑑み、AIシステムに関する各種問題についての内部統制システムの射程について、連結先に対する配慮を必要とすることを基本として明確にしておくべきである。

その場合自分が仮にAIシステムに関するガイドラインでどのようなことを考えなければならないかわかりにくい製造業者等が何を参照としたらよいかを考えると、上記連結先は連結元の内部統制の枠組みを連結元により連結先に周知させることで理解の共有を進めることとすることを確信的に明記することで周知を明確にすべきである。

上場企業においては殊にこの対応により、公開会社として自らのグループについての内部統制を自らの責任で投資家向けに配慮する動機づけができるであろうが、上場していない会社であっても、ある程度規模以上の経済主体については、連結主体を持つことを自覚して、自らの責任で内部統制の枠組みを踏まえて対応する動機づけができるであろう。

必要な情報は検討はガイドラインを踏まえて自らの責任で枠組みを理解することに必要な情報等を適宜集めたうえで検討することになるだろう。

ではそのような連結元を有さない中小企業等の主体はどうするか。

そのような対応は人的・物的にも必ずしも期待できるとは限らない。

従って、彼らの理解を促すために少なくとも製造活動・営業活動の客体・業種・業界による区分をガイドラインの策定においても意識した上で、中小企業向けの関連主体の理解を促す枠組みをガイドライン策定に内在させるべきである。

これがないと、上記対応が可能な主体とそれが無理な主体の間でこのガイドラインの存在自体が、情報の非対称性を促進することで、競争制限的な状況を深めることになりかねないのではないだろうか懸念するゆえんである。

少なくとも現行の独禁法的な視点を踏まえた下請け法的な取引先に対する配慮の視点を内在させたガイドラインの構造的な構築を目指すべきであると考えられる必要があるように思われる。

差出人: AIネットワーク(総務省情報通信政策研究所)  
宛先: [REDACTED]  
件名: FW: 【案件番号:145208852】 パブリックコメントに関する提出意見の配信(受付番号: 201701310000392117)  
日付: 2017年1月31日 22:05:53

---

-----Original Message-----

From: public-com-egov@e-gov.go.jp [mailto:public-com-egov@e-gov.go.jp]  
Sent: Tuesday, January 31, 2017 10:05 PM  
To: ai.network@soumu.go.jp  
Subject: 【案件番号:145208852】 パブリックコメントに関する提出意見の配信(受付番号:201701310000392117)

パブリックコメントに関する意見提出先窓口担当者 様

電子政府の総合窓口(<http://www.e-gov.go.jp/>)から貴府省宛に  
パブリックコメントに関する意見が提出されましたので、配信します。

受付番号:201701310000392117  
受信日付:2017/01/31 20:21:14

案件番号:145208852

案件名:

「AI開発ガイドライン」(仮称)の策定に向けて整理した論点に関する意見募集  
宛先府省名:総務省

郵便番号

住所

氏名

連絡先電話番号

利用者メールアドレス

提出意見:

以下、意見を行う。

「Don't be Evil」

だけで良いと考える。

簡単に、「AIが悪をなさぬように、AIを用いて悪をなされないように心がける」とし、軽量の法令や通知とするだけで良いのではないかと考える。

なお、「人間の尊厳」とは、悪事を行ったり、無駄話をしたりする事ではない事には釘を刺しておくべきであると考え。

「経理や人事や医療関係者がAIによって不正を指摘されたので、人間の尊厳(や最終利用者の利益)を害された」などとはならないという事であるが、厚生労働省所管事業分野や、犯罪の関係する分野において、犯罪者の天敵となりうるAIは「人間の尊厳」によってその適用・運用をねじ曲げられる可能性があるものである。

不正を排除し、世の中が効率的で安全になるように、AIが使われるようになる事を望むが、利便性・可用性は自ずから伸びるものであろうから、やはり指針としては「Don't be Evil」という様な簡単なものだけで良いのではないかと考える(1行で済ませるべきであるとは言わないが。)

意見は以上である。

-----  
電子政府の総合窓口



<http://www.e-gov.go.jp/>

-----  
メール識別No:0000336959

差出人: AIネットワーク(総務省情報通信政策研究所)  
宛先: [REDACTED]  
件名: FW: 【案件番号:145208852】 パブリックコメントに関する提出意見の配信(受付番号: 201701310000392127)  
日付: 2017年1月31日 22:05:54

---

-----Original Message-----

From: public-com-egov@e-gov.go.jp [mailto:public-com-egov@e-gov.go.jp]  
Sent: Tuesday, January 31, 2017 10:05 PM  
To: ai.network@soumu.go.jp  
Subject: 【案件番号:145208852】 パブリックコメントに関する提出意見の配信(受付番号:201701310000392127)

パブリックコメントに関する意見提出先窓口担当者 様

電子政府の総合窓口( <http://www.e-gov.go.jp/> )から貴府省宛に  
パブリックコメントに関する意見が提出されましたので、配信します。

受付番号:201701310000392127  
受信日付:2017/01/31 20:57:19

案件番号:145208852

案件名:

「AI開発ガイドライン」(仮称)の策定に向けて整理した論点に関する意見募集  
宛先府省名:総務省

郵便番号: [REDACTED]

住所: [REDACTED]

氏名: NPO日本ネットワークセキュリティ協会社会活動部会 [REDACTED]

連絡先電話番号: [REDACTED]

利用者メールアドレス: [REDACTED]

提出意見:

・「設計時点で想定し得なかった処理がなされることが社会にとっての脅威となるリスクを考慮し、社会的な安全を確保するためのフェールセーフ機構を設けるべき」  
と明確に設計者・開発者が考えなければならないこと(つまり、フェールセーフ機構の確保)を  
示すべきではないか？

・本ガイドラインが広く活用されるためにも、対象となるAIは定義すべきである。  
そうしなければ、開発者が参照すべきか否かも判断しにくい。

・利用者が留意すべき点について議論されること自体は価値があるが、「開発ガイドライン」である以上、  
重きを置くべきは開発者向けの内容で、特に利用者に過度な責任や義務が生じないよう配慮が必要である。

-----  
電子政府の総合窓口

<http://www.e-gov.go.jp/>  
-----

メール識別No:0000336969

差出人: AIネットワーク(総務省情報通信政策研究所)  
宛先: [REDACTED]  
件名: FW: 【案件番号:145208852】 パブリックコメントに関する提出意見の配信(受付番号: 201701310000391996)  
日付: 2017年2月1日 13:19:50

---

-----Original Message-----

From: public-com-egov@e-gov.go.jp [mailto:public-com-egov@e-gov.go.jp]  
Sent: Tuesday, January 31, 2017 6:05 PM  
To: ai.network@soumu.go.jp  
Subject: 【案件番号:145208852】 パブリックコメントに関する提出意見の配信(受付番号:201701310000391996)

パブリックコメントに関する意見提出先窓口担当者 様

電子政府の総合窓口( <http://www.e-gov.go.jp/> )から貴府省宛に  
パブリックコメントに関する意見が提出されましたので、配信します。

受付番号:201701310000391996  
受信日付:2017/01/31 15:16:37

案件番号:145208852

案件名:  
「AI開発ガイドライン」(仮称)の策定に向けて整理した論点に関する意見募集  
宛先府省名:総務省

郵便番号: [REDACTED]  
住所: [REDACTED]  
氏名:産業技術総合研究所 情報・人間工学領域 人工知能研究センター  
連絡先電話番号: [REDACTED]  
利用者メールアドレス: [REDACTED]  
提出意見:

1.制御可能性の原則について

「実験室等、閉鎖された空間でAIの制御可能性について実験」を行った場合に、一般社会で利用されるのと同条件での実験を行うことが入力や出力を一般社会の場合と同一にすることができない場合があり、本質的に制御可能性を判断するために十分な実験とならない可能性があります。例として、ユーザーからの入力によって、システムの制御が変化する場合、一般社会のユーザーがどのような入力を与えるかを全て想定することが困難(MSのTayが悪意のあるユーザーにより暴言を学習してしまった

ケース)な場合があります。

また、研究されるAIシステムに関して、開発された際に危険性が及ばないように予見しなければいけないことが研究者に対して求められることになると、研究者の学問の自由に支障をきたすおそれがあるのではないかと思います。例えば、災害時の被害を最小化するために研究されている避難シミュレーション技術は、その技術の使い方次第では、逆に被害を最大化するよう悪用される可能性もあります。人工知能学会で検討されています「人工知能研究者の倫理綱領(案)」でも、研究者倫理の必要性を認識しながら、自由な研究を縛ることにならないよう、現在も議論がすすんでいて、このような学会の動きとも連携していく必要があると思います。

こうしたことを踏まえ、開発原則に、利活用側の視点も含めた方が良いのではないのでしょうか。開発者のコントロールが効かないガイドラインで縛られることにより、開発現場が萎縮してしまうのではないのでしょうか。留意事項等を盛り込む必要があるのではないかと考えます。

## 2. プライバシー保護の原則について

「予め」「設計段階において」行ったプライバシーの影響評価が、一般社会におけるプライバシーの影響と同一にならない可能性があります。例えば、一般社会においてどのような入力を与えられるか、事前に想定できない場合は、設計段階で入力を制限したり、禁止することができない可能性があります。

こうしたことを踏まえ、開発原則には、利活用側の視点も含めた方が良いのではないのでしょうか。開発者のコントロールが効かないガイドラインで縛られることにより、開発現場が萎縮してしまうのではないのでしょうか。留意事項等を盛り込む必要があるのではないかと考えます。

## 3. セキュリティ確保の原則および安全保護の原則について

国際的に既にコンセンサスが得られている工業標準等と調和するようなガイドラインとなる必要があると思います。例えば、家電業界、自動車業界等では組み込まれたソフトウェアを安全にするため、設計する資料の残し方や試験の方法について、国際工業標準をふまえた手順で開発が行われていますが、AIを用いるシステムの開発においても、これら既存の規範を尊重しつつ足りないものを補うようなガイドラインとなっていることが望ましいと思います。

例えば、AIを「情報収集、データ分析、体系化、推論、判断、人や社会への働きかけを行う技術」等のように定義して、これらの技術を含む現状の様々な技術開発のガイドライン(情報システムの信頼性向上に関するガイドライン(経済産業省、平成21年)等)や技術を活用する立場としてのガイドライン(遺伝子組み換え技術に関する各種ガイドライン等)を参考とされているとは思いますが、これらとAIならではのガイドラインを明確に分けると良いかと思えます。(つまり、AI開発ガイドラインの体系を、AI共通と分野別に分ける前に、他の技術共通とAI固有に分けます。)

## 4. 透明性の原則について

使いながら学習できるというAIの特徴を備えた製品を開発し、競争力を持った製品を市場に出すためには、各分野の事情に鑑み、過度に製造者に責任を負わせないようにするガイドラインになることが必要ではないかと思えます。例えば、医療系ではAIは参考情報を提示するための支援システムという位置付けで、判断して使用する人間(医師)の側に責任があると捉えられています。一方で、自動運転の分野では外界のものを動かす最終判断が人間となるか機械となるか微妙な位置付けになっていく場合もあるかと思えます。このように分野ごとのさまざまな事情をふまえて、「検証可能性を確保する」ための「合理的な範囲・水準」が行き過ぎないようにすることがAIの研究開発のためには大切になるかと思えます。

以上

-----  
電子政府の総合窓口

<http://www.e-gov.go.jp/>  
-----

メール識別No:0000336880