調査概要

# 保険における責任ある人工知能の推進

THE GENEVA ASSOCIATION
INSURANCE FOR A BETTER WORLD

ベンノ・ケラー
ジュネーブ協会デジタル・イノベーション特別顧問

*保険における人工知能の利用はリスク・プーリングを改善しリスクの減少、低減、予防に貢献することを通じて、保険会社やその顧客だけでなく広く経済的、社会的にもメリットをもたらす可能性があります。AI システムの導入を促進し、AI 利用によるメリットを実現するためには、保険会社はこの新しい技術の責任ある利用により顧客の信頼を得る必要があります。透明性および説明可能性、公平性、安全性、説明責任、プライバシーなど責任ある AI が備えるべき主要な原則を周知するため、保険会社は、社内のガイドラインや方針を策定し、関連リスクに対処するための適切なガバナンス構造を導入し、従業員とエージェント向けの包括的な研修プログラムを策定、実施しなければなりません。*

## 保険における人工知能

人工知能（AI）は、1970 年代から 1990 年代の終わりにかけて挫折と失望を繰り返した「AI 冬の時代」を経て、その後目覚ましい進歩を遂げました。今日、多くの保険会社が、保険のバリューチェーン全体にわたって日常業務を自動化したり人間の意思決定を支援したりする AI システムを展開しつつあります。こういったシステムは新しいタイプの学習アルゴリズムを、インターネット・メディアやモノのインターネット（IoT）など新しい情報源から得たデータの分析に結合します。

例えば、自然言語処理の進歩により AI システムは人間と「対話」し交流することが可能になります。保険会社は、複雑な顧客の問い合わせを識別し応答することができる 24 時間 365 日利用可能な会話型エージェント（例えばチャットボット）の活用を次第に増加させています。さらに、AI システムは画像内の物を「見て」認識し、関連情報を抽出することができます。このようなコンピューター・ビジョンを活用して、保険会社は、引受や保険金請求のプロセスに使用するデータを書面や画像から抽出するといった定型的な手作業や認識業務を自動化することができます。

AI システムは、複雑なデータのパターンや相関関係を人間では困難または不可能な方法で検出することに優れています。こうして識別されたパターンは、分類、回帰、クラスタリングなど保険ビジネス・モデルで重要な役割を果たす分析的作業の基礎となります。従来保険で使用されてきたモデリング・アプローチ（一般化線形モデルなど）と比べると、AI システムは変数間の複雑な非線形関係を学習できる分、はるかに正確な予測を提供することができます。

学習アルゴリズムがさらに進歩すれば、より多くの分野で AI システムによる標準化された意思決定が、人間の監督の下で自律的に行われることができるようになります。

## 保険における AI 利用がもたらすメリット

AI システムの導入を促進し AI 利用によるメリットを実現するためには、保険会社はこの新しい技術の責任ある利用により顧客の信頼を得る必要があります。また AI は、保険会社がリスクの減少、低減、予防に関する役割を強化するのに貢献することもできます（表 1 参照）。

***表 1: 保険における AI 利用がもたらす社会経済的メリット***

| |
|---|
| **リスク・プーリング範囲の拡大** |
| • 特定個人向けの商品（例えば、持病のある個人のための生命保険）へのアクセスを容易にすることで、新規顧客層および今まで保険対象外だった顧客層に保険適用を拡張<br>• リスクに関する知見（サイバーリスク等）を高めることで、保険適用可能なリスクの範囲を拡大 |
| **リスク・プーリングのコスト削減** |
| • 特定の作業の自動化、リスク評価の改善、モラルハザード・逆選択（adverse selection）の減少により費用効率の高い保険を提供 |
| **リスクの低減および防止** |
| • リスクの低減、防止に役立つリスク関連の新たな知見を提供<br>• 損失の削減を可能にする早期警戒システムが可能になる |

*出典：ジュネーブ協会*

## 責任ある AI に関する諸原則

ここ数年、AI の責任ある利用のために何が必要かという点について激しい議論が続く中、さまざまな政府や非政府機関から AI の倫理的利用のためのガイドラインが発表されました。こういったガイドラインのどれを分析してもその内容は、責任ある AI に関する 5 つの基本原則に　収斂します。すなわち（1）*透明性および説明可能性*、（2）*公平性*、（3）*安全性*、（4）*説明責任*、（5）*プライバシー*です。ただし、ある特定の文脈の中で倫理的原則やガイドラインをどのように実施すべきかという点に関しては不確実な部分がかなり残っています。

保険の分野では、これらの基本原則は長い間重要な役割を果たしてきました。実際に、保険法、プライバシーおよびデータ保護に関する法律、差別禁止に関する法律、監督要件などさまざまな法律や規制が、保険会社の公平、透明かつ説明可能な活動のあり方について、またプライバシーの保護について規定しています。

しかし、AI の利用は保険会社に次のような難しい疑問を投げかけています。基本原則を実施した場合、どのようなトレード・オフが生じるのでしょうか。保険会社はどうすれば基本原則の順守を推進することができ、また順守していると実証できるのでしょうか。この目的を達成するため、現在のガバナンス体制やリスク管理の枠組みに何らかの変更を加える必要があるとすれば、それはどのような変更でしょうか。

保険監督当局も、これらの問題について検討を重ねているところです。

ジュネーブ協会の報告書「*保険における責任ある人工知能の推進*」は、公平な機械学習に関するコンピューター・サイエンスの発展の現状を念頭に置いて、AI 利用に関する主要な倫理ガイドラインを分析し、これらガイドラインを保険の分野で適用する方法について検討しています。上記の疑問に対する決定的な答えは出そうにありません。しかし、このジュネーブ協会の報告書が目指しているのは、保険の分野で責任ある AI の基本原則を適用する場合に生じる主なトレード・オフを特定し検討するという取組みに寄与することにあります。上で述べた諸原則はいずれも重要ですが、この報告書では特に 1）透明性および説明可能性と、2）公平性の原則に焦点を当てています。なぜなら、保険の分野ではこれらの原則が特に複雑な問題を提起しているからです。

## 透明性および説明可能性

透明性や説明可能性は、顧客やその他利害関係者との信頼関係を構築するうえで重要です。いくつかのガイドラインでは、個人が自分に影響のある決定に対し是正を要求できるためには透明性や説明可能性が重要であるとしています。ある決定が個人に与える影響が重大である場合には、説明の提供は特に重要です。

したがって、必要とされる説明可能性のレベルは、誤ったあるいは不正確なアウトプットによる影響がどういう環境で生じ、またどの程度重大なのかによって変わってきます。また、アルゴリズムの決定に関する解釈可能性は、AI システムの性能の評価とその継続的改善にとっての不可欠な前提条件であり、従って健全なデータ・サイエンスにとっての不可欠な前提条件となります。

意味のある説明を提供するというのは難題です。「ブラックボックス」アルゴリズムの一部は、より高い精度を得るのと引き換えに性質上複雑な構造になっているため、その解釈や説明は難しいのです。近年、コンピュータ・サイエンスでは「ブラックボックス」アルゴリズムに関する解釈と説明の難しさを克服するための取組みが相当程度行われてきました。例えば、*リバース・エンジニアリング・アプローチ*は、解釈可能なアルゴリズムの代替[1]を構築します。この代替という最近開発された技術に関しては今後さらに理解を深めていく必要があります。*設計的アプローチ*は、予測に関して一定の制約条件を課すという方法をとります。

ある意思決定に関してさまざまな変数の役割を説明することができない場合、信頼関係を醸成するためには、独立機関による AI システムの認証など他のアプローチを利用することもできます。とはいえ、解釈可能なモデルの適用が推奨されます。モデルの結果により、顧客が重大な影響を受ける場合には特にそうであると言えます。AI システムをリスク選択や料率設定のために利用する場合には、保険適用するリスクに関連するデータソースの利用を顧客が直感的に理解できるような方法で行うことにより、AI システムに対する信頼を醸成することができます。過度に複雑なモデルに利点があるからといって、そのことが解釈可能性の低下を正当化するわけではありません。

## 公平性

公平な決定に関しては相互に排他的な定義がいくつか存在しており、どうすれば公平性が確保されたといえるのかは非常に難しい問題です。通常、公平性は自由、尊厳、自己決定、プライバシー、無差別、平等、多様性といった多くの異なる価値と関連しています。これらの価値は多くの場合、文化的背景などの文脈の中で解釈される必要があります。したがって公平性に関して普遍的な基準を提示することは不可能です。

意思決定は、ある特定の個人又は個人の集団を差別せず、不利益を与えないという意味で公平であるべきです。したがって、AI が主導する決定において不公平なバイアスを排除または最小化することは、倫理ガイドラインの中核となる要件です。

人間には偏見がないわけではありません。実際には、人工知能の利用は状況によっては意思決定の公平性を高める可能性があります。しかし、大規模に展開された場合、AI システムのわずかなバイアスでも多数の個人に影響を及ぼしかねません。したがって、データ・サイエンティストにとっての課題は、特定の集団をシステム的に不利な立場に置く可能性のあるバイアスを特定し、測定し、低減することにあります。

保険に関しては、以下に述べる公平性の概念が特に重要です。

*保険数理上の公平性*は、同種のリスクは同じように取り扱われるべきであるとし、個人の支払う保険料は実際のリスクに対応していることを要求します。

*差別がない*とは、個人ではどうすることもできない要因がある場合、特にその要因が社会的に保護された集団に関するものである場合に、当該個人の支払う保険料がその要因に基づいて決定されていないことを意味します。差別がないことは集団的公平性（group fairness）の概念と関連しています。集団的公平性とは、保護属性（性別、民族、性的指向など）で定義づけられた集団が目指す目標であり、保護属性を持たない集団と同様の取扱いを受けまたは同様の結果を得ること、またセンシティブな属性に（部分的に）基づく決定と全く異なる取扱いを受けないことを意味します。集団的公平性は、差別的効果（disparate impact）の概念を補完するものです。差別的効果とは、特定のセンシティブな属性を持つ人々に度を超えた危害を加えるまたは利益を与えることを言います。

*連帯あるいは相互扶助*は保険の重要な特徴としてよく取り上げられます。AI システムとデータ分析によって保険の個別化が進むと、特定の集団に対して例えば手の出ないような高額の保険料を請求したり、保険による保障提供を完全に拒否したりするなどの不利益を与える可能性があります。

保険の種類や国・地域が異なれば、様々な公平性概念のもつ相対的な重要性も異なります。例えば、多くの国・地域では、連帯は医療保険の必須の特徴と考えられています。

潜在的バイアスを識別し、測定し、低減するためには、保険会社は AI を利用する都度、その利用環境特有の公平性を特定する必要があります。公平性は保険の分野では新しい概念ではないので、既存の枠組みや規範（保険数理倫理など）を活用すべきです。

1　アルゴリズムの代替とは、アルゴリズムからのアウトプットを模倣する、容易に解釈可能な近似的モデルであるといえます。

## 結論

組織内での AI の責任ある利用を促進するため、保険会社は以下に述べる 3 つの措置を検討すべきです。

### *1） AI 利用に関する社内ガイドライン・方針の策定*

保険における AI 利用に伴うメリット・リスクのトレード・オフに対する意識を高める上で、社内ガイドラインや方針は重要な役割を果たします。したがって、保険会社は、透明性・説明可能性および公平性の問題を取り扱う諸原則を含む関連ガイドライン・方針を策定し、適用しなければなりません。特にガイドラインは、AI を利用する場合のメリットとリスクをケース・バイ・ケースで評価する方法を明確にするのに役立ちます。アクチュアリー、リスク・マネジャー、データ・サイエンティスト、データ保護責任者は、このようなガイドラインおよび方針の策定、実行に関して緊密に協力しなければなりません。

これにより、保険会社は AI のガバナンスに関してリスクベース・アプローチを導入することができます。このアプローチは、個人に大きな影響を与える可能性のある AI システムの利用に対して特に焦点を当てるものです。大きな影響を与えるかどうかは、ある意思決定がその対象となる個人にもたらす結果と関係します。また AI が利用されたときの具体的状況によっても変わってきます。

### *2） 適切なガバナンス構造の導入*

保険における AI の責任ある利用が行われるためには、意思決定者（個人、委員会、部門）に説明責任を負わせるガバナンス構造が必要です。意思決定者は必要な能力、スキル、専門知識を持たなければなりません。また組織はトリガーやエスカレーションなど効果的なプロセスを確立すべきです。

ガバナンス・モデルは多種多様で、それぞれに長所と短所があります。最適な組織モデルは企業の構造と文化によって決まります。

### *3） 社内研修プログラムの策定と実施*

最後に、AI の責任ある利用を確実なものにするためには、さまざまな職務や管理レベルにおいて、AI の利用に伴うメリットとリスクについて認識される必要があります。この認識を高めるために、保険会社は、AI 利用のメリットとリスクおよび自社のガイドラインと方針について、包括的な研修プログラムの策定、実施を検討すべきです。このような研修プログラムは、様々な管理レベルの従業員、さらにエージェントやその他顧客対応に当たる職員を含め内部の意思決定に関わる様々な職員を対象とするのが理想です。