

在人工智能模型的保险应用中 避免不公平偏见



如今，人工智能（AI）技术正在不断地应用于保险价值链的许多领域中。公司可以使用人工智能技术来探索数据、收集到以往无法实现的洞察见解，或利用人工智能技术的高效率来加强或取代现有的模型。随着越来越多的公司探索人工智能并期望从中受益，他们可能会遇到关于此系统有用性、有效性和稳定性的问题。

随着人工智能应用的增加，一个已知的风险是可能出现不公平的偏见。了解人工智能系统中可能出现的不公平偏见的核心在于——定义什么是不公平偏见以及什么是公平。对于这项研究的目的来说，我们将不公平偏见定义为“对边缘化群体无法解释的不利结果”。

这项研究的目的是确定避免或缓解因使用人工智能模型而无意或加剧造成的不公平偏见的方法。本报告还提出了一个潜在的框架和方法，供保险公司在识别和减少其人工智能模型中的不公平偏见时考虑。框架和缓解性方法可能与刚刚开始进入人工智能领域的公司有关，也与那些在该领域更成熟的公司有关。虽然本报告中有些部分涵盖了广泛适用的框架和概念，但研究和建议还是以美国为发展重点所制定的。

本报告中的大部分建议都与拟议的模型开发框架中的某个领域相一致。此外，还提出了关于公司在人工智能开发和使用过程中采取的方法的基础性建议。这些基础性建议包括：

1. 监测监管环境以满足突发要求，不仅是来自保险领域，还有来自数据保护、隐私和贸易监管机构的要求。
2. 让利益相关者参与进来，树立有效监督的角色和责任。

Caveat and Disclaimer

The opinions expressed and conclusions reached by the authors are their own and do not represent any official position or opinion of the Society of Actuaries Research Institute, the Society of Actuaries or its members. The Society of Actuaries Research Institute makes no representation or warranty to the accuracy of the information.

3. 让员工掌握必要的工具和技能，以了解人工智能系统的复杂情况以及它们可能反映的偏见。
4. 进行模型风险评估，以定位风险位置。
5. 与模型风险管理相结合，以实现一致的治理方法。

由于不公平偏见可能发生在整个保险价值链的各个环节，因此企业可以采用一个人工智能治理模型，该模型不仅适用于当前的实践，而且能够对未来潜在的监管规则、客户期望和市场细分做出扩展和回应。

模型开发框架

这项研究还提出了一个拟议的模型开发框架，其中包括已确定的风险、有助于减少在人工智能中引入不公平偏见的可能性的潜在控制措施，以及一个建议的阶段式方法。该框架对以下几个部分进行了阐述：

1. 价值界定——确定建立模型是否有任何商业价值、模型的成功指标、模型的使用方法，以及模型的风险；
2. 价值探索——评价和定义成功标准、需要什么数据、测试公平性的措施，以及跟踪模型实验和版本的方法；
3. 价值交付——将模型应用到生产环境中，并转换到日常业务的生产状态；
4. 价值管理——捕捉模型的价值，并持续向高级管理层报告模型的性能没有退化以及为保持模型的稳健性所需的任何变动。

这个框架仅为模型开发框架提供思路，并非一概而论。

研究背景

人工智能（AI）技术正在不断地应用于保险价值链的许多领域中。在本报告的背景下，人工智能被定义为能够感知数字或实体世界的计算机系统，处理它所感知的信息，并在大多数情况下，采取那种通常可能需要人类智慧或推理的行动。公司可以使用人工智能技术来探索数据、收集到以往无法实现的洞察见解，或利用 AI 技术中的高效率来加强或取代现有的模型。这里的“模型”是指任何用于转换或处理现有数据的工具，以此支持新见解、新数据或业务建议。越来越多的公司探索人工智能并期望从中获益，但他们遇到的问题往往多于获得的答案。最近受到极大关注的一个基本问题是从业者如何识别并主动降低人工智能系统所引发的不公平偏见的风险？

在本研究中，我们将不公平偏见定义为对边缘化群体无法解释的不利结果。存在不同结果背后的原因在于人类的判断和人类一开始所做决定的方式的变化。若两个人接触到相同的事实和情况，却得出了不同的结论——这也是一种常见的情况。在一个保险设定的背景下，如果人类决策的这种差异对边缘化群体造成不公平的影响，这将被认为是不公平偏见。随着保险业越来越多地依靠人工智能模型来做出或建议此类决策，不公平偏见的风险正在从人类决策转移到人工智能模型和作为其基础的算法上。

研究介绍

人工智能的应用使各行业能够重新定义其经营方式——从金融科技等新兴市场到网上银行等新性能，各机构必须采用人工智能以维持市场竞争力。保险业也不例外。不断变化的客户期望、利润压力、成本的降低和流程的简化正在推动整个行业对人工智能的投入。

这些投入包括人工智能应用，如自适应对话代理人（即聊天机器人），在整个索赔过程中，聊天机器人的应用实现了以最低限度的人工参与重塑客户交互体验。在核保中提出风险等级的建议，使操作效率得到改善，或使评估更加个性化。自动化欺诈检测帮助保险公司通过识别索赔数据中的异常情况来降低风险。在整个业务职能中，保险公司正在使用人工智能来改善运营并产生新的价值主张。

但这些进步也给保险公司、商业领袖、监管者和技术专家带来了发展和应用人工智能的各种担忧。新的数据形式和组合的出现引起了数据质量、隐私和使用方面的问题。人工智能系统的复杂性引发了人们对其缺乏透明性和可解释性的担忧。社会变化使人质疑这些系统的长期稳定性，以及它们与不断变化的标准、期望和行为的一致性。对人工智能依赖性的增加需要对自主或人工智能辅助的决策进行审议。这些担忧包括不公平偏见的影响。

随着各公司在保险价值链的不同领域探索人工智能的新应用，整个保险业对人工智能的应用正在增加。目前，保险业内的人工智能从业人员数量不多，但保持持续增长的态势。正因为如此，目前存在着既缺乏建立新的人工智能模型的人才，又缺乏理解、验证和管理现有模型的人才的现象。越来越多的人期望保险公司能够认识到其有责任发现和最小化不公平偏见，这同时也给公司的治理职能带来了巨大的压力。虽然行业团体和美国监管机构都认识到解决这些问题的紧迫性，但可操作的标准和新的监管准则并没有跟上行业发展的步伐，使保险公司不得不在创新和负责任的发展之间进行权衡。保险公司对他们的客户、股东和社会负有责任，因此在如何使用和分析他们的数据方面需要做出负责任和合乎道德的决定。

保险公司有各种各样的产品，这些产品依赖于不同的数据源和数据集，以更好地了解其目标人群。人寿保险和年金、商业、个人财产和意外保险以及健康保险产品通常针对个人或个体公司。另外，团体保险和养老保险产品更有可能为个人群体服务。虽然分析所需的数据有一定的共性，但使用和依赖数据进行分析的方式可以有显著不同。即使在一个特定的产品系列中，分析方法和所依赖的数据源也有很大的差异。在人寿保险领域，保单的期限较长，通常与几年的时间或个人的全生命周期相联系，而且保险费率通常在保单期限内得到保证。这就要求保险公司保留长期的历史数据，并经常使用这些数据来预测未来的长期情况。另外，健康和财产险索赔每年都会被审核，并且保险费率可能每年都会有波动。对于团体保险来说，其对数据的需求甚至更大，在做出影响整个团体的决定之前，需要考虑某一特定团体或部分客户的数据。基于考虑到的团体规模和团体数量，这可能导致公司需要分析非常庞大的数据集。每个产品具体细微的差别都会改变数据在人工智能模型中的使用方式。因此，人工智能的不同应用可能会在不同类型的保险中无意中引入了不公平偏见。本文旨在解决保险公司在所有保险领域可能面临的挑战和风险，同时也评估了仅适用于部分保险产品系列的人工智能的具体使用案例。

本报告研究了在美国保险公司人工智能应用中引入不公平偏见所带来的高风险以及在模型开发过程中减少风险的方法。由于企业依靠数据来建立他们的人工智能模型，本报告还探讨了多个数据元素或数据源的组合，以及这些组合可能导致的不公平偏见。此外，该研究还涉及了公司如何开展与人工智能模型开发有关的治理做法，特别是与不公平偏见有关的治理做法。

About The Society of Actuaries Research Institute

Serving as the research arm of the Society of Actuaries (SOA), the SOA Research Institute provides objective, data-driven research bringing together tried and true practices and future-focused approaches to address societal challenges and your business needs. The Institute provides trusted knowledge, extensive experience and new technologies to help effectively identify, predict and manage risks.

Representing the thousands of actuaries who help conduct critical research, the SOA Research Institute provides clarity and solutions on risks and societal challenges. The Institute connects actuaries, academics, employers, the insurance industry, regulators, research partners, foundations and research institutions, sponsors and non-governmental organizations, building an effective network which provides support, knowledge and expertise regarding the management of risk to benefit the industry and the public.

Managed by experienced actuaries and research experts from a broad range of industries, the SOA Research Institute creates, funds, develops and distributes research to elevate actuaries as leaders in measuring and managing risk. These efforts include studies, essay collections, webcasts, research papers, survey reports, and original research on topics impacting society.

Harnessing its peer-reviewed research, leading-edge technologies, new data tools and innovative practices, the Institute seeks to understand the underlying causes of risk and the possible outcomes. The Institute develops objective research spanning a variety of topics with its [strategic research programs](#): aging and retirement; actuarial innovation and technology; mortality and longevity; diversity, equity and inclusion; health care cost trends; and catastrophe and climate risk. The Institute has a large volume of [topical research available](#), including an expanding collection of international and market-specific research, experience studies, models and timely research.

Society of Actuaries Research Institute
475 N. Martingale Road, Suite 600
Schaumburg, Illinois 60173
www.SOA.org